Research Topics

Duetting and turn-taking patterns of singing mammals: From genes to vocal plasticity, and beyond

Edited by
Patrice Adret, Dena Jane Clink and Charles T. Snowdon

Coordinated by
Sofya Dolotovskaya

Published in
Frontiers in Ecology and Evolution

ISSN 1664-8714
ISBN 978-2-8325-3681-0
DOI 10.3389/978-2-8325-3681-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Duetting and turn-taking patterns of singing mammals: From genes to vocal plasticity, and beyond

Topic editors
Patrice Adret — Universidad Autónoma Gabriel René Moreno, Bolivia
Dena Jane Clink — Cornell University, United States
Charles T. Snowdon — University of Wisconsin-Madison, United States

Topic Coordinator
Sofya Dolotovskaya — Primate Genetics Laboratory, German Primate Center, Germany

Citation
Adret, P., Clink, D. J., Snowdon, C. T., Dolotovskaya, S., eds. (2023). *Duetting and turn-taking patterns of singing mammals: From genes to vocal plasticity, and beyond*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3681-0

We respectfully dedicate this Research Topic to the late Emeritus Professor Charles T. Snowdon. As a distinguished scholar in the realms of Animal Behavior and Primatology, we were honored to have Chuck onboard for the most part of this project.

This Research Topic has been developed in collaboration with Sofya Dolotovskaya (German Primate Center, Göttingen, Germany).

Table of contents

frontiers | Frontiers in Ecology and Evolution

TYPE Editorial
PUBLISHED 14 September 2023
DOI 10.3389/fevo.2023.1279635

OPEN ACCESS

EDITED AND REVIEWED BY
Jordi Figuerola,
Spanish National Research Council
(CSIC), Spain

*CORRESPONDENCE
Patrice Adret
✉ patrice.adret@gmail.com

RECEIVED 18 August 2023
ACCEPTED 30 August 2023
PUBLISHED 14 September 2023

CITATION
Adret P, Clink DJ and Dolotovskaya S (2023) Editorial: Duetting and turn-taking patterns of singing mammals: from genes to vocal plasticity, and beyond. *Front. Ecol. Evol.* 11:1279635.
doi: 10.3389/fevo.2023.1279635

Editorial: Duetting and turn-taking patterns of singing mammals: from genes to vocal plasticity, and beyond

Patrice Adret[1]*, Dena J. Clink[2] and Sofya Dolotovskaya[3]

[1]Museo de Historia Natural Noel Kempff Mercado, Santa Cruz de la Sierra, Bolivia, [2]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States, [3]Behavioral Ecology and Sociobiology Unit, German Primate Center – Leibniz-Institute for Primate Research, Göttingen, Germany

KEYWORDS

primates, coordinated singing, turn-taking, social cooperation, evolutionary history, pairbonding, machine learning, language evolution

Editorial on the Research Topic
Duetting and turn-taking patterns of singing mammals: from genes to vocal plasticity, and beyond

One of the greatest challenges in evolutionary biology is tracing back the origins of human speech in the absence of fossilized vocal sounds. Since Darwin's (1871) landmark treatise on the evolution of spoken language and music, the search for phylogenetic precursors of these two intimately connected fields has remained a major endeavor of scientific research (ten Cate and Honing, 2023).

As a prime signaling channel, acoustic communication is above all socially interactive and can take many forms in the animal kingdom, thus providing an evolutionary substrate for the emergence of human musicality and conversational speech (Snowdon et al., 2015; Levinson, 2016; Snowdon, 2017, 2021; Savage et al., 2020); it is also a useful system for understanding the evolutionary processes that shape phenotypic variation. In the wake of a Research Topic entitled "*Turn-taking in Human Communicative Interaction*" (Holler et al., 2015), the present collection of 13 articles brings together 47 authors who share ideas, data and methods on the theme of vocal duetting (i.e., a coordinated vocal exchange between two individuals who alternate and/or overlap their contributions) and turn-taking (i.e., a vocal exchange based on active overlap avoidance between individuals who take turns as callers and listeners) in singing mammals.

Approximately 6,400 living species of mammals populate Earth (Burgin et al., 2018). Of those that have been the subject of detailed bioacoustics analyses, only few have evolved the capacity for singing and fewer still have been reported to coordinate song (i.e., a string of modulated vocal sounds delivered in a predictable pattern), either as an intra- or intersexual display (Figure 1). Arguably, the champions are the "singing primates" (De Gregorio et al., 2022), an assemblage of ~70 arboreal species – roughly 14% of all extant primates – distributed in Southeast Asia (tarsiers, a Mentawai langur, gibbons and siamang), Madagascar (indri, Milne-Edward's sportive lemur) and South America (titi monkeys).

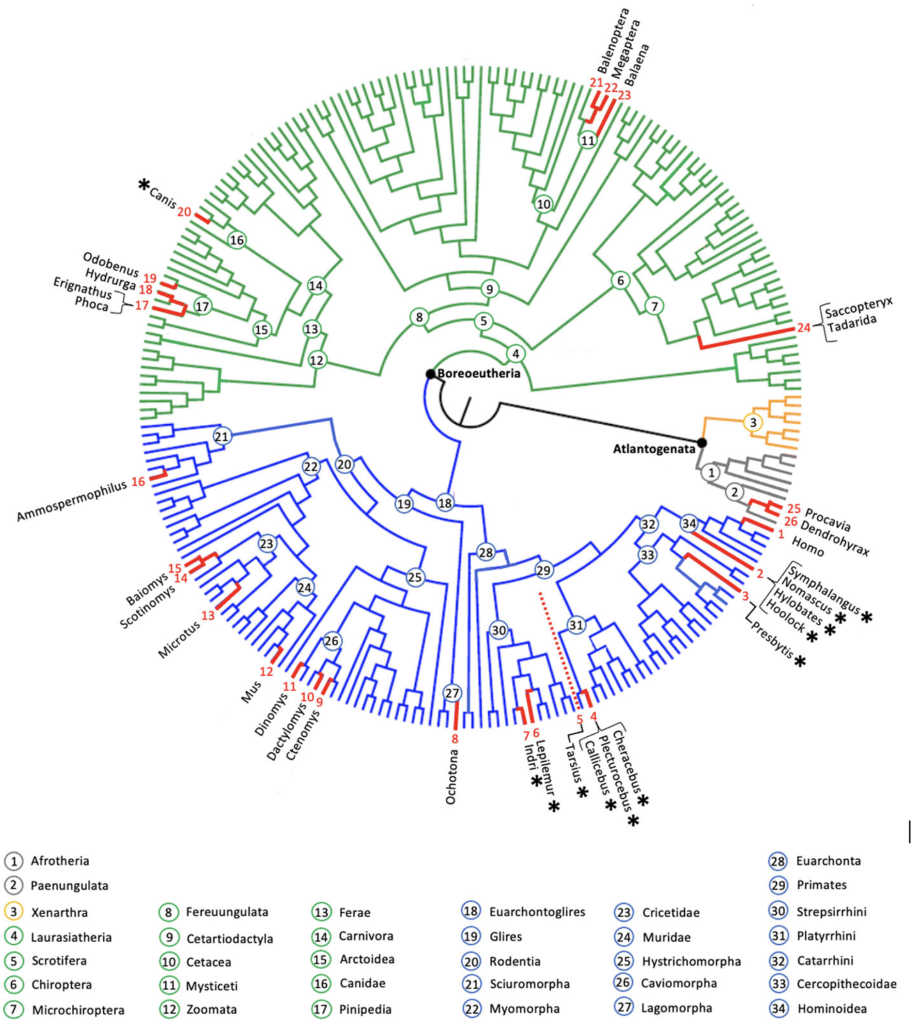


FIGURE 1

Ortho-phylogenetic tree of placental mammals (modified from Foley et al., 2023, with permission from the authors). Highlighted in red are mammalian genera endowed with species that have the ability to sing. Note that we consider howling in canids as a song-like vocalization performed as a duet by mated pairs or as a group chorus. In non-primate mammals, song appears to be a male prerogative (except in canids) leading some species to engage in intra-sexual counter-singing. Asterisks denote taxa in which mated pairs produce coordinated songs. The red dotted line corresponds to the Tarsiiformes, not included in the phylogenetic analysis of Foley et al. (2023). Where taxa differed from those originally reported in Foley et al. (2023), we elected the terminal branch which was most closely-related phylogenetically. Red numbers on the outer ring of the cladogram specify the relevant literature: 1- Mehr et al. (2019); 2- Geissmann (2002); 3- Tilson and Tenaza (1976); 4- Caselli et al. (2014); Adret et al. (2018); 5- Shekelle et al. (2019); 6- Méndez-Cárdenas and Zimmermann (2009); 7- Pollock (1986); 8- Somers (1973); 9- Amaya et al., (2016); 10- Emmons (1981); 11- Eisenberg (1974); 12- Holy and Guo (2005); 13- Rutovskaya (2020); 14- Banerjee et al. (2019); 15- Miller and Engstrom (2007); 16- Bolles (1988); 17- Ray et al. (1969); Fitch (2006); 18- Thomas and Golladay (1996); 19- Sjare et al. (2003); 20- Koler-Matznik et al. (2003); 21- Stafford et al. (2012); 22- Payne and McVay (1971); 23- Buchan et al. (2014); 24- Behr and von Helversen (2004); Bohn et al. (2009); 25- Demartsev et al. (2017); 26- Rosti et al. (2020).

While distinct phenotypes of coordinated acoustic signaling have been identified in a wide range of organisms (Pika et al., 2018; de Reus et al., 2021), a robust link between duetting, a long-lasting pair bond, and a non-migratory lifestyle marked by year-round territoriality is found primarily in homeothermic animals. To what extent these differences in communicative abilities are driven by genes, experience and the environment remains an active area of research. Notably, significant progress has been made by expanding the stage of experimentation to the emitter-receiver taken as the unit of investigation. Indeed, recent technological advances such as wireless dual recordings of vocalizations and concomitant brain activity can be considered as the gold standard to investigate social coordination in mammals (Rose et al., 2021), let alone the promising expectations from the field of Artificial Intelligence (Rutz et al., 2023).

Collectively, the contributions to this Frontiers Research Topic cover the four questions fundamental to behavioral research, namely causation, ontogeny, function and evolution. A final section is devoted to machine learning techniques with the goal of supporting primate conservation efforts.

1 Overview

Two articles introduce the Research Topic. Adret's mini-review provides a concise synthesis of developmental plasticity in the coordinated songs of songbirds and singing primates. The broad relevance of linking duetting behavior and its neural underpinnings is made. Similarities and differences between the two fields of research are highlighted to help guide ongoing research. In a perspective article, Vanderhoff and Bernal Hoverud focus on the coordinated vocal exchanges of non-primate mammals, pointing out inconsistencies in term usage such as duetting, antiphony, counter-singing and turn-taking. Moving on, the authors present a case study from the elusive South-American bamboo rats and encourage researchers to search for more examples of mammals that communicate via coordinated vocalizations.

2 Causation

In a thought-provoking contribution, Ravignani et al. hypothesize on the role of the corpus callosum (CC) in facilitating turn-taking and duetting (TTD) behavior in mammals. Drawing on similarities from animal behavior, language and music, the authors argue that CC and TTD likely co-evolved to speed up interhemispheric communication during vocal exchanges in eutherian mammals. They propose to test this hypothesis by comparing CC size in duetting and non-duetting pairs of closely related mammalian species.

3 Ontogeny

Four articles compose this section. Abreu and Pika thoroughly review the development and acquisition of turn-taking skills in non-human mammals. Using a top-down approach, the authors highlight four building blocks of conversational speech and identify research biases and gaps after methodically sifting key-articles in this emerging system. The authors pinpoint fruitful research avenues to stir more interest in this field that will improve our understanding of turn-taking for language evolution. Following a decade of field recordings collected from eight family groups of indris (*Indri indri*) in Madagascar, De Gregorio et al. track the social dynamics underlying this unique lemur song display, which combines elements of solo singing, duetting and chorusing. The authors report a clear stochastic process of vocal turns resulting from non-random dyads between group members. Interestingly, the study provides evidence that each parent alters its singing while interacting with an offspring. In a singular paper, Yi et al. highlight the occurrence of co-singing episodes between offspring and mothers in a "non-duetting" gibbon, *Hylobates moloch*. Twelve consecutive years of field observations revealed that these joint vocalizations are transiently expressed from two to seven years of age, with striking sex differences, after which mature individuals produce only sex-specific solo songs. Working in captive settings, Hradec et al. undertake an analysis of adult male songs in the Southern yellow-cheeked gibbon, *Nomascus gabriellae*. The authors highlight structural differences between unpaired and paired males although further studies are needed to disentangle the respective effects of age and social status on song structure.

4 Function

Two articles make up this section. Dolotovskaya and Heymann investigate the adaptive value of duetting with an observational study of six groups of coppery titi monkeys, *Plecturocebus cupreus*, from Peruvian Amazon. A systematic mapping of duet records during periods of female receptivity, gestation, and lactation allows the researchers to combine their data with relevant ecological variables. The ensuing multifactorial analyses support a resource defense mechanism as opposed to a mate guarding strategy. Experimenting at the National Primate Research Center in Davis (California, USA), Lau et al. provide preliminary data on duet song perception in female coppery titi monkeys, *Plecturocebus cupreus*. Audio playback tests conducted both before and after pairing reveal noticeable behavioral and hormonal changes linked to the reproductive cycle. The work adds an important component to the broad picture of primate duetting, especially from the listener perspective.

5 Evolution

Two articles cover this section. The evolution of signal design is central to the Comella et al. article focusing on the duets of Gursky's spectral tarsiers, *Tarsius spectrumgurskyae*, a nocturnal basal haplorrhine from Indonesia. Using unsupervised clustering methods, the authors show that individual pairs possess highly graded, sex-specific note repertoires, subject to morpho-physiological constraints between the rate of syllable repetition and note bandwidth. Such acoustic tradeoffs might represent one example of "species-universals" in vocal communication. Transcending the mechanistic view of duetting, Kaplan takes a multi-disciplinary and multimodal approach in formulating a "prosocial theory" for the evolution of human language. The author argues that the switch from self- to other-oriented behavior required expanding both cognitive and affective skills to foster intentional cooperation *after* social bonding has already been established. Within this framework, both gestural and vocal coordination were paramount to the emergence of human language.

6 Techniques

Passive acoustic monitoring (PAM) utilizes autonomous sensors for population surveys on broad spatial-temporal scales. On this basis, Clink et al. develop and test a machine learning approach for the automated detection and classification of female great calls in the Northern-grey gibbon, *Hylobates funereus*, on Malaysian Borneo. While performance of the open-source code for call detection was found satisfactory, the unsupervised clustering algorithm performed sub-optimally, thus impacting the ability to reliably discriminate individual females in the local population. Nonetheless, the proposed workflow constitutes a valuable effort on which further studies can build on. In a companion article using PAM, van Kuijk et al. investigate source level and detection range of duet songs in the cryptic red titi monkey, *Plecturocebus discolor*, of the Ecuadorian Amazon. To extract the target signal from audio recordings, the researchers apply a supervised template-based detection algorithm, which, compared with manual detection, significantly sped up data processing and will serve to implement future PAM studies of titi monkeys.

In conclusion, time will tell whether this Research Topic successfully achieved its goal to serve as a springboard for more empirical work in our quest to unravel pressing issues.

Author contributions

PA: Conceptualization, Writing – original draft, Writing – review & editing. DC: Writing – review & editing. SD: Writing – review & editing.

Acknowledgments

We wish to thank the Authors, Reviewers and invited Editors for their valuable contribution. PA is most grateful to Dr. Helen Kimbell and her team at Frontiers in Ecology and Evolution for inviting him to develop the theme of this article collection.

In memoriam

We respectfully dedicate this Research Topic to the late Emeritus Professor Charles T. Snowdon. As a distinguished scholar in the realms of Animal Behavior and Primatology, we were honored to have Chuck onboard for the most part of this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adret, P., Dingess, K. A., Caselli, C. B., Vermeer, J., Martínez, J., Luna Amancio, J. C., et al. (2018). Duetting patterns of titi monkeys (Primates: Callicebinae) and relationships with phylogeny. *Animals* 8, 178–211. doi: 10.3390/ani810 0178

Amaya, J. P., Areta, J. I., Valentinuzzi, V. S., and Zuffiaurre, E. (2016). Form and function of long-range vocalizations in a neotropical fossorial rodent: the anillaco tuco-tuco (Ctenomys sp.). *PeerJ* 4, e2559. doi: 10.7717/peerj.2559

Banerjee, A., Phelps, S. M., and Long, A. L. (2019). Singing mice. *Curr. Biol.* 29, R190–R191.

Behr, O., and von Helversen, O. (2004). Bat serenades–complex courtship songs of the sac-winged bat (Saccopteryx bilineata). *Behav. Ecol. Sociobiol.* 56, 106–115.

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., and Pollak, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *PloS One* 4 (8), e6746. doi: 10.1371/journal.pone.0006746

Bolles, K. (1988). Evolution and variation of antipredator vocalisations of antelope squirrels, Ammospermophilus (Rodentia: Sciuridae). *Z. Säugetierkunde* 53, 129–114.

Buchan, S. J., Hucke-Gaete, R., Rendell, L., and Stafford, K. M. (2014). A new song recorded from blue whales in the corcovado gulf, southern chile, and an acoustic link to the eastern tropical pacific. *Endang. Species Res.* 23, 241–252. doi: 10.3354/esr00566

Burgin, C. J., Colella, J. P., Kahn, P. L., and Upham, N. S. (2018). How many species of mammals are there? *J. Mammal.* 99, 1–14. doi: 10.1093/jmammal/gyx147

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., and Setz, E. Z. F. (2014). Vocal behavior of black-fronted titi monkeys (Callicebus nigrifrons): Acoustic properties and behavioral contexts of loud calls. *Am. J. Primatol.* 76, 788–800. doi: 10.1002/ajp.22270

Darwin, C. (1871). *The descent of man and selection in relation to sex* (London: John Murray, Albermarle Street).

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

Demartsev, V., Ilany, A., Kershenbaum, A., Geva, Y., Margalit, O., Schnitzer, I., et al. (2017). The progression pattern of male hyrax songs and the role of climactic ending. *Sci. Rep.* 7, 2794. doi: 10.1038/s41598-017-0 3035-x

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., et al. (2021). Rhythm in dyadic interactions. *Phil. Trans. R. Soc B* 376, 20200337. doi: 10.1098/rstb.2020.0337

Eisenberg, J. F. (1974). "The function and motivational basis of hystricomorph vocalizations," in *The biology of hystricomorph rodents*. Eds. I. W. Rowlands and B. J. Weir (London: Symposia of the Zoological Society), 211–247. 34:1-482.

Emmons, L. (1981). Morphological, ecological, and behavioral adaptations for arboreal browsing in *Dactylomys dactylinus* (Rodentia. Echimyidae). *J. Mammal.* 62, 183–189. doi: 10.2307/1380493

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Foley, N. M., Mason, V. C., Harris, A. J., Bredemeyer, K. R., Damas, J., Lewin, H. A., et al. (2023). A genomic timescale for placental mammal evolution. *Science* 380, 6643. doi: 10.1126/science.abl8189

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/s1464793101005826

Holler, J., Kendrick, K. H., Casillas, M., and Levinson, S. C. (2015). Editorial: Turn-taking in human communicative interaction. *Front. Psychol.* 6. doi: 10.3389/fpsyg.2015.01919

Holy, T. E., and Guo, Z. (2005). Ultrasonic songs of male mice. *PloS Biol.* 3, e386. doi: 10.1371/journal.pbio.0030386

Koler-Matznick, J., Brisbin, I., Feinstein, M., and Bulmer, S. (2003). An updated description of the new guinea singing dog (Canis hallstromi, troughton 1957). *J. Zool.* 261, 109–118. doi: 10.1017/S0952836903004060

Levinson, S. C. (2016). Turn-taking in human communication-origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting - a mechanism to strengthen pair bonds in a dispersed pair-living primate (Lepilemur edwardsi)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., et al. (2019). Universality and diversity in human song. *Science* 366 (6468), eaax0868. doi: 10.1126/science.aax0868

Miller, J. R., and Engstrom, M. D. (2007). Vocal stereotypy and singing behavior in baiomyine mice. *J. Mammal.* 88, 1447–1465. doi: 10.1644/06-MAMM-A-386R.1

Payne, R. S., and McVay, S. (1971). Songs of humpback whales: Humpbacks emit sounds in long, predictable patterns ranging over frequencies audible to humans. *Science* 173, 585–597. doi: 10.1126/science.173.3997.585

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* 285:20180598. doi: 10.1098/rspb.2018.0598

Pollock, J. I. (1986). The song of the indris (Indri indri; primates; lemuroidea): Natural history, form, and function. *Int. J. Primatol.* 7, 225–264. doi: 10.1007/BF02736391

Ray, C., Watkins, W. A., and Burns, J. (1969). The underwater song of erignathus (bearded seal). *Zoologica* 54, 79–83. doi: 10.5962/p.203251

Rose, M. C., Styr, B., Schmid, T. A., Elie, J. E., and Yartsev, M. M. (2021). Cortical representation of group social communication in bats. *Science* 374, eaba9584. doi: 10.1126/science.aba9584

Rosti, H., Pihlström, H., Bearder, S., Pellikka, P., and Rikkinen, J. (2020). Vocalization analyses of nocturnal arboreal mammals of the taita hills kenya. *Diversity* 12, 473. doi: 10.3390/d12120473

Rutovskaya, M. V. (2020). Acoustic communication in four species of subgenus alexandromys. *Russian J. Theriol.* 19, 21–36. doi: 10.15298/rusjtheriol.19.1.02

Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., and Blasi, D. E. (2023). Using machine learning to decode animal communication. *Science* 381 (6654), 152–155. doi: 10.1126/science.adg7314

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., et al. (2020). Music as a coevolved system for social bonding. *Behav. Brain Sci.* 44, e59. doi: 10.31234/osf.io/qp3st

Shekelle, M., Groves, C. P., Maryanto, I., Mittermeier, R. A., Salim, A., and Springer, M. S. (2019). A new tarsier species from the togean islands of central sulawesi, indonesia, with references to wallacea and conservation on sulawesi. *Primate Conserv.* 33, 65–73.

Sjare, B. L., Stirling, I., and Spencer, C. A. (2003). Structural variation in the songs of atlantic walruses breeding in the canadian high arctic. *Aquat. Mammals* 29, 297–318.

Snowdon, C. T. (2021). Animal signals, music and emotional well-being. *Animals* 11, 2670. doi: 10.3390/ani11092670

Snowdon, C. T. (2017). Learning from monkey "talk". *Science* 355, 1120–1122. doi: 10.3390/ani11092670

Snowdon, C. T., Zimmermann, E., and Altenmüller, E. (2015). Music evolution and neuroscience. *Prog. Brain Res.* 217, 17–34. doi: 10.1016/bs.pbr.2014.11.019

Somers, P. (1973). Dialects in southern rocky mountain pikas, ochotona princeps (Lagomorpha). *Anim. Behav.* 21, 124–137. doi: 10.1016/S0003-3472(73)80050-8

Stafford, K. M., Moore, S. E., Berchok, C. L., Wiig, ØVerifyt. a. t., Lydersen, C., Hansen, E., et al. (2012). Spitsbergen's endangered bowhead whales sing through the polar night. *Endangered Species Res.* 18 (2), 95–103. doi: 10.3354/esr00444

ten Cate, C., and Honing, H. (2023). "Precursors of music and language in animals. PsyArXiv," in *Oxford handbook of language and music*. Ed. D. Sammler (Oxford: Oxford University Press). doi: 10.31234/osf.io/4zxtr

Thomas, J. A., and Golladay, C. L. (1996). "Geographic variation in leopard seal (Hydrurga leptonyx) underwater vocalizationsInEds," in *Sensory systems of aquatic mammals*. Eds. R. Kastelein, J. A. Thomas and P. E. Nachtigall (Woerden, The Netherlands: DeSpil Publishers), 201–221.

Tilson, R. L., and Tenaza, R. R. (1976). Monogamy and duetting in an old world monkey. *Nature* 263 (5575), 320–321. doi: 10.1038/263320a0

frontiers
in Ecology and Evolution

MINI REVIEW
published: 25 March 2022
doi: 10.3389/fevo.2022.862196

Developmental Plasticity in Primate Coordinated Song: Parallels and Divergences With Duetting Songbirds

Patrice Adret*

Museo de Historia Natural Noel Kempff Mercado, Santa Cruz de la Sierra, Bolivia

Homeothermic animals (birds and mammals) are prime model systems for investigating the developmental plasticity and neural mechanisms of vocal duetting, a cooperative acoustic signal that prevails in family-living and pair-bonded species including humans. This review focuses on the nature of this trait and its nurturing during ontogeny and extending into adulthood. I begin by outlining the underpinning concepts of duet codes and pair-specific answering rules as used by birds to develop their learned coordinated song, driven by a complex interaction between self-generated and socially mediated auditory feedback. The more tractable avian model of duetting helps identify research gaps in singing primates that also use duetting as a type of intraspecific vocal interaction. Nevertheless, it has become clear that primate coordinated song—whether overlapping or antiphonal—is subject to some degree of vocal flexibility. This is reflected in the ability of lesser apes, titi monkeys, tarsiers, and lemurs to adjust the structure and timing of their calls through (1) social influence, (2) coordinated duetting both before and after mating, (3) the repair of vocal mistakes, (4) the production of heterosexual song early in life, (5) vocal accommodation in call rhythm, (6) conditioning, and (7) innovation. Furthermore, experimental work on the neural underpinnings of avian and mammalian antiphonal duets point to a hierarchical (cortico-subcortical) control mechanism that regulates, via inhibition, the temporal segregation of rapid vocal exchanges. I discuss some weaknesses in this growing field of research and highlight prospective avenues for future investigation.

Keywords: antiphonal, brain-to-brain coupling, development, duet code, singing primates, songbirds, vocal flexibility

Edited by:
Andrew James Jonathan MacIntosh,
Kyoto University, Japan

Reviewed by:
Allison Lau,
University of California, Davis, United States
Marco Gamba,
University of Turin, Italy

****Correspondence:***
Patrice Adret
patrice.adret@gmail.com

Specialty section:
This article was submitted to Behavioral and Evolutionary Ecology, a section of the journal Frontiers in Ecology and Evolution

Received: *25 January 2022*
Accepted: *28 February 2022*
Published: *25 March 2022*

Citation:
Adret P (2022) Developmental Plasticity in Primate Coordinated Song: Parallels and Divergences With Duetting Songbirds.
Front. Ecol. Evol. 10:862196.
doi: 10.3389/fevo.2022.862196

INTRODUCTION

"The development of communication is fundamentally embedded in social interactions across individual brains (Hasson et al., 2012)." Duetting, the coordinated sequences of acoustic signals exchanged between two individuals, has emerged as a remarkable phenotype of two brains wired to either cooperate or mitigate conflict (Fortune et al., 2011; Hoffmann et al., 2019; Okobi et al., 2019; Coleman et al., 2021). Whether this is a matter of hard or soft wiring remains an open question, but the diversity of mammalian and avian song duets holds great research promise for exploring how dyadic vocal interactions are shaped during ontogeny.

Here, I review the evidence for developmental plasticity in singing non-human primates[1], highlighting parallels and divergences with research on duetting songbirds. Collectively, these two phyla encompass tropical species that share similar socio-ecological characteristics, including putative sexual monogamy, family-living, and year-round territoriality with robust arboreal adaptations (Tobias et al., 2016; De Gregorio et al., 2022). However, they also differ in one key aspect, namely "vocal production learning," which is the ability to produce novel sounds from auditory experience (Janik and Slater, 2000; Vernes et al., 2021). While oscine songbirds (passerines) stand out as fine vocal learners, evidence of this is limited in non-human primates [Snowdon, 2017a; Janik and Knörnschild, 2021; but see Lameira (2017) who makes a strong case of vocal production learning in the voiceless calls of great apes].

THE NUTS AND BOLTS OF SONGBIRD DUETTING

The considerable progress in research on avian duetting is marked by several influential reviews (Farabaugh, 1982; Hall, 2009; Dahlin and Benedict, 2014). Duetting patterns in songbirds range from loosely coordinated song (Benedict and McEntee, 2009; Tobias and Seddon, 2009) to synchronized or antiphonal song[2] uttered with exquisite temporal precision (Wickler and Seibt, 1980; Templeton et al., 2013; Kovach et al., 2014) and combining alternation and synchrony (Mann et al., 2006).

Duet Codes and Answering Rules

Duetting behavior occurs at both the individual and pair levels (Levin, 1996), while Logue (2006) studied duetting from an operational perspective in which two individuals establish a shared set of rules. This led to the notion of a "duet code"—a set of answering rules one individual uses to answer its mate's song (Logue, 2006; Logue et al., 2008). While a duet is a pair-level property, a duet code is an individual attribute, and answers according to a duet code "adhere" to that code (Logue and Krupp, 2016). At its simplest, a single pairing rule, such as "answer F1 to M1,"[3] generates the cyclical duet [i-n(M1-F1)][4] produced by many songbirds (Levin, 1996; Rogers, 2005). A more complex duet code, such as "answer F1 to M1, F2 to M2, and F3 to M3," generates a non-repeated duet [i-(M1-F1-M2-F2-M3-F3)], as produced by an African weaver bird endowed with such a large syllable repertoire that both partners constantly switch between syllable types (Voigt et al., 2006; Lemazina et al., 2021). Logue's duet code concept opened up new avenues for measuring how code complexity and adherence vary across species (Logue and Krupp, 2016), whether duet codes are pair-specific (Mennill and Vehrencamp, 2005; Templeton et al., 2013), whether one sex or both adhere to these codes (Mann et al., 2003; Rivera-Cáceres, 2015), and whether duet codes emerge spontaneously in newly formed adult pairs or require vocal practice (Levin, 1996; Rivera-Cáceres et al., 2016). This begs the question: do young birds learn duet codes from their elders?

Duet Code Learning

Evidence that duet codes are learned from adults comes from observations of juveniles singing alongside their parents (Farabaugh, 1982; Hall, 2009). Such collective singing presumably allows juveniles to gain duetting experience, which not only requires learning what to answer and when but also mastering the duet rhythm in coordination with breathing given the rapid alternation (2–5 Hz) of male and female syllables (Hoffmann et al., 2019; Coleman et al., 2021). For example, song coordination in juvenile canebrake wrens improves over time via parental influence and independently of maturational effects, indicating a learning process (Rivera-Cáceres et al., 2018). Whether song acquisition results from copying a same-sex parent or integrating auditory information from both parental "tutors" remains unknown. There may also be alternative modes of code development with age. For example, a code might be retained throughout life ("close-ended"), whereby phrase-pairing rules remain constant regardless of partner identity (Levin, 1996); alternatively, mature individuals might re-learn a code each time they acquire a new mate ("open-ended"; Wickler, 1980). In the case of canebrake wrens, different pairs have distinct duetting rules, suggesting that learning in adulthood is likely. Indeed, removing and translocating individuals of well-established pairs confirmed that adult wrens re-learn pair-specific duet codes after re-mating, with males showing more flexibility in phrase-pairing rules than females (Rivera-Cáceres et al., 2016). Consequently, Rivera-Cáceres et al. (2018) proposed a three-step model for duet learning: (1) memorizing song material from auditory exposure, (2) rehearsing duet songs with both parents, and (3) re-learning to coordinate songs with a breeding partner (**Figure 1A**). Whether these two latter forms of sensorimotor learning share the same neural connections is the subject of future research (Nieder and Mooney, 2019).

Many songbirds co-sing in rapid turn on a syllable-to-syllable basis with sub-second latencies (Mann et al., 2009; Fortune et al., 2011; Rivera-Cáceres, 2015). To achieve such tight coordination, individuals rely on sensory information originating from two sources of auditory feedback—one generated by the bird's own voice (autogenous) and the other from its singing partner (heterogeneous). Owing to the velocity of sound, the longer the distance between the duetters the longer the delay for both receivers. Duetting songbirds adapt to these delays by altering the timing of their singing (Fortune et al., 2011) or by using visual cues in open habitats (Rek and Magrath, 2016, 2020). How, then, is auditory feedback encoded in the brain?

Neural Mechanisms

Neuroanatomical studies of duetting songbirds reveal the presence of well-developed brain nuclei dedicated to song production learning in both sexes, which contrasts with the females of species in which only males sing

---

[1] Singing primates are distributed in Southeast Asia (e.g., gibbons, tarsiers, and the Mentawai langur), Madagascar (e.g., indri and Milne Edwards' sportive lemurs), and South America (e.g., titi monkeys).

[2] A series of notes of different types, uttered following a hierarchical structure, and characterized by a frequency variation.

[3] Where F1 and M1 stand for female and male syllable types, respectively.

[4] Where "i" stands for the introductory notes, with $n > 1$.

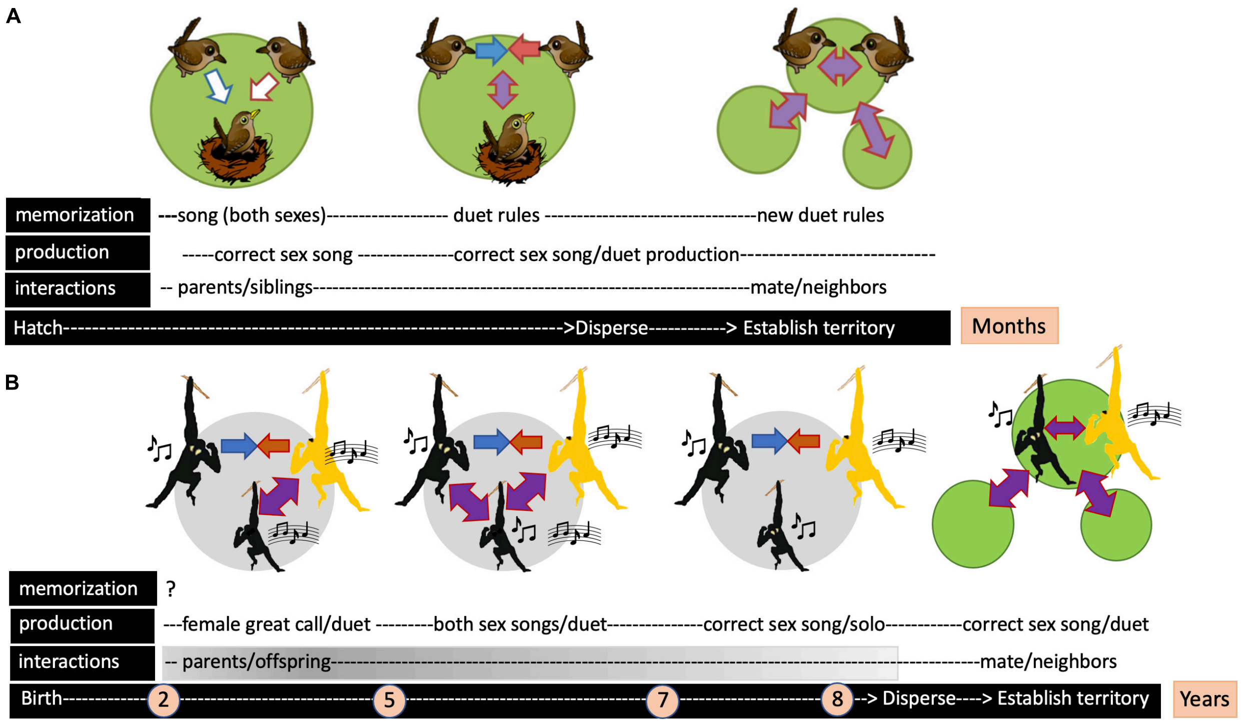


FIGURE 1 | Diagram illustrating the timing of song acquisition and the nature of social interactions that characterize the development of antiphonal duets in a songbird (canebrake wren) and a lesser ape (southern yellow-cheeked gibbon). **(A)** The songbird model of duet acquisition in which young birds first memorize sounds heard from both parents acting as tutors, then "learn" duet codes during mutual singing sessions with a same-sex parent; following post-natal dispersion, mature individuals "re-learn" to combine song elements with a new mate, performing a duet which is used to advertise territory ownership and/or pair bond strength. **(B)** The lesser ape model of duet acquisition in which a young male first develops a female-like great call while co-singing with his mother until reaching sexual maturity (3–5 years). Sexually mature daughters – not shown in the diagram – have acquired the basic pattern of the maternal song (Merker and Cox, 1999), which is then perfected during co-singing sessions with the mother until leaving the parental group (Koda et al., 2013). Mother-son vocal interactions continue at a decreasing rate (gray gradient) until adolescence. From 5 to 7 years, sons utter both male and female song elements and subsequently discard the female-like great call from their repertoire, retaining only male song (coda). The male coda consists of a multi-modulation note and a staccato note that develop in that sequence until at least 8 years of age (Hradec et al., 2021). In the absence of experimental evidence, gibbons are not considered vocal learners, but the memorization phase remains questionable. Arrows denote vocal interactions. Note the difference in the timeline between the two model systems. Green circles refer to observations made in the wild; gray circles depict events observed in zoo animals [adapted from Rivera-Cáceres and Templeton (2019), Hradec et al. (2021)].

(Nottebohm and Arnold, 1976; Brenowitz and Arnold, 1986; Deng et al., 2001; but see Lobato et al., 2015). Research into the neural underpinnings of antiphonal duetting targets the HVC (used as a proper name), a high-order forebrain song nucleus involved in sensorimotor learning (Nieder and Mooney, 2019). Contrary to neurophysiological data obtained for songbirds in which only males sing, extracellular recordings in the HVC of anesthetized wrens show strong responses to auditory presentations of both male and female song when played in isolation (Fortune et al., 2011). Furthermore, experimental manipulation of song stimuli shows a sensitivity of HVC neurons to inter-syllable intervals. Importantly, the response strength of HVC neurons to duet stimuli exceeds the sum of neuronal responses to each individual's song. This suggests that each participant not only knows what to sing but also develops an internal representation of the pair-specific duet (Fortune et al., 2011). Groundbreaking work in free-ranging African weavers further demonstrates the alternation of neuronal activity in each partner's HVC, with bursts temporally locked to syllable onsets (Hoffmann et al., 2019). This "on-off" pattern appears to be regulated by heterogeneous auditory feedback that reciprocally inhibits HVC premotor activity (Coleman et al., 2021). Such brain-to-brain coupling mechanisms ensure precise timing of dyadic vocal interactions, most likely through gamma-aminobutyric acid-ergic inhibition (Benichov and Vallentin, 2020). For comprehensive reviews on this topic, see Elie et al. (2019) and Rivera-Cáceres and Templeton (2019).

DUETTING STYLES IN SINGING PRIMATES

Worldwide, singing primates comprise 72 species, some of which are nocturnal and others diurnal; most share a family-living and territorial social system mediated by loud, coordinated calls emitted at predictable times, usually around dawn and/or dusk (De Gregorio et al., 2022). The gibbons' "great-call sequence" combines the female great call and male coda, often repeated alternately [i-n(F1-M1)], with a pronounced sexual divocalism (Marshall and Marshall, 1976; Geissmann, 2002). Sexually dimorphic species duet antiphonally, whereas in monomorphic taxa, singers tend to overlap (Deputte, 1982). The duet songs of lemurs, tarsiers, and the Mentawai langur overlap, except in *Lepilemur edwardsi* (Méndez-Cárdenas and Zimmermann, 2009) and *Tarsius niemitzi* (Shekelle et al., 2019). Sexually monomorphic indris advertise with duets and choruses[5] in which the paired males and females overlap more than any other dyad while dominant and non-dominant individuals avoid overlapping (Gamba et al., 2016). In each of these lineages, sex-differentiated calls often occupy a different frequency register, making them readily distinguishable on spectrograms (Tilson and Tenaza, 1976; Nietsch, 1999; Torti et al., 2013). In contrast, Neotropical titi monkey duets overlap extensively, both in the time and frequency domains, with male and female contributions exhibiting an anti-phase-locked pattern of phrase coordination devoid of discrete turns (Robinson, 1979; Müller and Anzenberger, 2002; Caselli et al., 2014; Adret et al., 2018a; Clink et al., 2019, 2022). In each of these primate lineages, there is increasing evidence of vocal malleability for this trait, long thought to be subject to strong genetic constraints (Brockelman and Schilling, 1984; Tenaza, 1985; Hammerschmidt and Fischer, 2008).

FLEXIBILITY IN THE COORDINATED SONG OF SINGING PRIMATES

Vocal flexibility, the capacity for modifying vocalizations according to context, can affect call structure, amplitude, timing, duration, and rhythm. For duetting animals, this includes individuals adjusting their singing to either their partner's or neighbors' vocal outputs.

Interactive Group Singing

Neighboring groups of singing primates often call antiphonally (Kinzey et al., 1977; Marler and Tenaza, 1977; Raemaekers and Raemaekers, 1985) and counter-sung solos and duets are longer than solos and duets sung alone (Tenaza, 1976; Mitani, 1985). In support of the flexible timing of vocal output, active counter-singing and singing motivation have been experimentally corroborated (Chivers and MacKinnon, 1977; Mitani, 1988; Dooley and Judge, 2007). Studies of communication networks showing that siamangs are sensitive to their neighbors' group disruption (Morino et al., 2021) are likely to unveil further instances of vocal flexibility in the future.

Within-Pair Vocal Coordination and Repair

Individual gibbons flexibly time their contributions relative to their mates' during the great-call sequence. Guided by subtle changes in female introductory notes that signal an impending great call, the male suspends phonation; cued by her post-climax descending notes, he resumes singing with a coda phrase according to a precise turn-taking pattern (Terleph et al., 2018a). Flexibility is needed given individual variability in the female great call (Terleph et al., 2015, 2016). Adjustment made by hylobatids in response to a mate's vocal "mistakes" are termed "*repairs*," a universal principle of human conversation (Schegloff et al., 1977; Dingemanse et al., 2015). Repairs have been scrutinized for self-corrected, stalled, and aborted great calls (Haimoff, 1988; Haraway and Maples, 1998; Terleph et al., 2018a). Such studies confirm the existence of duet codes and answering rules in lesser apes. Non-adherence to the duet code (e.g., production of atypical notes or unexpected call timing) may result in duet interruption and song reset by the mate.

Vocal Accommodation in Call Rhythm

Coordinated singing and rhythm dynamics are not necessarily tied (Ravignani et al., 2014). For example, inter-onset call intervals extracted at each level of the indri's song organization (i.e., units and phrases), reveal music-like categorical rhythmicity (De Gregorio et al., 2021a). Both in adults and young individuals, females exhibit more flexibility than males, with a sensitivity to chorus size (Gamba et al., 2016; De Gregorio et al., 2019, 2021b). Sex-related "divergence" in indri song rhythm contrasts with titis and tarsiers. In a cross-sectional study of duetting pairs of titi monkeys, partners were found to adjust pulse rate and phrase duration to one another, showing call "convergence" (Clink et al., 2019). A longitudinal study with newly formed pairs of titis might establish whether vocal learning is involved through convergence in the spectral features of calls, as reported in marmosets (Elowson and Snowdon, 1994; Snowdon and Elowson, 1999; Zürcher et al., 2021). Likewise, male and female tarsiers flexibly adjust call rhythm relative to their partner through simultaneous accelerations and decelerations (Clink et al., 2020). Within-pair convergence in duet tempo might be achieved by entrainment, i.e., spontaneous responsiveness to a perceived rhythmic signal (de Reus et al., 2021).

Parental Influence

Immature individuals singing jointly with their elders have long sparked research attention (Deputte, 1982; Raemaekers et al., 1984; Pollock, 1986; Reichard, 2003). A longitudinal study of mother-daughter vocal interactions in gibbons revealed the acquisition of correct note sequencing over time (5–30 months; Merker and Cox, 1999). In a cross-sectional study of free-ranging family groups, an inverse relationship was found between mother-daughter co-singing rates and call synchronization; less proficient daughters co-sang at higher rates. Interestingly, mothers adjusted their song to a more stereotyped pattern when co-singing than when singing alone, suggesting a "teaching role" of mothers (Koda et al., 2013). While sexually mature females sing an adult-like maternal song (Brockelman and Schilling, 1984; Merker and Cox, 1999; Koda et al., 2013), males master the multi-part coda phrase years later (Hradec et al., 2021) via an intriguing developmental trajectory (**Figure 1B**).

Production of Heterosexual Song

Spontaneous production of a female-like great call by immature males has been reported in several gibbon species (Koda et al., 2014; Hradec et al., 2016, 2017, 2021). A triggering role of the maternal call in young males, possibly associated with low

---

[5] Coordinated song uttered by more than two individuals within a family group.

androgen levels, has also been invoked (Koda et al., 2014). Immature individuals producing male calls potentially face aggression from the father (Hradec et al., 2021) and there is evidence that the stress hormone cortisol may negatively interact with testosterone in influencing the expression of secondary sexual traits (Puts et al., 2016). Close monitoring of hormone levels would be worthwhile in order to determine the impact of parent–offspring relationships on gibbon song development (Burns and Judge, 2016).

Acquisition of a Pair-Specific Duet Code

To reproduce outside their natal groups, mature individuals must coordinate their song with a prospective mate "having both different genetic parentage and a different history of developmental experience than their own" (Haraway and Maples, 1998). In indris, spectral-temporal features of descending phrases correlate with genetic distance in males, whereas females are less constrained (Torti et al., 2017). Thus, indri choruses may inform conspecifics about individuals' genetic relatedness. Such an effect is less apparent in titi duets (Clink et al., 2022). Consistent with vocal flexibility and duet code learning, the duets of long-term mates are better coordinated than those of newly formed pairs (Geissmann, 1986, 1999; Maples et al., 1989; Müller and Anzenberger, 2002).

Conditioning

Robust conditioned responses are obtained in lesser apes via reinforcement and extinction procedures in which song presentation is contingent upon an individual's own vocalization (Haraway et al., 1981; Maples and Haraway, 1982; Maples et al., 1988). Moreover, both in lemurs and gibbons, phonation can be brought under volitional control in response to an arbitrary visual signal (Wilson, 1975; Koda et al., 2007), thus demonstrating voluntary control over call timing.

Innovation

Captive siamangs can alter their calls using various "tricks," including the production of hand- modulated and echoing calls (Badraun et al., 1998). Geissmann (2009) observed one female gibbon who amplified her duet contribution by slamming the sliding door of her sleeping quarters at the climax of her great call.

Causal Mechanisms

As renowned "soprano singers" (Koda et al., 2012), gibbons produce pure-tonal melodious song that requires appropriate hormonal and neural machinery for pitch control. Contrasting with humans, however, higher androgen levels result in calls with a higher pitch (Barelli et al., 2013; Puts et al., 2016). Experiments in a helium-oxygen atmosphere revealed that the unshifted call fundamental frequency is strongly attenuated and the first harmonic is emphasized, suggesting that the sound source (larynx) operates independently of the supralaryngeal vocal tract (Koda et al., 2012). Thus, call flexibility can be achieved by controlling laryngeal function and/or the resonance filter configuration (Gamba et al., 2011, 2017; Fitch et al., 2016), but the challenge is to account for the larynx development (Zhang et al., 2020). Importantly, bipolar excitation in the inferior portion of the precentral gyrus in the left hemisphere yields adduction of the vocal folds (Mott et al., 1911). This suggests that, in the gibbon brain, a direct pathway exists from the laryngeal representation in the primary motor cortex to the laryngeal motoneurons of the nucleus ambiguus, which controls the muscles of the larynx for vocal production (Simonyan, 2014). This might explain why gibbons can be trained to call on command (Koda et al., 2007; but see Hage and Nieder, 2013).

DISCUSSION AND FUTURE DIRECTIONS

From strepsirrhines to lesser apes, the duetting patterns of singing primates provide compelling evidence of developmental plasticity extending into adulthood. This is consistent with the view that non-human primates exhibit more flexibility in their vocal behavior than is generally acknowledged

TABLE 1 | Parallels and divergences in vocal plasticity between duetting songbirds and singing primates.

	Acronym	Duetting songbirds	Singing primates
Parallels	COS	yes	yes
	CTS	yes	yes
	CTXL	yes	yes
	HET	yes	yes
	NFP vs. WEP	yes	yes
	REP	yes	yes
Divergences	CONV	?	yes
	CDT	?	yes
	INN	?	yes
	VPL	yes	?
	MEM	yes	?
Strengths and weaknesses	TDA	months	years
	ONT	weak	strong
	RIP	strong	absent
	NEULAB	strong	weak
	NEUTEL	strong	absent
	VOCTEL	strong	absent

Strengths and weaknesses identify several methodological approaches for which songbirds have proved to be more tractable experimentally. Note that, despite a protracted developmental period for duet acquisition, intensive studies have been carried out on the ontogeny of coordinated song in singing primates. CDT, conditioning; CONV, vocal convergence; COS, parent-offspring co-singing; CTS, inter-group counter-singing; CTXL, contextual learning; HET, production of heterosexual song in sexually dimorphic species; INN, vocal innovation; MEM, song memorization; NEULAB, neural investigations in captive animals; NEUTEL, neural telemetry in freely ranging animals, which is achieved, for instance, by trapping songbirds and implanting electrodes in a target brain nucleus to obtain chronic recordings of the neural activity via telemetry once the bird is released into the wild; NFP vs. WEP, newly formed pairs vs. well-established pairs; ONT, ontogeny of vocal duetting; REP, vocal repair; RIP, removal experiment and interactive playback in which one pair member is first captured; subsequently, the experimenter tries to elicit a duet with the lone, territorial individual by playing back his/her mate's song contribution (unaltered or manipulated); TDA, timeline for duet acquisition; VPL, vocal production learning; VOCTEL, vocal telemetry in freely ranging animals equipped with a lightweight, backpacked miniature microphone (songbirds) or with the sensor fitted to the subject's throat, in close apposition with the larynx (primates).

(Snowdon, 2009, 2017a,b, 2018). Promising areas of ongoing research include (1) vocal convergence as a learning process, linked to pair-bond strength (Clink et al., 2019, 2020), (2) sex-dependent mechanisms regulating "acquisition" of categorical duet rhythms (De Gregorio et al., 2021a), and (3) the potential for parental tutoring and vocal production learning in gibbons (Koda et al., 2013; Koda, 2016; Terleph et al., 2018b; Hradec et al., 2021).

Striking similarities have emerged in duet acquisition between songbirds and singing primates (**Table 1**). In both phyla, young individuals co-sing extensively with their elders, although timescales can widely differ (**Figure 1**). Furthermore, in species with sex-specific repertoires, males and females can produce heterosexual song (Geissmann, 1983; Chen et al., 2008; Rivera-Cáceres and Templeton, 2019; Hradec et al., 2021). The production of heterosexual song early in life suggests a pre-existing or learned auditory template (Adret, 2004; Cheyne et al., 2007), possibly engaging a mirror-neuron system (Newman, 2014).

Research currently tends to focus on antiphonal duets, given their potential as precursors of turn-taking conversations in humans (Levinson, 2016). At the same time, bio-acoustics research in titi monkey duets has been hampered by extensive call overlap (Caselli et al., 2014; Adret et al., 2018a; Clink et al., 2019); cracking the code will require radio-tracking calls with miniature voice detectors (Adret et al., 2018b), as has been elegantly demonstrated in songbirds (Hoffmann et al., 2019; Lemazina et al., 2021). Another solution is conducting studies in captive (or wild) populations for which high speed video of vocalizing animals can be paired with high quality audio to ensure caller identity (Haimoff, 1981). Performant computational methods also allow effective clustering of acoustic signatures at multiple levels within animal vocal repertoires (Sainburg et al., 2020). A machine-learning approach to acoustic stream segregation might further help resolve the "cocktail party problem" (Elhilali, 2017). Developmental studies of duet acquisition in singing primates are also needed to investigate vocal flexibility in response to anthropogenic noise (Duarte et al., 2017).

While the neuroscience of pair-bonding in socially monogamous mammals is well documented (Bales et al., 2017; Potretzke and Ryabinin, 2019), a significant gap in knowledge concerns the neural mechanisms of duetting in singing primates. Integrating respiratory functions associated with coordinated song is also necessary to account for the generation of rhythmic patterns (Laje and Mindlin, 2003; Amador et al., 2005). Neuroimaging studies provide a powerful, non-invasive approach to mapping brain areas activated by antiphonal calling (Takahashi et al., 2021). Singing rodents, which offer a genetically tractable model system, produce antiphonal duets, which, much like duetting songbirds, reveals a hierarchical (cortico-subcortical) control mechanism that regulates the temporal segregation of rapid vocal exchanges via inhibition (Okobi et al., 2019). Emergence, deep in the evolutionary past, of an interlocking mechanism derived from sender-listener brain coupling (Hasson et al., 2012) may have been a key step in the evolution of human conversation.

AUTHOR CONTRIBUTIONS

PA conceived and wrote the article and approved the final version of the manuscript.

ACKNOWLEDGMENTS

I would like to thank both reviewers for insightful comments and Editage (www.editage.com) for English language editing.

REFERENCES

Adret, P. (2004). In search of the song template. *N. Y. Acad. Sci.* 1016, 303–324. doi: 10.1196/annals.1298.005

Adret, P., Dingess, K. A., Caselli, C. B., Vermeer, J., Martínez, J., Luna Amancio, J. C., et al. (2018a). Duetting patterns of titi monkeys (Primates: Callicebinae) and relationships with phylogeny. *Animals* 8, 178–211. doi: 10.3390/ani8100178

Adret, P., Cochran, J. S., and Suarez Rado, M. (2018b). Airborne vs. radio-transmitted vocalizations in two primates: a technical report. *Bioacoustics* 27, 273–294. doi: 10.1080/09524622.2017.1335617

Amador, A., Trevisan, M. A., and Mindlin, G. B. (2005). Simple neural substrate predicts complex rhythmic structure in duetting birds. *Phys. Rev. E* 72:031905. doi: 10.1103/PhysRevE.72.031905

Badraun, J. C., Mootnick, A. R., Deaner, R. O., Agoramoorthy, G., and McNeese, K. M. (1998). Hand modulation of vocalization in Siamangs, *Hylobates syndactylus*. *lnt. Zoo Yb* 36, 84–89. doi: 10.1111/j.1748-1090.1998.tb02888.x

Bales, K. L., Arias, del Razoa, R., Conklina, Q. A., Hartman, S., Mayera, H. S., et al. (2017). Titi monkeys as a novel non-human primate model for the neurobiology of pair bonding. *Yale J. Biol. Med.* 90, 373–387.

Barelli, C., Mundry, R., Heistermann, M., and Hammerschmidt, K. (2013). Cues to androgens and quality in male gibbon songs. *PLoS One* 8:e82748. doi: 10.1371/journal.pone.0082748

Benedict, L., and McEntee, J. P. (2009). Context, structural variability and distinctiveness of California towhee (*Pipilo crissalis*) vocal duets. *Ethology* 115, 77–86. doi: 10.1111/j.1439-0310.2008.01583.x

Benichov, J. I., and Vallentin, D. (2020). Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nat. Commun.* 11:221. doi: 10.1038/s41467-019-13938-0

Brenowitz, E. A., and Arnold, A. (1986). Interspecific comparisons of the size of neural song control regions and song complexity in duetting birds: evolutionary implications. *J. Neurosci.* 6, 2875–2879. doi: 10.1523/JNEUROSCI.06-10-02875.1986

Brockelman, W. Y., and Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature* 312:634. doi: 10.1038/312634a0

Burns, B. L., and Judge, D. S. (2016). The varied path to adulthood: plasticity in developmental timing in hylobatids. *Am. J. Primatol.* 78, 610–625. doi: 10.1002/ajp.22523

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., and Setz, E. Z. F. (2014). Vocal behavior of black-fronted titi monkeys (*Callicebus nigrifrons*): Acoustic properties and behavioral contexts of loud calls. *Am. J. Primatol.* 76, 788–800. doi: 10.1002/ajp.22270

Chen, H.-C., Kamolnorranath, S., and Kaplan, G. (2008). Female crested gibbons (*genus Nomascus*) sing male song. *Vietnamese J. Primatol.* 2, 47–53.

Cheyne, S., Chivers, D. J., and Sugardjito, J. (2007). Covariation in the great calls of rehabilitant and wild gibbons (*Hylobates albibarbis*). *Raffles Bull. Zool.* 55, 201–207.

Chivers, D. J., and MacKinnon, J. P. (1977). On the behavior of siamang after playback of their calls. *Primates* 18, 943–948. doi: 10.1007/BF02382944

Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour* 156, 1–24. doi: 10.1163/1568539X-00003575

Clink, D. J., Lau, A. R., Kanthaswamy, S., Johnson, L. M., and Bales, K. L. (2022). Moderate evidence for heritability in the duet contributions of a South American primate. *J. Evol. Biol.* 35, 51–63. doi: 10.1111/jeb.13962

Clink, D. J., Tasirin, J. S., and Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Curr. Zool.* 66, 173–186. doi: 10.1093/cz/zoz035

Coleman, M. J., Day, N. F., Rivera-Parrac, P., and Fortune, E. S. (2021). Neurophysiological coordination of duet singing. *PNAS* 118:e2018188118. doi: 10.1073/pnas.2018188118

Dahlin, C. R., and Benedict, L. (2014). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 120, 1–10. doi: 10.1111/eth.12182

De Gregorio, C., Daria Valente, Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., et al. (2021a). Categorical rhythms in a singing primate. *Curr. Biol.* 31, R1379–R1380. doi: 10.1016/j.cub.2021.09.032

De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., et al. (2021b). Born to sing! Song development in a singing primate. *Curr. Zool.* 67, 585–596. doi: 10.1093/cz/zoab018

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* doi: 10.1080/03949370.2021.2015451

De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., et al. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Curr. Zool.* 65, 89–97. doi: 10.1093/cz/zoy058

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., et al. (2021). Rhythm in dyadic interactions. *Phil. Trans. R. Soc. B* 376:20200337. doi: 10.1098/rstb.2020.0337

Deng, C., Kaplan, G., and Rogers, L. J. (2001). Similarity of the song nuclei of male and female Australian magpies (*Gymnorhina tibicen*). *Behav. Brain Res.* 123, 89–102. doi: 10.1016/s0166-4328(01)00200-5

Deputte, B. (1982). "Duetting in male and female songs of the white-cheeked gibbon (*Hylobates concolor* leucogenys)," in *Primate Communication*, eds C. T. Snowdon, C. H. Brown, and M. R. Petersen (Cambridge: Cambridge University Press), 67–93.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal principles in the repair of communication problems. *PLoS One* 10:e0136100. doi: 10.1371/journal.pone.0136100

Dooley, H., and Judge, D. (2007). Vocal responses of captive gibbon groups to a mate change in a pair of white-cheeked gibbons (*Nomascus leucogenys*). *Folia Primatol.* 78, 228–239. doi: 10.1159/000102318

Duarte, M., Kaizer, M., Young, R., Rodrigues, M., and Sousa-Lima, R. (2017). Mining noise affects loud call structures and emission patterns of wild black-fronted titi monkeys. *Primates* 59, 89–97. doi: 10.1007/s10329-017-0629-624

Elhilali, M. (2017). "Modelling the cocktail party problem," in *The Auditory System at the Cocktail Party*, eds J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Berlin: Springer Handbook of Auditory Research), doi: 10.1007/978-3-319-51662-2_5

Elie, J. E., Hoffmann, S., Dunning, J. L., Coleman, M. J., Fortune, E. S., and Prather, J. F. (2019). From perception to action: the role of auditory input in shaping vocal communication and social behaviors in birds. *Brain Behav. Evol.* 94, 51–60. doi: 10.1159/000504380

Elowson, A. M., and Snowdon, C. T. (1994). Pygmy marmosets, Cebuella pygmaea, modify vocal structure in response to changed social environment. *Anim. Behav.* 47, 1267–1277. doi: 10.1006/anbe.1994.1175

Farabaugh, S. M. (1982). "The ecological and social significance of duetting," in *Acoustic Communication in Birds*, eds D. E. Kroodsma and E. H. Miller (New York, NY: Academic press), 85–124.

Fitch, W. T., de Boer, B., Mathur, N., and Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Sci. Adv.* 2:e1600723. doi: 10.1126/sciadv.1600723

Fortune, E. S., Rodríguez, C., Li, D., Ball, G. F., and Coleman, M. J. (2011). Neural Mechanisms for the coordination of duet singing in wrens. *Science* 334, 666–670. doi: 10.1126/science.1209867

Gamba, M., Favaro, L., Araldi, A., Matteucci, V., Giacoma, C., and Friard, O. (2017). Modeling individual vocal differences in group-living lemurs using vocal tract morphology. *Curr. Zool.* 63, 467–475. doi: 10.1093/cz/zox023

Gamba, M., Favaro, L., Torti, V., Sorrentino, V., and Giacoma, C. (2011). Vocal tract flexibility and variation in the vocal output in wild indris. *Bioacoustics* 20, 251–265. doi: 10.1080/09524622.2011.9753649

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., et al. (2016). The Indris have got rhythm! timing and pitch variation of a primate song examined between sexes and age classes. *Front. Neurosci.* 10:249. doi: 10.3389/fnins.2016.00249

Geissmann, T. (1983). Female capped gibbon (*Hylobates pileatus* Gray 1891) sings male song. *J. Human Evol.* 12, 667–671. doi: 10.1016/S0047-2484(83)80006-2

Geissmann, T. (1986). Mate change enhances duetting activity in the Siamang gibbon (*Hylobates syndactylus*). *Behaviour* 96, 17–27. doi: 10.1163/156853986x00199

Geissmann, T. (1999). Duet songs of the siamang, *Hylobates syndactylus*: ii. testing the pair-bonding hypothesis during a partner exchange. *Behaviour* 136, 1005–1039. doi: 10.1163/156853999501694

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/s1464793101005826

Geissmann, T. (2009). Door slamming: tool-use by a captive white-handed gibbon (*Hylobates lar*). *Gibbon J.* 5, 53–60. doi: 10.5167/UZH-20113

Hage, S. R., and Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Com.* 4:2409. doi: 10.1038/ncomms3409

Haimoff, E. H. (1981). Video analysis of siamang (*Hylobates syndactylus*) songs. *Behaviour* 76, 128–151. doi: 10.2307/4534093

Haimoff, E. H. (1988). The organization of repair in the songs of gibbons. *Semiotica* 68, 89–120. doi: 10.1515/semi.1988.68.1-2.89

Hall, M. L. (2009). A review of vocal duetting in birds. *Adv. Stud. Behav.* 40, 67–121. doi: 10.1016/S0065-3454(09)40003-2

Hammerschmidt, K., and Fischer, J. (2008). "Constraints in primate vocal production," in *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability*, eds D. K. Oller and U. Griebel (Cambridge, MA: MIT Press), 93–120. doi: 10.7551/mitpress/9780262151214.003.0005

Haraway, M. M., and Maples, E. G. (1998). Flexibility in the species-typical songs of gibbons. *Primates* 39, 1–12. doi: 10.1007/BF02557739

Haraway, M. M., Maples, E. G., and Tolson, S. (1981). Taped vocalization as a reinforcer of vocal behavior in a Siamang gibbon (*Symphalangus syndactylus*). *Psychol. Rep.* 49, 995–999. doi: 10.2466/pr0.1981.49.3.995

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-Brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

Hoffmann, S., Trost, L., Voigt, C., Leitner, S., Lemazina, A., Sagunsky, H., et al. (2019). Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. *Nat. Com.* 10:2577. doi: 10.1038/s41467-019-10593-3

Hradec, M., Bolechová, P., and Svobodova, I. (2016). Production of a female-specific great call in an immature male gibbon, the *Nomascus genus*. *Primates* 57, 445–448. doi: 10.1007/s10329-016-0569-564

Hradec, M., Illmann, G., Bartoš, L., and Bolechová, P. (2021). The transition from the female–like great calls to male calls during ontogeny in southern yellow–cheeked gibbon males (*Nomascus gabriellae*). *Sci. Rep.* 11:22040. doi: 10.1038/s41598-021-01648-x

Hradec, M., Linhart, P., Bartoš, L., and Bolechová, P. (2017). The traits of the great calls in the juvenile and adolescent gibbon males *Nomascus gabriellae*. *PLoS One* 12:e0173959. doi: 10.1371/journal.pone.0173959

Janik, V. M., and Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Phil. Trans. R. Soc. B* 376:20200244. doi: 10.1098/rstb.2020.0244

Janik, V. M., and Slater, P. J. B. (2000). The different roles of social learning in communication. *Anim. Behav.* 60, 1–11. doi: 10.1006/anbe.2000.1410

Kinzey, W. G., Rosenberger, A. L., Heisler, P. S., Prowse, D. L., and Trilling, J. S. (1977). A preliminary field investigation of the Yellow handed titi monkey, *Callicebus torquatus* torquatus, in Northern Peru. *Primates* 18, 159–181. doi: 10.1007/bf02382957

Koda, H. (2016). "Gibbon songs: understanding the evolution and development of this unique form of vocal communication," in *Evolution of Gibbons and Siamang. Developments in Primatology: Progress and Prospects*, eds U. Reichard, H. Hirai, and C. Barelli (New York: Springer), 347–357. doi: 10.1007/978-1-4939-5614-2_15

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS One* 8:e71432. doi: 10.1371/journal.pone.0071432

Koda, H., Masataka, N., Kato, A., and Oyakawa, C. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour* 144, 681–692. doi: 10.1163/156853907781347817

Koda, H., Nishimura, T., Tokuda, I. T., Oyakawa, C., Nihonmatsu, T., and Masataka, N. (2012). Soprano singing in gibbons. *Am. J. Phys. Anthropol.* 149, 347–355. doi: 10.1002/ajpa.22124

Koda, H., Oyakawa, C., Kato, A., Shimizu, D., Rizaldi, Koyama, Y., et al. (2014). Immature male gibbons produce female-specific songs. *Primates* 55, 13–17. doi: 10.1007/s10329-013-0390-2

Kovach, C., Hall, M. L., Verhencamp, S. L., and Mennill, D. J. (2014). Timing isn't everything: responses of tropical wrens to coordinated duets, uncoordinated duets and alternating solos. *Anim. Behav.* 95, 101–109. doi: 10.1016/j.anbehav.2014.06.012

Laje, R., and Mindlin, G. B. (2003). Highly structured duets in the song of the South American Hornero. *Phys. Rev. Lett.* 91:258104. doi: 10.1103/PhysRevLett.91.258104

Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrate for speech evolution. *Neurosci. Biobehav. Rev.* 83, 429–439. doi: 10.1016/j.neubiorev.2017.09.021

Lemazina, A., Trost, L., Gahr, M., and Hoffmann, S. (2021). The multifaceted vocal duets of white-browed sparrow weavers are based on complex duetting rules. *J. Avian Biol.* 52:e02703. doi: 10.1111/jav.02703

Levin, R. N. (1996). Song behaviour and reproductive strategies in a duetting wren, *Thryothorus nigricapillus*: I. removal experiments. *Anim. Behav.* 52, 1093–1106.

Levinson, S. C. (2016). Turn-taking in human communication-origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Lobato, M., Vellema, M., Gahr, C., Leitão, A., de Lima, S. M. A., Geberzahn, N., et al. (2015). Mismatch in sexual dimorphism of developing song and song control system in blue-capped cordon-bleus, a songbird species with singing females and males. *Front. Ecol. Evol.* 3:117. doi: 10.3389/fevo.2015.00117

Logue, D. M. (2006). The duet code of the female black-bellied wren. *Condor* 108, 326–335. doi: 10.1093/condor/108.2.326

Logue, D. M., and Krupp, D. B. (2016). Duetting as a collective behavior. *Front. Ecol. Evol.* 4:7. doi: 10.3389/fevo.2016.00007

Logue, D. M., Chalmers, C., and Gowland, H. (2008). The behavioural mechanisms underlying temporal coordination in black-bellied wren duets. *Anim. Behav.* 75, 1803–1808. doi: 10.1016/j.anbehav.2007.10.036

Mann, N. I., Dingess, K. A., and Slater, P. J. B. (2006). Antiphonal four-part synchronized chorusing in a neotropical wren. *Biol. Lett.* 2, 1–4. doi: 10.1098/rsbl.2005.0373

Mann, N. I., Dingess, K. A., Barker, F. K., Graves, J. A., and Slater, P. J. B. (2009). A comparative study of song form and duetting in neotropical *Thryothorus wrens*. *Behaviour* 146, 1–43. doi: 10.1163/156853908X390913

Mann, N. I., Marshall-Ball, L., and Slater, P. J. B. (2003). The complex song duet of the plain wren. *Condor* 105, 672–682. doi: 10.1650/7208

Maples, E. G. Jr., and Haraway, M. M. (1982). Taped vocalization as a reinforcer of vocal behavior in a female agile gibbon (*Hylobates agilis*). *Psychol. Rep.* 493, 995–999. doi: 10.2466/pr0.1982.51.1.95

Maples, E. G. Jr., Haraway, M. M., and Hutto, C. W. (1989). Development of coordinated singing in a newly formed siamang pair (*Hylobates syndactylus*). *Zoo Biol.* 8, 367–378. doi: 10.1002/zoo.1430080407

Maples, E. G., Haraway, M. M., and Collie, L. (1988). Interactive singing of a male Mueller's gibbon with a simulated neighbor. *Zoo Biol.* 7, 115–122. doi: 10.1002/zoo.1430070204

Marler, P., and Tenaza. (1977). "Signaling behavior of apes with special reference to vocalization," in *How Animals Communicate*, ed. T. A. Sebeok (Bloomington: Indiana University Press).

Marshall, J., and Marshall, E. (1976). Gibbons and their territorial songs. *Science* 193, 235–237. doi: 10.1126/science.193.4249.235

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting—a mechanism to strengthen pair bonds in a dispersed pair-living primate (Lepilemur edwardsi)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Mennill, D. J., and Vehrencamp, S. L. (2005). Sex differences in singing and duetting behavior of neotropical rufous-and-white wrens (*Thryothorus rufalbus*). *Auk* 122, 175–186. doi: 10.1093/auk/122.1.175

Merker, B., and Cox, C. (1999). Development of the female great call in Hylobates gabriellae: a case study. *Folia Primatol.* 70, 97–106. doi: 10.1159/000021680

Mitani, J. C. (1985). Gibbon song duets and intergroup spacing. *Behaviour* 92, 59–96. doi: 10.1163/156853985X00389

Mitani, J. C. (1988). Male gibbon (*Hylobates agilis*) singing behavior: natural history, song variations, and function. *Ethology* 79, 177–194. doi: 10.1111/j.1439-0310.1988.tb00710.x

Morino, L., Pasquaretta, C., Sueur, C., and MacIntosh, A. J. J. (2021). Communication network reflects social instability in a wild siamang (*Symphalangus syndactylus*) population. *Int. J. Primatol.* 42, 618–639. doi: 10.1007/s10764-021-00227-1

Mott, F. W., Schuster, E., and Sherrington, C. S. (1911). Motor localisation in the brain of the gibbon, correlated with a histological examination. *Proc. Roy. Soc. B London* 84, 67–74. doi: 10.1098/rspb.1911.0047

Müller, A. E., and Anzenberger, G. (2002). Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. doi: 10.1159/000064788

Newman, J. D. (2014). Vocal coordination and vocal imitation: a role for mirror neurons? *Behav. Brain Sci.* 37, 211–212. doi: 10.1017/S0140525X13002410

Nieder, A., and Mooney, R. (2019). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Phil. Trans. R. Soc. B* 375:20190054. doi: 10.1098/rstb.2019.0054

Nietsch, A. (1999). Duet vocalizations among different populations of *Sulawesi tarsiers*. *Int. J. Primatol.* 20, 567–583. doi: 10.1023/A:1020342807709

Nottebohm, F., and Arnold, A. P. (1976). Sexual dimorphism in vocal control areas of the songbird brain. *Science* 194, 211–213. doi: 10.1126/science.959852

Okobi, D. E. Jr., Banerjee, A., Matheson, A. M. M., Phelps, S. M., and Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988. doi: 10.1126/science.aau9480

Pollock, J. I. (1986). The song of the indris (Indri indri; Primates: Lemuroidea): natural history, form, and function. *Int. J. Primatol.* 7, 225–264. doi: 10.1007/BF02736391

Potretzke, S., and Ryabinin, A. E. (2019). The prairie vole model of pair-bonding and its sensitivity to addictive substances. *Front. Psychol.* 10:2477. doi: 10.3389/fpsyg.2019.02477

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., et al. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proc. R. Soc. B* 283:20152830. doi: 10.1098/rspb.2015.2830

Raemaekers, J. J., and Raemaekers, P. M. (1985). Long-range vocal interactions between groups of gibbons (*Hylobates lar*). *Behaviour* 95, 26–44. doi: 10.1163/156853985X00037

Raemaekers, J. J., Raemaekers, P. M., and Haimoff, E. H. (1984). Loud calls of the gibbon repertoire, (*Hylobates lar*): organization and context. *Behaviour* 91, 146–189. doi: 10.1163/156853984X00263

Ravignani, A., Bowling, D. L., and Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Front. Psychol.* 5:1118. doi: 10.3389/fpsyg.2014.01118

Reichard, U. H. (2003). "Social monogamy in gibbons: the male perspective," in *Monogamy. Mating Strategies and Partnerships in Birds, Humans and Other Mammals*, eds U. H. Reichard and C. Boesch (Cambridge: Cambridge University), 190–213. doi: 10.1017/cbo9781139087247.013

Rek, P., and Magrath, R. D. (2016). Multimodal duetting in magpie-larks: how do vocal and visual components contribute to a cooperative signal's function? *Anim. Behav.* 117, 35–42. doi: 10.1016/j.anbehav.2016.04.024

Rek, P., and Magrath, R. D. (2020). Visual displays enhance vocal duet production and the perception of coordination despite spatial separation of partners. *Anim. Behav.* 168, 231–241. doi: 10.1016/j.anbehav.2020.08.002

Rivera-Cáceres, K. D. (2015). Plain wrens *Cantorchilus modestus* zeledoni adjust their singing tempo based on self and partner's cues to perform precisely coordinated duets. *J. Avian Biol.* 46, 361–368. doi: 10.1111/jav.00575

Rivera-Cáceres, K. D., and Templeton, C. N. (2019). A duetting perspective on avian song learning. *Behav. Processes* 163, 71–80.

Rivera-Cáceres, K. D., Quirós-Guerrero, E., Araya-Salas, M., and Searcy, W. A. (2016). Neotropical wrens learn new duet rules as adults. *Proc. R. Soc. B* 283:20161819. doi: 10.1098/rspb.2016.1819

Rivera-Cáceres, K. D., Quirós-Guerrero, E., Araya-Salas, M., Templeton, C. N., and Searcy, W. A. (2018). Early development of vocal interaction rules in a duetting songbird. *R. Soc. Open Sci.* 5:171791. doi: 10.1098/rsos.171791

Robinson, J. G. (1979). An analysis of the organization of vocal communication in the titi monkey, *Callicebus moloch*. *Z. Tierpsychol.* 49, 381–403. doi: 10.1111/j.1439-0310.1979.tb00300.x

Rogers, A. C. (2005). Male and female song structure and singing behavior in the duetting eastern whipbird, *Psophodes olivaceus*. *Aust. J. Zool.* 53, 157–166. doi: 10.1071/ZO04083

Sainburg, T., Thielk, M., and Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* 16:e1008228. doi: 10.1371/journal.pcbi.1008228

Schegloff, E., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 362–382. doi: 10.2307/413107

Shekelle, M., Groves, C. P., Maryanto, I., Mittermeier, R. A., Salim, A., and Springer, M. S. (2019). A new tarsier species from the Togean islands of central Sulawesi, Indonesia, with references to Wallacea and conservation on Sulawesi. *Primate Conserv.* 33, 65–73.

Simonyan, K. (2014). The laryngeal motor cortex: its organization and connectivity. *Curr. Opin. Neurobiol.* 28, 15–21. doi: 10.1016/j.conb.2014.05.006

Snowdon, C. T. (2009). Plasticity of communication in nonhuman primates. *Adv. Stu. Behav.* 40, 239–276. doi: 10.1016/S0065-3454(09)40007-X

Snowdon, C. T. (2017a). Learning from monkey "talk". *Science* 355, 1120–1122. doi: 10.1126/science.aam7443

Snowdon, C. T. (2017b). "Vocal communication in family living and pair-bonded primates," in *Primate Hearing and Communication*, eds R. M. Quam, M. A. Ramsier, R. R. Fay, and A. N. Popper (Berlin: Springer), 141–174. doi: 10.1007/978-3-319-59478-1_6

Snowdon, C. T. (2018). Cognitive components of vocal communication: a case study. *Animals* 8:126. doi: 10.3390/ani8070126

Snowdon, C. T., and Elowson, A. M. (1999). Pygmy marmosets modify vocal structure when paired. *Ethology* 105, 893–908. doi: 10.1046/j.1439-0310.1999.00483.x

Takahashi, D. Y., El Hady, A., Zhang, Y. S., Liao, D. A., Montaldo, G., Urban, A., et al. (2021). Social-vocal brain networks in a non-human primate. *bioRxiv [Preprint]* doi: 10.1101/2021.12.01.470701

Templeton, C. N., Mann, N. I., Ríos-Chelén, A. A., Quiros-Guerrero, E., Macías Garcia, C., and Slater, P. J. B. (2013). An experimental study of duet integration in the happy wren, *Pheugopedius felix*. *Anim. Behav.* 86, 821–827. doi: 10.1016/j.anbehav.2013.07.022

Tenaza, R. R. (1976). Songs, choruses, and countersinging of Kloss' gibbons (*Hylobates klossii*) in Siberut Island, Indonesia. *Z. Tierpsychol.* 40, 37–52. doi: 10.1111/j.1439-0310.1976.tb00924.x

Tenaza, R. R. (1985). Songs of hybrid gibbons (*Hylobates lar* x *N. muelleri*). *Am. J. Primatol.* 8, 249–253. doi: 10.1002/ajp.1350080307

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2015). Lar gibbon (*Hylobates lar*) great call reveals individual caller identity. *Am. J. Primatol.* 77, 811–821. doi: 10.1002/ajp.22406

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2016). Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *BMC Evol. Biol.* 16:4. doi: 10.1186/s12862-015-0578-8.

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018a). Male white-handed gibbons flexibly time duet contributions. *Behav. Ecol. Sociobiol.* 72:16. doi: 10.1007/s00265-017-2432-z

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018b). An analysis of white-handed gibbon male song reveals speech-like phrases. *Am. J. Phys. Anthropol.* 2018, 1–12. doi: 10.1002/ajpa.23451

Tilson, R. L., and Tenaza, R. R. (1976). Monogamy and duetting in an Old World monkey. *Nature* 263, 320–321. doi: 10.1038/263320a0

Tobias, J. A., and Seddon, N. (2009). Signal jamming mediates sexual conflict in a duetting bird. *Curr. Biol.* 19, 577–582. doi: 10.1016/j.cub.2009.02.036

Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., and Nakagawa, S. (2016). Territoriality, social bonds, and the evolution of communal signaling in birds. *Front. Ecol. Evol.* 4:74. doi: 10.3389/fevo.2016.00074

Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Randrianarison, R. M., Friard, O., et al. (2017). An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. *Sci. Rep.* 7:10140. doi: 10.1038/s41598-017-10656-9

Torti, V., Gamba, M., Rabemananjara, Z. H., and Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Ital. J. Zool.* 80, 596–607. doi: 10.1080/11250003.2013.845261

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., ten Cate, C., et al. (2021). The multi-dimensional nature of vocal learning. *Phil. Trans. R. Soc. B* 376:20200236. doi: 10.1098/rstb.2020.0236

Voigt, C., Leitner, S., and Gahr, M. (2006). Repertoire and structure of duet and solo songs in cooperatively breeding white-browed sparrow weavers. *Behaviour* 143, 159–182. doi: 10.1163/156853906775900739

Wickler, W. (1980). Vocal dueting and the pair bond. I. Coyness and partner commitment: a hypothesis. *Z. Tierpsychol.* 52, 201–209.

Wickler, W., and Seibt, U. (1980). Vocal dueting and the pair bond. II. Unisono dueting in the African forest weaver, *Symplectes bicolor*. *Z. Tierpsychol.* 52, 217–226. doi: 10.1111/j.1439-0310.1980.tb00713.x

Wilson. (1975). Discriminative conditioning of vocalizations in *Lemur catta*. *Anim. Behav.* 23, 432–436. doi: 10.1016/0003-3472(75)90091-3

Zhang, Y. S., Takahashi, D. Y., Liao, D. A., Ghazanfar, A. A., and Elemans, C. P. H. (2020). Vocal state change through laryngeal development. *Nat. Comm.* 10:4592. doi: 10.1038/s41467-019-12588-

Zürcher, Y., Willems, E. P., and Burkart, J. M. (2021). Trade–offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Sci. Rep.* 11:15683. doi: 10.1038/s41598-021-95101-8

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fevo.2022.898509

Coordinated Singing in Coppery Titi Monkeys (*Plecturocebus cupreus*): Resource or Mate Defense?

***Sofya Dolotovskaya*[1,2,3]* *and Eckhard W. Heymann*[1]**

[1] Behavioral Ecology and Sociobiology Unit, German Primate Center - Leibniz-Institute for Primate Research, Göttingen, Germany, [2] Primate Genetics Laboratory, German Primate Center - Leibniz-Institute for Primate Research, Göttingen, Germany, [3] Laboratory of Comparative Ethology and Biocommunication, Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia

OPEN ACCESS

Edited by:
Charles T. Snowdon,
University of Wisconsin-Madison,
United States

Reviewed by:
Allison Lau,
University of California, Davis,
United States
Sarie Van Belle,
University of Texas at Austin,
United States

****Correspondence:***
Sofya Dolotovskaya
s.dolotovskaya@gmail.com

Specialty section:
This article was submitted to
Behavioral and Evolutionary Ecology,
a section of the journal
Frontiers in Ecology and Evolution

Received: *17 March 2022*
Accepted: *26 April 2022*
Published: *26 May 2022*

Citation:
Dolotovskaya S and Heymann EW
(2022) Coordinated Singing
in Coppery Titi Monkeys
(Plecturocebus cupreus): Resource or
Mate Defense?
Front. Ecol. Evol. 10:898509.
doi: 10.3389/fevo.2022.898509

Coordinated singing, performed as duets by mated pairs and often joined by offspring to form choruses, is a distinctive behavioral attribute of the social system of pair-living and pair-bonded Neotropical titi monkeys. Duets and choruses are presumed to be associated with mate or territorial defense, but no consensus has yet been reached regarding their function. Here, we examined temporal and spatial patterns of coordinated singing in eight wild groups of coppery titi monkeys, *Plecturocebus cupreus*, in Peruvian Amazonia to test predictions of the joint resource and mate defense. We investigated singing rates in relation to female reproductive state, fruit consumption and demographic context using a dataset based on 227 observation days and analyzed temporal and spatial distribution of songs using a dataset based on 150 songs, collected between June 2017 and September 2021. Titi monkeys sang least frequently when females were likely to be sexually receptive and most frequently when females were likely to be pregnant. Groups also sang slightly more often when fruits were consumed more intensively, although this association did not reach statistical significance. The duration of songs was not associated with female reproductive state or fruit consumption, but songs were longer during inter-group encounters compared to non-encounter contexts. Songs were not concentrated in the core areas of home ranges; rather, they were distributed throughout the home ranges in concordance with its use. Finally, songs were concentrated around dawn. Our results provide support for a function in joint resource defense and inter-group communication of coordinated songs in coppery titi monkeys. The function of coordinated songs for mate defense in the form of paternity guarding, on the other hand, was not supported by our findings.

Keywords: duets, coordinated singing, titi monkeys, *Plecturocebus*, resource defense, mate defense

INTRODUCTION

Duets are joint acoustic displays where two individuals coordinate their songs so that they alternate or overlap (Hall, 2009, 2004; De Gregorio et al., 2022). Among vertebrates, duetting is taxonomically most widespread among birds (although only a minority of birds duet; Hall, 2009). In mammals, duetting is mainly, but not exclusively, found in primates (Tilson and Norton, 1981; Paula and Monticelli, 2021). Both in birds and in primates, duetting is most common in pair-living species with long-term pair bonds and year-round territoriality (Malacarne et al., 1991; Benedict, 2008; Adret et al., 2018). Duetting primates include the Neotropical titi monkeys (genera *Callicebus*, *Plecturocebus*, and *Cheracebus*), south-east Asian gibbons (all species except *Hylobates moloch* and *Hylobates klossii*), Malagasy indris (*Indri indri*), Mentawai langurs (*Presbytis potenziani*), and Sulawesi tarsiers (genus *Tarsius*) (Tilson and Tenaza, 1976; Haimoff, 1986; Nietsch, 1999; Méndez-Cárdenas and Zimmermann, 2009; Adret et al., 2018; Bonadonna et al., 2020).

Duets are usually produced by paired females and males (although there are exceptions, such as male-male duets in lekking manakins of the genus *Chiroxiphia*: Snow, 1977) (Hall, 2004). In some species, offspring can join the adults' duets to form choruses — e.g., in titi monkeys (Adret et al., 2018) and yellow-cheeked gibbons, *Nomascus gabriellae* (Merker and Cox, 1999). It has been suggested that the participation of the offspring in the adult singing may represent a form of practicing toward a development of adult-like song (De Gregorio et al., 2022).

Although duets have been studied extensively in birds, it remains unclear why in some species an individual coordinates its song with those of its partner instead of singing independently (Hall, 2009). Most avian studies agree that duets are largely cooperative displays, where song coordination provides one or several benefits to both individuals (Mennill and Vehrencamp, 2008; Hall, 2009). Several hypotheses have been proposed to describe the nature of these benefits.

According to the joint resource defense hypothesis, duets function as a cooperative display to outsiders, advertising an ownership of a home range and/or resources (Hall, 2004; Logue, 2005; Rasoloharijaona et al., 2006; Caselli et al., 2015). According to the mate defense hypothesis, an individual participates in coordinated singing to advertise its partner's or its own mated status to outsiders and repel potential rivals (Robinson, 1981; Levin, 1996; Appleby et al., 1999; Hall, 2004). Within a mate defense function, an individual joins its partner's song in a duet to advertise a partner's mated status to outsiders and repel potential rivals (Robinson, 1981; Levin, 1996; Appleby et al., 1999; Hall, 2004). There are various forms of mate defense: an individual may prevent same-sex rivals from pairing with its partner or reinforce a partner's position within the pair by advertising its status to opposite-sex outsiders. One specific type of mate defense is paternity guarding, where a male joins his mate's singing to repel rival males seeking extra-pair copulations (Hall, 2004). In addition, an individual contribution to duetting has been proposed to be directed at a partner to function for pair-bond reinforcement or signaling commitment to a partner (Hall, 2004; Méndez-Cárdenas and Zimmermann, 2009).

Predictions for these hypotheses can be divided into two groups, one explaining the function of singing regardless of whether it is coordinated or not and the other explaining why duetting in each of these contexts is more effective than solo or uncoordinated singing in achieving the corresponding effect on listeners. For example, if duetting or uncoordinated singing functions for joint resource defense, it can be expected to be more frequent when defensible and valuable resources, such as fruits, are available. However, there are many features of duets that are consistent with a joint resource defense but do not distinguish duetting from uncoordinated singing — i.e., duets are often loud and performed in counter-singing interactions with neighbors or in response to intrusion (Hall, 2004, 2009). To show that duetting itself has a function for joint resource defense, it is necessary to show that coordination of songs plays a role in this function over and above that achieved by uncoordinated songs (Hall, 2009). Specifically, duets should be more threatening displays than uncoordinated songs and partners should be more likely to coordinate their songs into duets than to sing alone when faced with outsiders (Hall, 2004). Similarly, if singing functions, for example, for male mate defense in the form of paternity guarding, it should be more frequent when females are sexually receptive. However, to show that duetting is more efficient in achieving this function than uncoordinated singing, it is necessary to demonstrate that duets are initiated by females and that males join more of their partners' songs to form duets when females are sexually receptive (Hall, 2004).

Titi monkeys are Neotropical primates living in groups comprising one reproductive pair and one to three young (Kinzey, 1981; Kinzey and Robinson, 1983; Bicca-Marques and Heymann, 2013; Fernandez-Duque et al., 2013). They exhibit strong long-term pair bonds, year-round territoriality and biparental infant care (Anzenberger, 1988; Bicca-Marques and Heymann, 2013; Fernandez-Duque et al., 2013; Van Belle et al., 2016). Titi monkey pairs regularly sing in duets that are often joined by juvenile and subadult group members to form choruses (Caselli et al., 2014; Adret et al., 2018). Duets are composed of partially overlapping sequences sung by both mates with no sex-specificity in song components (Robinson, 1979a; Müller and Anzenberger, 2002). However, as shown by research on captive *Plecturocebus cupreus* (previously *Callicebus cupreus*), there is individuality in duet contributions of each mate, with moderate heritability of some song characteristics, and duets are pair-specific as a result of a summation of individual attributes of the two mates (Müller and Anzenberger, 2002; Lau et al., 2020; Clink et al., 2022). There is also evidence for vocal convergence among mates in some features of duets and for changes in duet elements correlated with pair-bond duration (Müller and Anzenberger, 2002; Clink et al., 2019).

So far, few studies investigated the function duets and choruses in wild titi monkeys. Generally, the results of these studies seem to be more compatible with the joint resource defense rather than the mate defense hypothesis. In observational and playback studies of black-fronted titi monkeys (*Callicebus nigrifrons*), coordinated songs were produced more often in months with higher fruit availability, while groups did not sing more often during receptive periods of females; duets were initiated by either partner, and individuals did not show sex-specific responses to the playback of solo songs (Caselli et al., 2014, 2015). In *Plecturocebus toppini* (previously *Callicebus brunneus*), males reacted stronger to playbacks in the high-used versus low-used parts of the home range (Lawrence, 2007), also supporting the resource defense hypothesis. On the other hand, *Plecturocebus ornatus* (previously *Callicebus moloch*) showed sex-specific reactions to playbacks, and males often sang alone, lending some support to the mate defense hypothesis (Robinson, 1981).

Here, we examine temporal and spatial patterns of coordinated singing in eight wild groups of coppery titi monkeys, *Plecturocebus cupreus*, to test some of the predictions of the joint resource and mate defense hypotheses for the function of coordinated singing. If singing is more important in *joint resource defense*, we predict that songs would be produced more frequently and/or be longer when defensible and valuable resources, such as fruits, are consumed more intensively; and/or when groups need more resources, e.g., when females are pregnant or lactating and/or when there are more group members. If, on the other hand, singing is more important in *mate defense* (in the form of paternity guarding), we predict that songs would be produced more frequently and/or be longer when females are sexually receptive. In addition, we analyzed the spatial distribution of coordinated songs to see whether they are concentrated at the territory borders or are produced throughout the home range in concordance with its use. Finally, we examined temporal distribution of singing throughout the day to see whether songs are concentrated round dawn when sound propagation is optimal (Brown and Handford, 2003). If songs are concentrated around dawn, it would suggest that singing is used for inter-group communication, as opposed to pair-bond reinforcement where songs would be expected to be produced throughout the day. Because this study was observational, we could not analyze the reaction of listeners to different types of songs in different contexts to test the functions of duetting itself as opposed to uncoordinated singing. However, in the Discussion we put our findings in the context of previous experimental studies on wild and captive titi monkeys in order to explore the functions of coordinated singing in this taxon.

TABLE 1 | Observation periods, group compositions, and observation times for eight studied groups.

Group	Study period	Group composition[(1)]	Days observed
1	Jun–Dec 2017	AM, AF, SM, Juv	56
	Sep–Dec 2018	AM, AF, SF, Juv	33
	March–May 2019	AM, AF, SF, Juv	13
	Sep 2021	AM, AF, SM	5
2	Sep–Oct 2017	AM, AF	17
	Jul 2018 2018	AM, AF	5
	Nov–Dec 2018	AM, AF, Inf	25
	March–Jul, Sep 2019	AM, AF, Juv	20
3	Oct–Dec 2017	AM, AF, Juv	26
	Jun-Jul, Oct–Nov 2018	AM, AF, SF, SF	26
	Nov 2018	AM, AF, SF, SF, Inf	3
4	Jun–Jul 2018, Sep–Oct 2018	AM, AF, Juv	50
5	Jun–Oct 2018	AM, AF, Juv	54
	Apr–Jun 2019	AM, AF, SM, Inf	9
	Sep 2019	AM, AF, Juv	13
	Sep 2021	AM, AF, S*	4
6	Aug–Oct 2017	AM, AF, SM, SF, Juv, Inf	44
	Jul-Nov 2018	AM, AF, SM, SF, Juv	33
	Sep 2021	AM, AF, SF, S*, Juv	2
7	June-Aug, Oct 2018	AM, AF	20
	Nov-Dec 2018	AM, AF, Inf	19
11	Oct–Nov 2018	AM, AF	4
	Sep 2021	AM, AF, Juv	9

[(1)]AM, adult male; AF, adult female; SM, subadult male; SF, subadult female (distinguishable by size from adults); Juv, juvenile (>4.5 months); Inf, infant (<4.5 months). Sex could not be determined for juveniles and infants due to small genital size.
**Sex unknown.*

METHODS

Study Site and Animals

The study was conducted at the Estacioìn Bioloìgica Quebrada Blanco (EBQB) in the Peruvian Amazon (4°21′S, 73°09′W; for details of EBQB see Heymann et al., 2021). We studied eight habituated titi monkey groups in June–December 2017, June–December 2018, March–July and September 2019, and September 2021 (see **Table 1** for observation periods, group compositions, and observation times for each group). Group 1 had been studied intermittently since 1997 (e.g., Tirado Herrera and Heymann, 2004) and was well habituated to the presence of humans by the start of this study; the other groups were habituated in 2017–2018.

Data Collection

Each group was followed by a team of two observers in blocks of 5–6 days. We followed a group from the early morning when the animals left a sleeping tree (most often between 5:30 h and 6:30 h) or from when we located the group until the late afternoon when the animals retired to a sleeping tree (most often between 16:00 h and 17:00 h) or until we lost sight of them. During follows, we used instantaneous scan sampling to record the activity of all non-infant (i.e., independently moving) group members at 10-min intervals, allowing 2 min for the location of the animals. During feeding, when possible, we specified the type of food as fruits, leaves, flowers, arthropods, or soil (from termite nests). During each scan point, we recorded the location of a focal group using a handheld GPS unit (Garmin GPSMAP 62s or 64s).

Starting from June 2018, we recorded duration, location, and context (whether a song was produced during an intergroup encounter or not) of all coordinated songs (duets and choruses) produced by focal groups, i.e., groups followed at the moment. A coordinated song was defined as overlapping stereotyped singing of two or more individuals (Adret et al., 2018; Clink et al., 2019). A singing bout was defined as singing interrupted by pauses of less than 5 min; when all group members were silent for $\geq$ 5 min in between singing bouts, we scored two singing bouts as independent. When calculating the duration of each singing bout, we did not exclude the duration of pauses because when the pauses were shorter than 5 min (our cut-off for scoring two independent bouts) animals usually stayed agitated and continued singing intermittently. We scored encounters whenever the focal group came within a visual contact with another group and responded to its presence by singing and/or chasing. We considered two encounters to be independent when all participants stopped singing and chasing for more than 30 min. In addition, in June–December 2017, we opportunistically recorded duration and location of some (but not all) of songs made by focal groups.

Data Analysis

We made two datasets for our analyses. Dataset 1 comprised data on presence or absence of singing during observation days. This dataset was used to examine whether songs were more frequent on days when fruits were consumed more intensively, when females were sexually receptive, or when groups had unweaned infants or more group members. We included data from June–December 2018, September 2019, and September 2021 in this dataset, as data on presence/absence of singing was only collected systematically during these periods. As animals were active for 10 h a day on average, we excluded days with less than 5 h of observation (less than 30 scan points) from our analyses. We further excluded days for which observations started after 10:00 h as most of songs occurred before this time. This resulted in a dataset of 227 days across eight study groups. Dataset 2 was used to analyze the temporal and spatial distribution of songs and to examine whether songs were longer on days when fruits were consumed more intensively, when females were sexually receptive, or when groups had unweaned infants or more group members, or during intergroup encounters. This dataset included 159 singing bouts across eight study groups for which data were collected starting from June 2017.

To estimate female reproductive state, we used data on infant birth dates (N = 15 births in seven study groups between June 2017 and September 2021) and observed copulations. An infant's date of birth was estimated as the midpoint between the dates when a group was last seen without and first seen with an infant. The difference between these dates varied between 0 and 26 days. We supplemented these data with visual estimation of infant age based on its body size and tail coloration. To estimate periods when females were likely pregnant, we counted back from an estimated birth date using the average gestation length of 128 days in captive *P. cupreus* (Valeggia et al., 1999). We estimated that females started to be sexually receptive again 196 days after an estimated birth date as 196 days was the average duration of lactational anovulation in captive *P. cupreus* (Valeggia et al., 1999).

To estimate when fruits were consumed more intensively, we used a mean monthly proportion of feeding time allocated to fruits by each study group. We first calculated the daily proportions of feeding time allocated to fruits by each adult animal by dividing the number of scan points allocated to fruits by the total daily number of scan points. We then averaged these values for each month and each group. To make these data comparable between groups and months, we used data only for breeding adults because group composition varied between different groups and periods.

What Affected the Probability of Singing?

To test whether the probability of singing was affected by female receptivity, the presence of unweaned infant in a group, the proportion of feeding time allocated to fruits, or group size, we used Dataset 1 and ran a generalized linear mixed model (GLMM; Baayen, 2008) with binomial error structure and logit link function. As a response variable, we used the presence or absence of singing (yes/no) on an observation day. As test predictors, we used female reproductive state (receptive/pregnant/lactating), mean monthly proportion of feeding time allocated to fruits by each study group, and group size (number of non-infants, i.e., independently moving group members). To control for the possible effect of seasonality, we included rainfall data [monthly averages in mm at the nearest meteorological station in Tamshiyacu (4°00′10.7″S 73°09′38.2″W), ca. 40 km north of EBQB, data available at[1]] as a control predictor. To account for repeated observations, we used group ID as a random-effect predictor. The sample size for this model was 171 observation days across seven groups.

Prior to fitting the model, we z-transformed fruit proportion, group size, and rainfall to make a model more easily interpretable (Schielzeth, 2010) and make model convergence more likely. To rule out collinearity, we determined Variance Inflation Factors [VIF, Quinn and Keough (2002)] for a standard linear model excluding the random effects; all VIFs were close to 1 and thus not of issue. After fitting the model, we assessed model stability by excluding the levels of a random effect one by one from the full model. As an overall test of the effect of the test predictors, we compared the full model with a null model lacking the fixed-effect predictors but otherwise having the same structure as the full model (Forstmeier and Schielzeth, 2011) using a likelihood ratio test (Dobson et al., 2008). To test the effect of the individual predictors, we applied likelihood ratio tests (Barr et al., 2013) using R function drop1. We fitted the model in R [version 4.1.1; R Core Team (2021)] using the function lmer of the package lme4 [version 1.1–27.1; Bates et al. (2015)]. We determined VIFs using the function vif of the package car [version 3.0–11; Fox and Weisberg (2011)]. We assessed model stability and bootstrapped model estimates using functions kindly provided by Roger Mundry. The model was fairly stable for all the estimates.

What Affected the Duration of Singing Bouts?

To test whether the duration of singing bouts was affected by female fertility, the presence of an unweaned infant in a group, the proportion of feeding time allocated to fruits, or the context of singing (song produced during an encounter or not), we used Dataset 2 and ran a linear mixed model (LMM; Baayen, 2008). As a response variable, we used the duration of singing bouts (in min). As test predictors, we used female reproductive state (receptive/pregnant/lactating), mean monthly proportion of feeding time allocated to fruits by each study group, group size, and the context of singing (encounter vs. non-encounter).

---

[1] https://www.senamhi.gob.pe

As in the previous model, we included rainfall data as a control predictor. To account for the repeated observations, we used group ID as a random-effect predictor. The sample size for this model was 114 songs across seven groups.

Prior to fitting the model, we square-root-transformed the response to achieve an approximately symmetrical distribution and z-transformed fruit proportion and rainfall to make a model more easily interpretable (Schielzeth, 2010) and make model convergence more likely. To rule out collinearity, we determined Variance Inflation Factors [VIF, Quinn and Keough (2002)] for a standard linear model excluding the random effects; all VIFs were close to 1 and thus not of issue. After fitting the model, we assessed model stability by excluding the levels of a random effect one by one from the full model. As an overall test of the effect of the test predictors, we compared the full model with a null model lacking the fixed-effect predictors but otherwise having the same structure as the full model (Forstmeier and Schielzeth, 2011) using a likelihood ratio test (Dobson et al., 2008). To test the effect of the individual predictors, we applied likelihood ratio tests [Barr et al. (2013)] using R function drop1. We fitted the model in R [version 4.1.1; R Core Team (2021)] using the function lmer of the package lme4 [version 1.1–27.1; Bates et al. (2015)]. We determined VIFs using the function vif of the package car [version 3.0–11; Fox and Weisberg (2011)]. We assessed model stability and bootstrapped model estimates using functions kindly provided by Roger Mundry. The model was fairly stable with the exception for the estimate of the effect of reproductive state.

Spatial Distribution of Songs

To analyze the spatial distribution of songs, we first estimated the home range and core areas of each group using the fixed kernel density method and the reference smoothing factor *h* (Erran Seaman and Powell, 1996) in the R package adehabitatHR (Calenge, 2006). We defined home ranges as the 95% fixed kernel contour and the core areas as the 50% fixed kernel contour (Asensio et al., 2012; Holzmann et al., 2012). In addition, we drew an inner 25 m border area within the 95% home range of each group using QGIS 3.22.3 (QGIS Development Team, 2022). We then mapped the locations of singing bouts onto the home ranges and used Fisher's exact tests to compare the frequency of singing in the core area and in the rest of the 95% home range with expected values for each group separately. We did not analyze the border areas separately because the number of songs given in these areas was too low ($N = 1$–6 songs per group). Expected values were calculated under the null hypothesis of songs being evenly distributed across the home ranges, taking into account the time spent in each area; the time spent in each area was calculated in QGIS 3.22.3 as the proportion of GPS points taken within each area. First, we did these analyses for all songs; next, we repeated the analyses excluding songs given during the intergroup encounters ($N = 52$) because many of these songs were concentrated near the border areas and could potentially bias the result. Statistical tests were 2-tailed, with $p \leq 0.05$.

Temporal Distribution of Songs

To see how songs were distributed over the activity period and whether songs were concentrated in the morning, we divided each observation day into 1-h intervals relative to the time of sunrise to account for seasonal variability in sunrise times. We then calculated the number of recorded singing bouts that began in each interval.

RESULTS

What Affected the Probability of Singing?

The probability of singing was affected by female reproductive state (comparison of full model with reduced model not comprising reproductive state: $\chi^2 = 15.697$, df = 2, $P = 0.0004$; **Table 2** and **Figures 1**, **2**). Specifically, the probability of singing was lowest when females were sexually receptive, slightly higher when females were lactating (i.e., when there was an unweaned infant in a group), and considerably higher when females were pregnant. The probability of singing was also somewhat affected by fruit proportion, with groups singing slightly more frequently when fruit proportion in their diet was higher, although this effect did not reach statistical significance (comparison of full model with reduced model not comprising fruit proportion: $\chi^2 = 3.442$, df = 1, $P = 0.064$; **Table 2** and **Figures 1**, **2**). Group size and rainfall did not affect singing probability (full-reduced model comparisons for group size: $\chi^2 = 1.471$, df = 1, $P = 0.225$; for rainfall: $\chi^2 = 0.189$, df = 1, $P = 0.664$).

What Affected the Duration of Singing?

The duration of singing bouts was affected only by context, with bouts being longer during encounters (comparison of full model with reduced model not comprising context: $\chi^2 = 14.555$, df = 1, $P = 0.0001$; **Table 3** and **Figure 3**). Fruit proportion or female reproductive state did not affect the duration of singing bouts (comparisons of full model with reduced model not comprising fruit proportion: $\chi^2 = 0.948$, df = 1, $P = 0.330$, reproductive state: $\chi^2 = 1.024$, df = 2, $P = 0.599$; **Table 3**).

Spatial Distribution of Songs

Song produced outside of the intergroup encounter context were distributed throughout the home ranges in concordance with its use (**Figure 4**). The observed frequencies of singing within core areas and the rest of the home ranges did not significantly differ from the expected frequencies calculated under the null hypothesis of songs being distributed throughout the home range in concordance with its use, both when analyzing all songs and when excluding songs given during the intergroup encounters from the analyses (Fisher's exacts tests for all songs: group 1, $P = 0.094$; group 2, $P = 0.176$; group 3, $P = 1$; group 4, $P = 1$; group 5, $P = 1$; group 6, $P = 0.417$; group 7, $P = 0.608$; group 11, $P = 0.444$).

Temporal Distribution of Songs

Singing showed a clear peak around dawn (**Figure 5**). Approximately half from all 159 recorded singing bouts (75

TABLE 2 | Results of the model of the effects of female reproductive state, fruit proportion in the diet, and group size on the probability to of singing: estimates, together with standard errors, confidence intervals, test results, and minimum and maximum of model estimates derived by dropping levels of random effects one at a time.

Term	Estimate	SE	Lower CI	Upper CI	*z*-value	*P*-value	min	max
Intercept	−2.040	0.517	−3498	−1.096	–	–	−2.190	−1.900
Group size(1)	−0.304	0.273	−0.889	0.197	−1.115	0.225	−0.516	−0.187
Fruit proportion(2)	0.422	0.238	0.013	1.019	1.777	0.064	0.278	0.776
Reproductive state (lactation)(3)	1.180	0.616	−0.059	2.791	1.916	0.055	0.482	1.816
Reproductive state (pregnancy)(3)	2.507	0.697	1.251	4.511	3.595	0.0003	2.125	2.971
Rainfall(4)	−0.083	0.193	−0.550	0.291	−0.431	0.664	0.000	0.639

(1) *z-transformed to a mean of 0 and a standard deviation of 1; mean and SD of the original variable were 3.22 and 1.01, respectively.*
(2) *z-transformed to a mean of 0 and a standard deviation of 1; mean and SD of the original variable were 0.88 and 0.12, respectively.*
(3) *Reproductive state was dummy coded with receptivity being as the reference level, and comparisons here are with the reference level of receptivity; the difference between lactation and pregnancy was estimated as 1.327 ± 0.569, z = 2.334, P = 0.020).*
(4) *z-transformed to a mean of 0 and a standard deviation of 1; mean and SD of the original variable were 165.48 and 67.81, respectively.*

bouts, 47%) started within an hour before or after sunrise, and most singing bouts (131 bouts, 82%) started within 1 h before or 3 h after sunrise. When analyzing songs produced in non-encounter context separately ($N = 107$), the dawn peak became even more pronounced: more than half of singing bouts (66 bouts, 62%) started within an hour before or after the sunrise and almost all singing bouts (101 bouts, 94%) started within 1 h before or 3 h after the sunrise; only six bouts were recorded later than 3 h after the sunrise and only one bout was recorded after midday. Songs produced during encounters ($N = 52$) were more evenly distributed throughout the day, with only a slight increase around dawn. Around half of the encounter singing bouts (29 bouts, 55%) were produced within 3 h after the sunrise and then frequency of encounter songs decreased slowly toward the end of the day.

DISCUSSION

Our findings suggest that coordinated songs in coppery titi monkeys function for joint resource defense and inter-group communication. The function of coordinated songs for mate defense in the form of paternity guarding, on the other hand, was not supported by our results. Groups sang least frequently when females were likely to be sexually receptive, more frequently when females were lactating, and most frequently when females were likely to be pregnant. There was also a tendency for groups to sing more often when fruits were consumed more intensively, although this trend was not statistically significant. The duration of songs was not associated with female reproductive state or fruit consumption, but songs were longer during inter-group encounters compared to non-encounter contexts. Songs were not concentrated in the core areas of home ranges; rather, they were distributed throughout the home ranges in concordance with its use. Finally, songs were concentrated around dawn, supporting their function for inter-group communication.

Joint Resource Defense Hypothesis

In line with our predictions of the joint resource defense, songs were more frequent when females were lactating or likely to be pregnant—i.e., when groups were expected to have increased energetic demands. Higher duetting rates during lactation were also demonstrated in Milne Edwards' sportive lemurs, *Lepilemur edwardsi* (Méndez-Cárdenas and Zimmermann, 2009). Lactation is considered to be the most energetically costly part of mammal reproduction (Clutton-Brock et al., 1989; Altmann and Samuels, 1992). Indeed, in our previous study on the same population, females increased their feeding time and consumed more arthropods (presumably rich in proteins) during lactation (Dolotovskaya and Heymann, 2020), suggesting increased requirements for energy and nutrients. Gestation is generally not as energetically demanding as lactation (Gittleman and Thompson, 1988), and in some primate species, pregnant and cycling females do not even differ in energy intake (e.g., white-faced capuchins, *Cebus capucinus:* McCabe and Fedigan, 2007). However, energy intake in pregnant females is still often increased compared to cycling or anestrous non-cycling females, and in several primate species pregnant females consume more or higher quality food (e.g., chimpanzees, *Pan troglodytes*: Murray et al., 2009); red-ruffed lemurs, *Varecia rubra*: Vasey, 2004, 2005). If the frequency of singing reflects the animals' need of resources and motivation to defend them, then singing can be expected to be more frequent during lactation, not pregnancy. It is possible in titi monkeys, where infants are carried not by females but by males most of the time (Wright, 1984; Tirado Herrera and Heymann, 2004; Lawrence, 2007; Fernandez-Duque et al., 2013; Spence-Aizenberg et al., 2016), the male's contribution relieves a mother from some of the costs of infant care, making an increase in energetic demands during lactation less pronounced than in species where infants are carried by mothers.

To date, no studies have assessed energy demands of lactation vs. pregnancy in titi monkeys. In Azara's owl monkeys (*Aotus azarae*), who are similar to titi monkeys in size, social system, and patterns of biparental care (infants are carried by males most of the time: Fernandez-Duque et al., 2020), fecal cortisol in both sexes was highest during gestation compared to lactation, possibly indicating higher energetic costs of gestation compared to lactation in this species (Corley et al., 2021). However, this effect might have been due to seasonality, as gestation in *A. azarae* takes place during the southern winter in the Argentinean Chaco, where the study was conducted (Corley et al., 2021). At our site reproduction is not strictly seasonal, with births having been

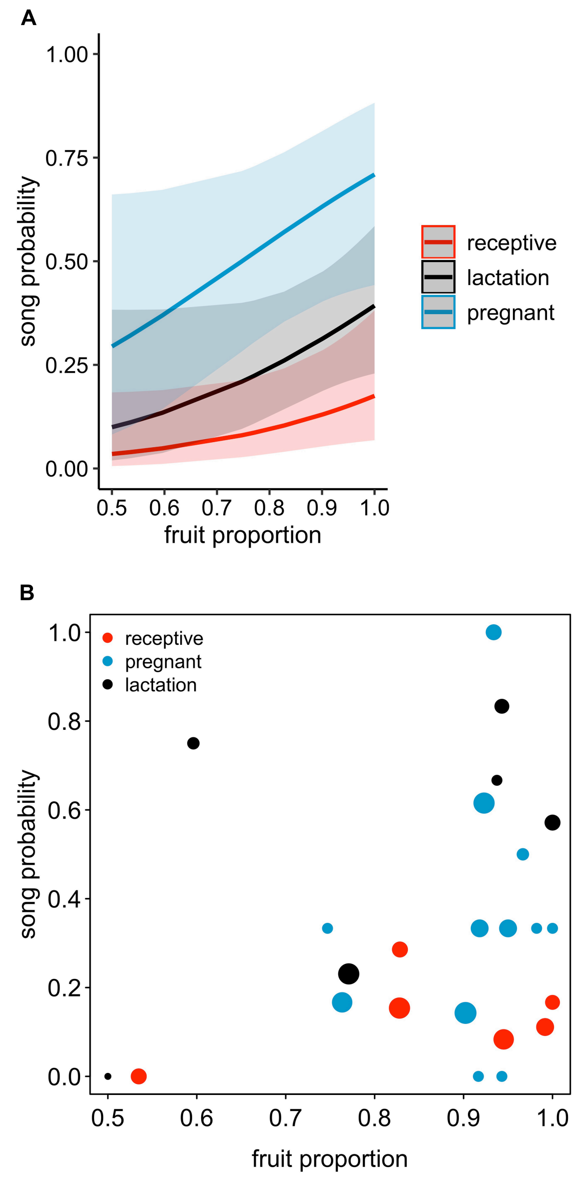


FIGURE 1 | (A) Song probability as a function of female reproductive state and proportion of feeding time allocated to fruits, shown separately for each female reproductive state. The lines depict the fitted model (based on group size at its average), and blue, black, and red areas show the corresponding 95% confidence intervals. **(B)** Proportions of singing days for different female reproductive states and different fruit proportions, with the area of the dots increasing linearly with the respective sample size (1 to 14 observation days per each combination of reproductive state and fruit proportion, total N = 171 observation days).

recorded in July, August, and September (dry season) as well as October through February (rainy season). To see whether our finding of higher singing frequency during gestation is explained by higher energetic demands of pregnant females, it will be necessary to study activity budgets or variation in cortisol levels during different reproductive periods and to have a larger dataset to separate the effects of seasonality and reproductive stage.

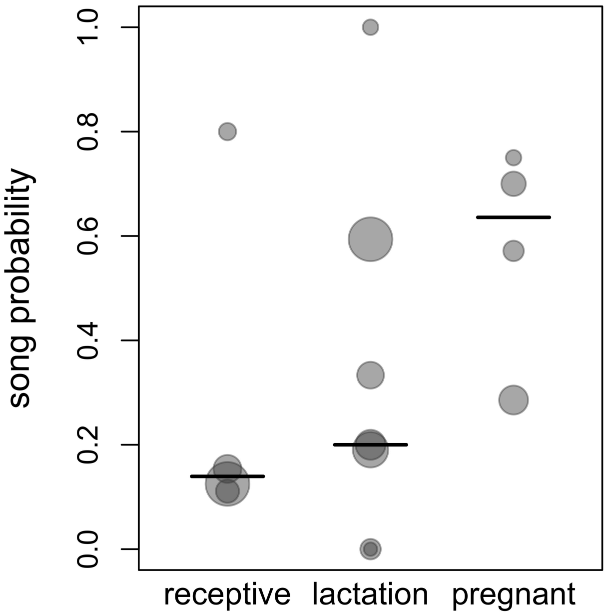


FIGURE 2 | Song probability for different female reproductive states. Shown are proportions of singing days of all observation days, with each dot corresponding to one group ID and the area of the dots increasing linearly with the respective sample size for a given group and a given reproductive state (3 to 32 observation days per each combination of reproductive state and group ID, total N = 171 observation days). The lines depict the median values for each reproductive state.

Our study groups also sang slightly more frequently when their fruit consumption was higher, although this result did not reach statistical significance. Average monthly feeding time allocated to fruits was used in this study as a proxy for seasonal changes in fruit availability, as we did not have direct measures of fruit availability available for our site. The use of this proxy is reasonable because, according to the optimal foraging theory, high-quality food items should increase in the diet as their abundance increases (Stephens and Krebs, 1986). And indeed, higher fruit consumption in times of higher fruit availability was shown for black-fronted titi monkeys, *Callicebus nigrifrons* (Caselli and Setz, 2011), as well as other primates, e.g., *Hoolock hoolock* (Neha et al., 2020).

Higher singing rates in months with higher fruit availability was also shown in *C. nigrifrons* (Caselli et al., 2014), *P. toppini* (Wright, 2013), in *Hylobates* gibbons (Cowlishaw, 1996), and in Milne Edwards' sportive lemurs (Méndez-Cárdenas and Zimmermann, 2009). More frequent singing during higher fruit availability is usually interpreted as a defense of valuable resources (Cowlishaw, 1996). Another possible explanation is that since singing is likely to be energetically costly (Cowlishaw, 1996; Wich and Nunn, 2002), animals sing less when less high-quality food is available. During months with lower fruit availability, groups of *P. toppini* not only sang less but also had shorter daily path lengths (DPL) (Wright, 2013). Shorter DPL during fruit scarcity was also shown in *C. nigrifrons* (Nagy-Reis and Setz, 2017) and in Coimbra-Filho's titi monkeys (*Callicebus*

TABLE 3 | Results of the model of the effects of song context, female reproductive state, and fruit proportion in the diet, on the duration of singing bouts: estimates, together with standard errors, confidence intervals, test results, and minimum and maximum of model estimates derived by dropping levels of random effects one at a time.

Term	Estimate	SE	Lower CI	Upper CI	*t*-value	min	max
Intercept	1.370	0.222	0.884	1.797	–	1.179	1.528
Context	0.676	0.171	0.316	0.989	3.940	0.618	0.732
Fruit proportion(1)	−0.095	0.093	−0.278	0.076	−1.027	−0.191	−0.003
Reproductive state (lactation)(2)	0.262	0.253	−0.203	0.794	1.037	0.039	0.333
Reproductive state (pregnancy)(2)	0.139	0.270	−0.374	0.703	0.517	−0.015	0.366

(1) *z-transformed to a mean of 0 and a standard deviation of 1; mean and SD of the original variable were 0.87 and 0.12, respectively.*
(2) *Reproductive state was dummy coded with receptivity being as the reference level, and comparisons here are with the reference level of receptivity.*

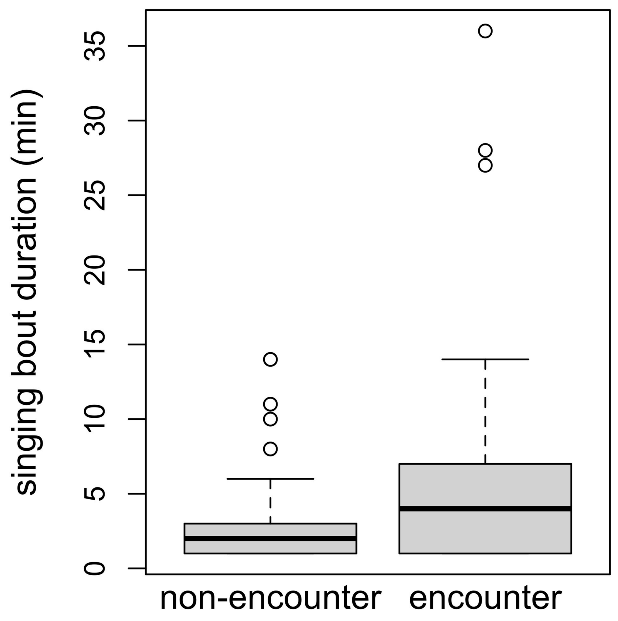


FIGURE 3 | Singing bout durations (in minutes) for encounter and non-encounter contexts. Boxes depict median and lower and upper quartiles.

coimbrai) (although only in a small forest fragment, whereas in a larger fragment this association was reversed (Souza-Alves et al., 2021). The hypothesis that singing is energetically costly is also indirectly supported by findings that gibbons sing less often at higher altitudes (i.e., at lower temperatures: (Cowlishaw, 1996), after cold nights and after rainy nights (*Hylobates klossii*: (Whitten, 1982). However, decreased singing rates after rainy nights might be also related to decreased sound transmission due to the background noise produced by dripping water rather than to energetic constraints. In this study, we showed that average monthly rainfall did not affect singing rates. We did not, however, record daily rainfall and temperature and our data was not sufficient to analyze the relationship between DPL and fruit consumption. To better understand the links between singing rates, fruits consumption, and rainfall, it will be necessary to estimate energetic costs of singing compared to other daily activities and relative to the energy input and environmental conditions.

While findings from our and other observational studies support the joint resource defense function of singing, they do not address a function of coordinated singing specifically. This issue can be disentangled by playback studies testing whether duets are more threatening to listeners than solo songs and whether partners are more likely to coordinate their songs into duets than to sing alone when faced with outsiders (Hall, 2004). In line with the latter prediction, wild *C. nigrifrons* and *P. ornatus* pairs consistently replied with duets to simulated intruders (although they did not react differently to playbacks of duets vs. solo male or female songs) (Robinson, 1981; Caselli et al., 2014). The coordinated nature of response to outsiders' songs is further supported by the behavior of young animals who join the adults to produce choruses both in our study population and in an observational and in a playback studies on *C. nigrifrons* (Caselli et al., 2014, 2015) and by coordinated behaviors displayed by captive *P. cupreus* in intruder tests (Mercier et al., 2020).

Mate Defense Hypothesis

Our findings did not support the mate defense hypothesis in the form of paternity guarding, as singing was not more frequent when females were estimated to be sexually receptive — in fact, singing was the least frequent during these periods. Similarly, in *C. nigrifrons*, groups did not sing more frequently when females were likely to be sexually receptive (Caselli et al., 2014). Moreover, predictions for this hypothesis are that duets are initiated by females (while males answer more of their partners' songs to form duets when females are receptive) (Hall, 2004). However, in *C. nigrifrons* duets were started either simultaneously by both individuals or with a short time difference (Caselli et al., 2014). Unfortunately, in our study, we were not able to identify the individual that initiated duetting (in *C. nigrifrons*, too, it was only possible using the spectrogram inspection of recorded songs).

However, paternity guarding is not the only form of mate defense, other forms being mate defense by either males or females *via* the defense of their own positions or their mates' positions within the partnership and commitment signaling, in which an individual prevents its partner from deserting (Hall, 2004, 2009). In these contexts, individuals are expected to show sex-specific responses to songs, and solo songs should be more threatening to listeners than duets (Hall, 2004). These predictions, impossible to test in an observational study, were addressed in a playback study in wild *C. nigrifrons* (Caselli et al., 2014). The study did not provide support for these forms of

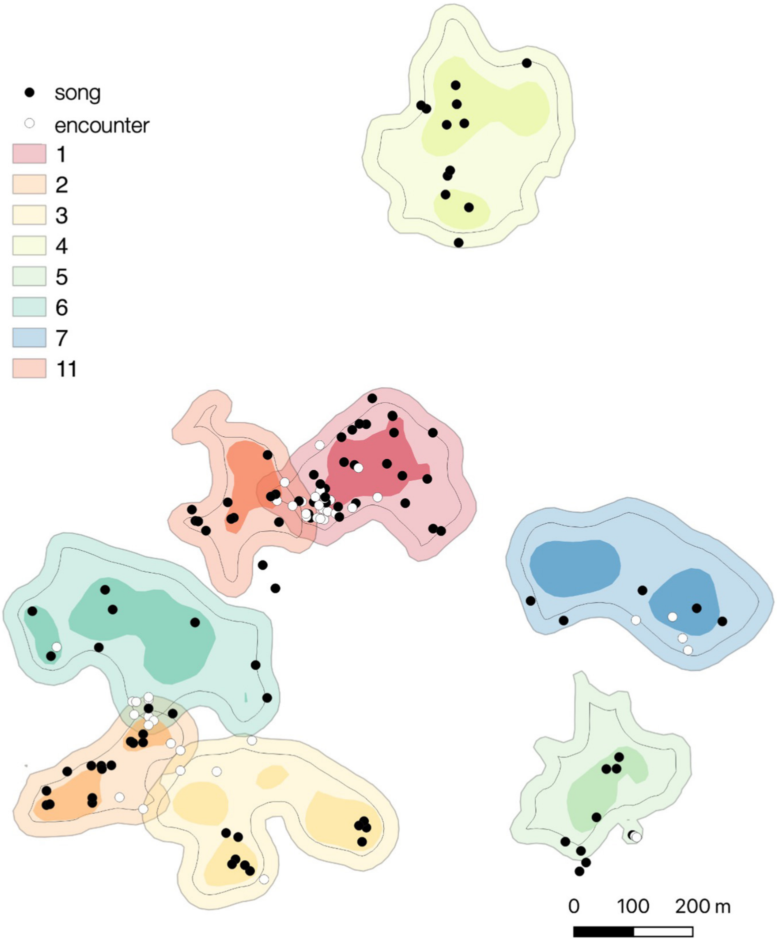


FIGURE 4 | Spatial distribution of songs produced in the intergroup encounter context (white dots) and non-intergroup encounter contexts (black dots) for eight study groups. Light areas depict the 95% fixed kernel home ranges, darker areas depict the 50% fixed kernel core areas. Black lines within each home range depict 25-m inner border areas.

mate defense: individuals did not show sex-specific responses to duets or male and female solos and did not react stronger to duets than to solo songs. However, in an earlier playback study on *P. ornatus*, reactions to solo playbacks were sex-specific: males initiated duetting more often in response to male solo song, while females initiated duetting more often in response to female solo song, possibly indicating both male and female mate defense (Robinson, 1981). Moreover, in *P. ornatus*, males often sang alone (Robinson, 1981). The differences between the two playback studies might be related to population characteristics. *P. ornatus* were studied in a much higher-density population than *C. nigrifrons*, which could have increased intrasexual competition and enhanced potential for extra-pair copulations (Caselli et al., 2014).

On the other hand, intruder tests with captive *P. cupreus* showed that males react more consistently to same-sex intruders than females and show more behavioral arousal to strangers compared to females (Cubicciotti and Mason, 1978; Fernandez-Duque et al., 2000, 1997). Similar sex differences were demonstrated in three wild titi monkey populations, including our study population, where males were more active in inter-group encounters (Robinson, 1981; Wright, 1984; Lawrence, 2007; Dolotovskaya et al., 2020). Whether this sex-specific defense results from conflicting male and female interests regarding male and female intruders, or to common benefits of division of labor (related, for example, to body size dimorphism) (Marshall-Ball et al., 2006), remains an open question. It should be noted, however, that at least in two titi monkey species, males are also more active in anti-predator behaviors (*P. cupreus*: Dolotovskaya et al., 2019; *P. discolor*: De Luna et al., 2010), suggesting that they might be generally more involved in defense of their territory and their group. In *P. toppini,* for example, males reacted stronger to playbacks in the high-use versus low-use parts of the home range (Lawrence, 2007), supporting the resource defense hypothesis and possibly indicating more active male involvement in resource defense.

Spatial and Temporal Distribution of Songs

Our study groups sang throughout their home ranges proportional to its use. The same pattern was observed in

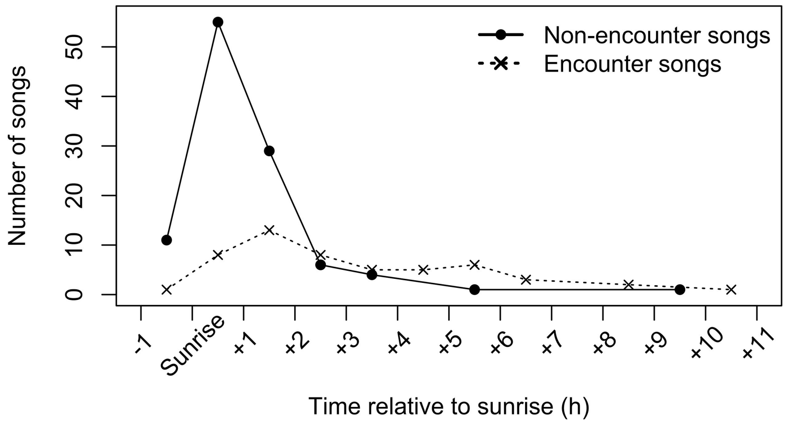


FIGURE 5 | Temporal distribution of songs. Shown are numbers of songs that started in each 1-h interval relative to sunrise time on a given day, separately for the encounter (total *N* = 52 songs) and non-encounter (total *N* = 107 songs) contexts, for eight study groups.

four other titi monkey species, *P. discolor* (Van Belle et al., 2021), *Plecturocebus modestus* and *Plecturocebus olallae* (Martinez and Wallace, 2017), and *Callicebus personatus* (Price and Piedade, 2001). This suggests that intergroup spacing mechanism of titi monkeys involves regular advertisement of the occupancy of the entire home range, as shown also in indris, *Indri indri* (Bonadonna et al., 2020) and black howler monkeys, *Alouatta caraya* (da Cunha and Byrne, 2006; Van Belle et al., 2013).

An alternative spacing mechanism involves signaling visitation to either border or core areas and was demonstrated, e.g., in brown howlers, *Alouatta guariba,* where howling was concentrated almost exclusively at the home range borders (Da Cunha and Jalles-Filho, 2007). Singing mostly from border areas of the home range was also shown in *P. ornatus* (Robinson, 1979b, 1981). Robinson (1979b) hypothesized that the spatial distribution of songs reflects spatial tensions between neighboring groups, where groups with small home ranges would engage in patrolling and singing at home range borders, while in groups with larger home ranges, border patrolling would be too energy-demanding and therefore, these groups would sing from more central areas. A preliminary study comparing titi monkey groups with different home range sizes (*P. ornatus* with smaller home ranges and *Cheracebus* cf. *lucifer* (*previously Callicebus torquatus*) with a larger home range) suggested that groups with smaller home ranges seem to sing and participate in intergroup encounters at the borders more often (Kinzey and Robinson, 1983). The study, however, included only one group with "large" home range and three groups with "small" home ranges. A larger sample will be needed to see whether titi groups with smaller home ranges are indeed more involved in patrolling behavior. It also remains to be studied whether songs of neighboring groups regulate movement decisions in listeners, as shown, for example, in black howlers (Van Belle and Estrada, 2020).

In our study, songs produced outside of the intergroup encounter context were concentrated around dawn. At this time of day, background noise in an Amazonian lowland forest is reduced, increasing communication distance (Ellinger et al., 2003). This further supports the notion that intergroup spacing mechanism of coppery titis involves regular advertisement of the occupancy of the entire home range rather than vocal border patrolling. Interestingly, in brown howlers, where howling is concentrated at home range borders, no dawn peak in howling was observed (Da Cunha and Jalles-Filho, 2007).

While songs produced during encounters did not show such a clear morning peak as songs produced outside of the encounters, around half of them were still produced within 3 h after the sunrise. Since most songs are produced around dawn, it is possible that neighboring groups are attracted to them and approach the borders of their home range to engage in intergroup encounters. It remains to be studied whether groups change their movement patterns in response to the songs of their neighbors. In this study, we did not distinguish between spontaneous songs and songs produced in response to neighbors' songs, because it was not always possible, while following a focal group, to determine unequivocally which of the neighboring groups was singing. Moreover, we cannot exclude the possibility that hearing distance might be higher for titis than for human observers and that a song which appears spontaneous to human observers might be in fact given in response to another group's song. Nevertheless, by following several neighboring groups in parallel, it should be possible in the future to study how groups react to their neighbors' singing.

Other Possible Functions of Songs: Pair-Bond Reinforcement?

Pair-bond reinforcement has been proposed as another possible function of coordinated singing, and in Milne Edwards' sportive lemur, pair mates were shown to synchronize their activities after duetting (Méndez-Cárdenas and Zimmermann, 2009). However, sportive lemurs live in dispersed pairs, foraging solitary but sleeping together, and duetting likely helps pair mates to localize each other and coordinate activities. Titi monkey pairs, on the other hand, are highly cohesive, and pair mates spend most of the day within a few meters from each other (Kinzey and Wright, 1982; Spence-Aizenberg et al., 2016; Dolotovskaya et al., 2020). Even if songs in titis function partly as a pair-bond reinforcing behavior, it is unlikely to be its main function. This is further supported by our finding that songs are concentrated around dawn and not produced randomly throughout the day, as could be expected if they were primarily serving for pair-bond reinforcement.

Conclusion

In sum, our findings, as well as results of other observational and playback studies, generally provide more support for the joint resource defense function of duetting in titi monkeys than for the mate defense function. However, there are several issues that still need to be explored by future studies. First, although pair mates in playback studies consistently replied with duets to simulated intruders, supporting the joint resource defense hypothesis, they did not react differently to playbacks of duets vs. solo songs as can be expected under this hypothesis. This issue can be disentangled by playback experiments comparing responses to duets with its temporal coordination artificially manipulated or by comparing playback of duets and playback of solo songs, using songs by the same two individuals. Second, playback studies that investigated the mate defense function of duets provided conflicting results: sex-specific responses were observed in *P. ornatus* but not in *C. nigrifrons*. To address this issue, future studies should evaluate the influence of population density on singing and on listeners' responses to it. And third, the more active male reaction to outsiders, observed both in wild and in captive titi monkeys, still needs to be explained within either joint resource or mate defense functions. To address this question, male and female vocal strategies (initiating song or joining a mate to form duets) in response to playback should be studied in more detail.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

This work was conducted under all necessary permissions and ethical guidelines from the relevant authorities of Peru (research permit no. 249-2017-SERFOR/DGGSPFFS from the Servicio Nacional Forestal y de Fauna Silvestre of the Peruvian Ministry of Agriculture) and the German Primate Center.

AUTHOR CONTRIBUTIONS

SD collected and analyzed the data and prepared the tables and figures. Both authors designed the research and wrote the manuscript.

FUNDING

This work was supported by the German Primate Center, the Leakey Foundation, the Deutsche Forschungsgemeinschaft (DFG) (grant HE 1870/29-1), the International Primatological Society, and the Primate Action Fund.

ACKNOWLEDGMENTS

We would like to thank Camilo Flores Amasifuén, Migdonio Huanuiri Arirama, Ney Shahuano Tello, Sarah Walker, Mathieu Maréchal, and all other assistants who worked with us in the field. We are also grateful to Ilya Filippov (Yugra State University, Russia) for his invaluable help with the spatial analyses. We thank the Servicio Forestal y de Fauna Silvestre (SERFOR) of the Ministry of Agriculture in Lima for the research permit (no. 249-2017-SERFOR/DGGSPFFS). We are also grateful to two reviewers and to the guest editor CS for their helpful feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.898509/full#supplementary-material

REFERENCES

Adret, P., Dingess, K., Caselli, C., Vermeer, J., Martínez, J., Luna Amancio, J., et al. (2018). Duetting patterns of titi monkeys (Primates, Pitheciidae: Callicebinae) and relationships with phylogeny. *Animals* 8:178. doi: 10.3390/ani8100178

Altmann, J., and Samuels, A. (1992). Costs of maternal care: infant-carrying in baboons. *Behav. Ecol. Sociobiol.* 29, 391–398. doi: 10.1007/BF00170168

Anzenberger, G. (1988). The pairbond in the titi monkey (*Callicebus moloch*): intrinsic versus extrinsic contributions of the pairmates. *Folia Primatol.* 50, 188–203.

Appleby, B. M., Yamaguchi, N., Johnson, P. J., and Macdonald, D. W. (1999). Sex-specific territorial responses in Tawny Owls *Strix aluco*. *Ibis* 141, 91–99. doi: 10.1111/j.1474-919x.1999.tb04267.x

Asensio, N., Schaffner, C. M., and Aureli, F. (2012). Variability in core areas of spider monkeys (*Ateles geoffroyi*) in a tropical dry forest in Costa Rica. *Primates* 53, 147–156. doi: 10.1007/S10329-011-0288-9/FIGURES/6

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Benedict, L. (2008). Occurrence and life history correlates of vocal duetting in North American passerines. *J. Avian Biol.* 39, 57–65. doi: 10.1111/j.2008.0908-8857.04103.x

Bicca-Marques, J. C., and Heymann, E. W. (2013). "Ecology and behavior of titi monkeys (genus *Callicebus*)," in *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, eds A. Barnett, L. M. Veiga, S. F. Ferrari, and M. A. Norconk (New York, NY: Cambridge University Press), 196–207.

Bonadonna, G., Zaccagno, M., Torti, V., Valente, D., De Gregorio, C., Randrianarison, R. M., et al. (2020). Intra- and intergroup spatial dynamics of a pair-living singing primate, *Indri indri*: a multiannual study of three indri groups in Maromizaha Forest, Madagascar. *Int. J. Primatol.* 41, 224–245. doi: 10.1007/s10764-019-00127-5

Brown, T. J., and Handford, P. (2003). Why birds sing at dawn: the role of consistent song transmission. *Ibis* 145, 120–129. doi: 10.1046/j.1474-919X.2003.00130.x

Calenge, C. (2006). The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. *Ecol. Modell.* 197, 516–519. doi: 10.1016/J.ECOLMODEL.2006.03.017

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., and Setz, E. Z. F. (2014). Vocal behavior of black-fronted titi monkeys (*Callicebus nigrifrons*): acoustic properties and behavioral contexts of loud calls. *Am. J. Primatol.* 76, 788–800. doi: 10.1002/ajp.22270

Caselli, C. B., Mennill, D. J., Gestich, C. C., Setz, E. Z. F., and Bicca-Marques, J. C. (2015). Playback responses of socially monogamous black-fronted titi monkeys to simulated solitary and paired intruders. *Am. J. Primatol.* 77, 1135–1142. doi: 10.1002/ajp.22447

Caselli, C. B., and Setz, E. Z. F. (2011). Feeding ecology and activity pattern of black-fronted titi monkeys (*Callicebus nigrifrons*) in a semideciduous tropical forest of southern Brazil. *Primates* 52, 351–359. doi: 10.1007/s10329-011-0266-2

Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour* 156, 1471–1494. doi: 10.1163/1568539X-00003575

Clink, D. J., Lau, A. R., Kanthaswamy, S., Johnson, L. M., and Bales, K. L. (2022). Moderate evidence for heritability in the duet contributions of a South American primate. *J. Evol. Biol.* 35, 51–63. doi: 10.1111/jeb.13962

Clutton-Brock, T. H., Albon, S. D., and Guinness, F. E. (1989). Fitness costs of gestation and lactation in wild mammals. *Nature* 337, 260–262. doi: 10.1038/337260a0

Corley, M. K., Perea-Rodriguez, J. P., Valeggia, C., and Fernandez-Duque, E. (2021). Associations between fecal cortisol and biparental care in a pair-living primate. *Am. J. Phys. Anthropol.* 176, 295–307. doi: 10.1002/ajpa.24368

Cowlishaw, G. (1996). Sexual selection and information content in gibbon song bouts. *Ethology* 102, 272–284. doi: 10.1111/j.1439-0310.1996.tb01125.x

Cubicciotti, D. D., and Mason, W. A. (1978). Comparative studies of social behavior in *Callicebus* and *Saimiri*: heterosexual jealousy behavior. *Behav. Ecol. Sociobiol.* 3, 311–322. doi: 10.1007/BF00296316

da Cunha, R. G. T., and Byrne, R. W. (2006). Roars of black howler monkeys (*Alouatta caraya*): evidence for a function in inter-group spacing. *Behaviour* 143, 1169–1199. doi: 10.1163/156853906778691568

Da Cunha, R. G. T., and Jalles-Filho, E. (2007). The roaring of southern brown howler monkeys (*Alouatta guariba* clamitans) as a mechanism of active defence of borders. *Folia Primatol.* 78, 259–271. doi: 10.1159/000105545

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

De Luna, A. G., Sanmiguel, R., Di Fiore, A., and Fernandez-Duque, E. (2010). Predation and predation attempts on red titi monkeys (*Callicebus discolor*) and equatorial sakis (*Pithecia aequatorialis*) in Amazonian Ecuador. *Folia Primatol.* 81, 86–95. doi: 10.1159/000314948

Dobson, A. J., Barnett, A. G., and Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*. New York, NY: Chapman and Hall. doi: 10.1201/9781584889519

Dolotovskaya, S., Flores Amasifuen, C., Haas, C. E., Nummert, F., and Heymann, E. W. (2019). Active anti-predator behaviour of red titi monkeys (*Plecturocebus cupreus*). *Primate Biol.* 6, 59–64. doi: 10.5194/pb-6-59-2019

Dolotovskaya, S., Walker, S., and Heymann, E. W. (2020). What makes a pair bond in a Neotropical primate: female and male contributions. *R. Soc. Open Sci.* 7:191489. doi: 10.1098/rsos.191489

Dolotovskaya, S., and Heymann, E. W. (2020). Do less or eat more: strategies to cope with costs of parental care in a pair-living monkey. *Anim. Behav.* 163, 163–173. doi: 10.1016/j.anbehav.2020.03.012

Ellinger, N., Hödl, W., and Ellinger, N. (2003). Habitat acoustics of a neotropical lowland rainforest. *Bioacoustics* 13, 297–321. doi: 10.1080/09524622.2003.9753503

Erran Seaman, D., and Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology* 77, 2075–2085. doi: 10.2307/2265701

Fernandez-Duque, E., Di Fiore, A., and de Luna, A. G. (2013). "Pair-mate relationships and parenting in equatorial saki monkeys (*Pithecia aequatorialis*) and red titi monkeys (*Callicebus discolor*) of ecuador," in *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, eds L. M. Veiga, A. A. Barnett, S. F. Ferrari, and M. A. Norconk (New York, NY: Cambridge University Press), 295–302.

Fernandez-Duque, E., Huck, M., Van Belle, S., and Di Fiore, A. (2020). The evolution of pair-living, sexual monogamy, and cooperative infant care: insights from research on wild owl monkeys, titi monkeys, sakis, and tamarins. *Am. J. Phys. Anthropol.* 171, 118–173. doi: 10.1111/j.1365-2958.2003.03935.x

Fernandez-Duque, E., Mason, W. A., and Mendoza, S. P. (1997). Effects of duration of separation on responses to mates and strangers in the monogamous titi monkey (*Callicebus moloch*). *Am. J. Primatol.* 43, 225–237.

Fernandez-Duque, E., Valeggia, C. R., and Mason, W. A. (2000). Effects of pair-bond and social context on male-female interactions in captive titi monkeys (*Callicebus moloch*, Primates: Cebidae). *Ethology* 106, 1067–1082. doi: 10.1046/j.1439-0310.2000.00629.x

Forstmeier, W., and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* 65, 47–55. doi: 10.1007/s00265-010-1038-5

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage.

Gittleman, J. L., and Thompson, S. D. (1988). Energy allocation in mammalian reproduction. *Integr. Comp. Biol.* 28, 863–875. doi: 10.1093/icb/28.3.863

Haimoff, E. H. (1986). Convergence in the duetting of monogamous old world primates. *J. Hum. Evol.* 15, 51–59. doi: 10.1016/S0047-2484(86)80065-3

Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behav. Ecol. Sociobiol.* 55, 415–430. doi: 10.1007/s00265-003-0741-x

Hall, M. L. (2009). A review of vocal duetting in birds. *Adv. Study Behav.* 40, 67–121. doi: 10.1016/S0065-3454(09)40003-2

Heymann, E. W., Dolotovskaya, S., and Herrera, E. R. T. (2021). Estación Biológica Quebrada Blanco. *Ecotropica* 23, 4–7. doi: 10.30427/ecotrop202101

Holzmann, I., Agostini, I., and Di Bitetti, M. (2012). Roaring behavior of two syntopic howler species (*Alouatta caraya* and *A. guariba clamitans*): evidence supports the mate defense hypothesis. *Int. J. Primatol.* 33, 338–355. doi: 10.1007/s10764-012-9583-6

Kinzey, W. G. (1981). "The titi monkeys, genus *Callicebus*," in *Ecology and Behavior of Neotropical Primates*, eds A. F. Coimbra-Filho and R. A. Mittermeier (Rio de Janeiro: Academia Brasileira de Ciencias), 241–276.

Kinzey, W. G., and Robinson, J. G. (1983). Intergroup loud calls, range size, and spacing in *Callicebus torquatus*. *Am. J. Phys. Anthropol.* 60, 539–544.

Kinzey, W. G., and Wright, P. C. (1982). Grooming behavior in the titi monkey (*Callicebus torquatus*). *Am. J. Primatol.* 3, 267–275. doi: 10.1002/ajp.1350030124

Lau, A., Clink, D. J., and Bales, K. L. (2020). Individuality in the vocalizations of adult and infant coppery titi monkeys (*Plecturocebus cupreus*). *J. Acoust. Soc. Am.* 82:e23134. doi: 10.1121/1.5136592

Lawrence, J. (2007). *Understanding the Pair Bond in Brown Titi Monkeys (Callicebus brunneus): Male and Female Reproductive Interests*. Ph.D. thesis. New York, NY: Columbia University.

Levin, R. N. (1996). Song behaviour and reproductive strategies in a duetting wren, *Thryothorus nigricapillus*: II. Playback experiments. *Anim. Behav.* 52, 1107–1117. doi: 10.1006/ANBE.1996.0258

Logue, D. (2005). Cooperative defence in duet singing birds. *Cogn. Brain Behav.* 9:e510.

Malacarne, G., Cucco, M., and Camanni, S. (1991). Coordinated visual displays and vocal duetting in different ecological situations among western palearctic non-passerine birds. *Ethol. Ecol. Evol.* 3, 207–219. doi: 10.1080/08927014.1991.9525369

Marshall-Ball, L., Mann, N., and Slater, P. J. B. (2006). Multiple functions to duet singing: hidden conflicts and apparent cooperation. *Anim. Behav.* 71, 823–831. doi: 10.1016/J.ANBEHAV.2005.05.021

Martinez, J., and Wallace, R. B. (2017). Ecological and behavioural factors influencing territorial call rates for the bolivian titi monkeys, *Plecturocebus modestus* and *Plecturocebus olallae*. *Folia Primatol.* 87, 279–290. doi: 10.1159/000448710

McCabe, G. M., and Fedigan, L. M. (2007). Effects of reproductive status on energy intake, ingestion rates, and dietary composition of female *Cebus capucinus* at Santa Rosa, Costa Rica. *Int. J. Primatol.* 28, 837–851. doi: 10.1007/s10764-007-9159-z

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting — a mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi*)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Mennill, D. J., and Vehrencamp, S. L. (2008). Context-dependent functions of avian duets revealed by microphone-array recordings and multispeaker playback. *Curr. Biol.* 18, 1314–1319. doi: 10.1016/J.CUB.2008.07.073

Mercier, F., Witczak, L. R., and Bales, K. L. (2020). Coppery titi monkey (*Plecturocebus cupreus*) pairs display coordinated behaviors in response to a simulated intruder. *Am. J. Primatol.* 82:e23141. doi: 10.1002/ajp.23141

Merker, B., and Cox, C. (1999). Development of the female great call in *Hylobates gabriellae*: a case study. *Folia Primatol.* 70, 97–106. doi: 10.1159/000021680

Müller, A. E., and Anzenberger, G. (2002). Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. doi: 10.1159/000064788

Murray, C. M., Lonsdorf, E. V., Eberly, L. E., and Pusey, A. E. (2009). Reproductive energetics in free-living female chimpanzees (*Pan troglodytes schweinfurthii*). *Behav. Ecol.* 20, 1211–1216. doi: 10.1093/beheco/arp114

Nagy-Reis, M. B., and Setz, E. Z. F. (2017). Foraging strategies of black-fronted titi monkeys (*Callicebus nigrifrons*) in relation to food availability in a seasonal tropical forest. *Primates* 58, 149–158. doi: 10.1007/s10329-016-0556-9

Neha, S. A., Habiba Khatun, M. U., and Ul Hasan, M. A. (2020). Feeding behavior of the western hoolock gibbon (*Hoolock hoolock*) in Bangladesh: response to temporal variation of food sources. *Primate Conserv.* 2020, 185–194.

Nietsch, A. (1999). Duet vocalizations among different populations of Sulawesi tarsiers. *Int. J. Primatol.* 204, 567–583. doi: 10.1023/A:1020342807709

Paula, B., and Monticelli, P. (2021). Maned wolf duet? The first record of two maned wolves vocalizing simultaneous. *Acad. Lett.* doi: 10.20935/al1720

Price, E. C., and Piedade, H. M. (2001). Ranging behavior and intraspecific relationships of masked titi monkeys (*Callicebus personatus personatus*). *Am. J. Primatol.* 53, 87–92.

QGIS Development Team (2022). *QGIS Geographic Information System. Open Source Geospatial Foundation Project*. Available Online at: http://qgis.osgeo.org (accessed January 11, 2022)

Quinn, G. P., and Keough, M. J. (2002). *Experimental Designs and Data Analysis for Biologists*. Cambridge: Cambridge University Press.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rasoloharijaona, S., Randrianambinina, B., Braune, P., and Zimmermann, E. (2006). Loud calling, spacing, and cohesiveness in a nocturnal primate, the Milne Edwards' sportive lemur (*Lepilemur edwardsi*). *Am. J. Phys. Anthropol.* 129, 591–600. doi: 10.1002/ajpa

Robinson, J. G. (1979a). An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Z. Tierpsychol.* 49, 381–405.

Robinson, J. G. (1979b). Vocal regulation of use of space by groups of titi monkeys *Callicebus moloch*. *Behav. Ecol. Sociobiol.* 5, 1–15. doi: 10.1007/BF00302691

Robinson, J. G. (1981). Vocal regulation of inter- and intragroup spacing during boundary encounters in the titi monkey, *Callicebus moloch*. *Primates* 22, 161–172. doi: 10.1007/BF02382607

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* 1, 103–113. doi: 10.1111/j.2041-210X.2010.00012.x

Snow, D. W. (1977). "Duetting and other synchronised displays of the blue-backed manakins, *Chiroxiphia spp*.," in *Evolutionary Ecology*, eds B. Stonehouse and C. Perrins (London: Palgrave), 239–250 doi: 10.1007/978-1-349-05226-4_20

Souza-Alves, J. P., Chagas, R. R. D., Santana, M. M., Boyle, S. A., and Bezerra, B. M. (2021). Food availability, plant diversity, and vegetation structure drive behavioral and ecological variation in Endangered Coimbra-Filho's titi monkeys. *Am. J. Primatol.* 83:e23237. doi: 10.1002/ajp.23237

Spence-Aizenberg, A., Di Fiore, A., and Fernandez-Duque, E. (2016). Social monogamy, male–female relationships, and biparental care in wild titi monkeys (*Callicebus discolor*). *Primates* 57, 103–112. doi: 10.1007/s10329-015-0489-8

Stephens, D. W., and Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Tilson, R. L., and Norton, P. M. (1981). Alarm duetting and pursuit deterrence in an African antelope. *Am. Nat.* 118, 455–462.

Tilson, R. L., and Tenaza, R. R. (1976). Monogamy and duetting in an old world monkey. *Nature* 263, 320–321. doi: 10.1038/263320a0

Tirado Herrera, E. R., and Heymann, E. W. (2004). Does mom need more protein? Preliminary observations on differences in diet composition in a pair of red titi monkeys (*Callicebus cupreus*). *Folia Primatol.* 75, 150–153. doi: 10.1159/000078304

Valeggia, C. R., Mendoza, S. P., Fernandez-Duque, E., Mason, W. A., and Lasley, B. (1999). Reproductive biology of female titi monkeys (*Callicebus moloch*) in captivity. *Am. J. Primatol.* 47, 183–195.

Van Belle, S., and Estrada, A. (2020). The influence of loud calls on intergroup spacing mechanism in black howler monkeys (*Alouatta pigra*). *Int. J. Primatol.* 41, 265–286. doi: 10.1007/s10764-019-00121-x

Van Belle, S., Estrada, A., and Garber, P. A. (2013). Spatial and diurnal distribution of loud calling in black howlers (*Alouatta pigra*). *Int. J. Primatol.* 34, 1209–1224. doi: 10.1007/s10764-013-9734-4

Van Belle, S., Fernandez-Duque, E., and Di Fiore, A. (2016). Demography and life history of wild red titi monkeys (*Callicebus discolor*) and equatorial sakis (*Pithecia aequatorialis*) in Amazonian Ecuador: a 12-year study. *Am. J. Primatol.* 78, 204–215. doi: 10.1002/ajp.22493

Van Belle, S., Porter, A., Fernandez-Duque, E., and Di Fiore, A. (2021). Ranging behavior and the potential for territoriality in pair-living, sexually monogamous titi monkeys (*Plecturocebus discolor*). *Am. J. Primatol.* 83:e23225. doi: 10.1002/ajp.23225

Vasey, N. (2004). Circadian rhythms in diet and habitat use in red ruffed lemurs (*Varecia rubra*) and white-fronted brown lemurs (*Eulemur fulvus albifrons*). *Am. J. Phys. Anthropol.* 124, 353–363. doi: 10.1002/ajpa.10357

Vasey, N. (2005). Activity budgets and activity rhythms in red ruffed lemurs (*Varecia rubra*) on the Masoala Peninsula, Madagascar: seasonality and reproductive energetics. *Am. J. Primatol.* 66, 23–44. doi: 10.1002/ajp.20126

Whitten, A. J. (1982). The ecology of singing in Kloss gibbons (*Hylobates klossii*) on Siberut Island, Indonesia. *Int. J. Primatol.* 3, 33–51. doi: 10.1007/BF02693489

Wich, S. A., and Nunn, C. L. (2002). Do male "long-distance calls" function in mate defense? A comparative study of long-distance calls in primates. *Behav. Ecol. Sociobiol.* 52, 474–484. doi: 10.1007/s00265-002-0541-8

Wright, P. C. (1984). "Biparental care in *Aotus trivirgatus* and *Callicebus moloch*," in *Female Primates: Studies by Women Primatologists*, ed. M. F. Small (New York, NY: Alan R. Liss), 59–75.

Wright, P. C. (2013). "*Callicebus* in Manu National Park: territory, resources, scent marking and vocalizations," in *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, eds L. M. Veiga, A. A. Barnett, S. F. Ferrari, and M. A. Norconk (Cambridge: Cambridge University Press), 232–239.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Ecology and Evolution

PERSPECTIVE
published: 07 July 2022
doi: 10.3389/fevo.2022.906546

Perspectives on Antiphonal Calling, Duetting and Counter-Singing in Non-primate Mammals: An Overview With Notes on the Coordinated Vocalizations of Bamboo Rats (*Dactylomys* spp., Rodentia: Echimyidae)

E. Natasha Vanderhoff[1]* and N. Bernal Hoverud[2,3]

[1] Department of Biology and Marine Science, Jacksonville University, Jacksonville, FL, United States, [2] Graduate Program, Department of Biological Sciences, Texas Tech University, Lubbock, TX, United States, [3] Wildlife Conservation Society, Bolivia Program, La Paz, Bolivia

Edited by:
Patrice Adret,
Universidad Autónoma Gabriel René Moreno, Bolivia

Reviewed by:
Vlad Demartsev,
University of Konstanz, Germany
Charles T. Snowdon,
University of Wisconsin-Madison, United States

***Correspondence:**
E. Natasha Vanderhoff
nvander4@ju.edu

Specialty section:
This article was submitted to Behavioral and Evolutionary Ecology, a section of the journal Frontiers in Ecology and Evolution

Received: 28 March 2022
Accepted: 15 June 2022
Published: 07 July 2022

Citation:
Vanderhoff EN and Bernal Hoverud N (2022) Perspectives on Antiphonal Calling, Duetting and Counter-Singing in Non-primate Mammals: An Overview With Notes on the Coordinated Vocalizations of Bamboo Rats (Dactylomys spp., Rodentia: Echimyidae). Front. Ecol. Evol. 10:906546. doi: 10.3389/fevo.2022.906546

Temporally coordinated interactive vocalizations are important means of communication between individuals in various animal taxa. In mammals, interactive calling and singing can be highly synchronized to create either overlapping or antiphonal duets while in others, competitors antagonistically vocalize, engaging in counter-singing. Among non-primate mammals these vocalizations are considered rare and poorly understood. We provide an overview of antiphonal calling, duetting and counter-singing in non-primate mammals. Many of these coordinated vocalizations play a role in social interactions and allow mammals to convey information to other members of the social unit in visually inaccessible environments. South American Bamboo rats *Dactylomys* spp. are arboreal bamboo specialists found in dense bamboo thickets in Bolivia, Peru, Ecuador, Brazil and Colombia. These nocturnal rodents are rarely seen but can be easily heard because of their loud and distinctive staccato vocalizations. We provide some evidence that Bamboo rats engage in duetting, and as such they provide another case of a mammalian species, in which to investigate temporally coordinated interactive singing. We urge researchers to work toward common definitions of temporally coordinated vocalizations and to search for more mammals that utilize such vocalizations.

Keywords: antiphonal vocalizations, mammal duets, counter-singing, *Dactylomys* spp., temporally coordinated interactive vocalizations

INTRODUCTION

Mammals produce a diverse array of vocalizations ranging from single note contact calls to highly synchronized, multiple note songs (Fitch, 2006). Whether simple or complex, mammal vocalizations convey vital information to conspecifics. While calls, such as contact and alarm calls, are short with few notes and function in a variety of daily activities songs are typically more complex consisting of several notes, and used in mating and territorial displays (Boughman and Moss, 2003; Marler, 2004). Interactive vocal communication has evolved in several mammalian groups, including primates, rodents, bats, and cetaceans (Logue and Krupp, 2016; Vernes, 2017; Terleph et al., 2018; de Reus et al., 2021; De Gregorio et al., 2022). Antiphonal vocalizations, counter-singing, and duetting have been thoroughly documented and studied in birds (for a review of duetting in birds see Hall, 2009), but less so in mammals, with most of the attention given to primates (Adret et al., 2018; Clink and Lau, 2020). Broadening research to include mammals other than primates will allow us to test hypotheses on the function and evolution of these vocalizations.

Across all taxa, interactive vocalizations and turn-taking (Levinson, 2016), have received more attention in recent years, with researchers trying to assess appropriate frameworks to study and analyze interactive vocalizations. The disparity in definitions of these types of vocalizations makes comparative studies difficult, with some researchers suggesting a focus on the stepwise process between individuals involved in back-and-forth communication (who, when and how does turn-taking occur), whereas others suggest a focus on the temporal and rhythmic elements of vocalizations (Pika et al., 2018; Ravignani et al., 2019; de Reus et al., 2021). We suggest and utilize the term *temporally coordinated interactive vocalizations* (TCIVs hereafter) to encompass both the dynamics between signaling mammals as well as the rhythmic and temporal component of the vocalizations themselves. We provide an overview of the literature on non-primate mammal vocalizations that fall under this umbrella. Across the literature different terms are utilized and we adhere to the terms utilized by individual researchers. We classify TCIVs into three categories (antiphonal vocalizations, duets, and counter-vocalizations). These categories are based on the literature we examined but we realize that some overlap may occur and may not reflect current definitions. Antiphonal vocalizations are the broadest category and involve call and response occurring at regular intervals between two or more individuals (Yoshida and Okanoya, 2005; Filippi et al., 2019). Counter-vocalizations occur between specific individuals and territorial mammals may engage in counter-calling (or counter-singing), in which rival individuals call (or sing) back and forth in a non-overlapping fashion (Banerjee et al., 2019). These types of vocalizations may or may not involve turn-taking in which individuals adjust vocalizations based on the behavior of the other participant, including overlap avoidance (Demartsev et al., 2018; Pika et al., 2018; Okobi et al., 2019). Duetting has been examined extensively in birds and early definitions drawn from the avian literature frequently described duetting as coordinated vocalizations in mated pairs (Todt and Naguib, 2000; Hall, 2009). Duets may or may not overlap, and current definitions of duetting vary (see Pika et al., 2018). As noted by Langmore (2002), a more accurate definition should focus on acoustic features of duets and not just the participants. Our definition of duetting follows the literature that characterizes duets as coordinated, predictable, repetitive, stereotypical vocalizations between two or more individuals, often bonded individuals (Langmore, 2002; Clink et al., 2020; Nieder and Mooney, 2020).

ANTIPHONAL VOCALIZATIONS, DUETS AND COUNTER-VOCALIZATIONS IN NON-PRIMATE MAMMALS

Mammals from diverse orders, with varying habitats, activity periods, and social and mating systems engage in TCIVs (**Table 1**). The functions of these calls vary greatly, and categories were selected *post-hoc* after examining the literature. Researchers used behavioral data, including temporal and spatial aspects of vocalizations and individuals, and playback experiments to elucidate function. Below is a brief discussion of the different vocalizations and their functions in non-primate mammals.

Antiphonal Vocalizations

In mammals, antiphonal vocalizations (or the use of the term antiphonal) is more prevalent in the literature than duetting and counter-singing, and antiphonal calls serve as contact calls that can encode information about individual identity and condition. For example, naked-mole rats *Heterocephalus glaber* can identify individuals as well as social rank within their colony (Yosida and Okanoya, 2009). In group living mammals, antiphonal vocalizations can also help individuals coordinate with conspecifics. White-winged vampire bats *Diaemus youngi* can discern individuals and thus their spatial relationship in the colony *via* antiphonal calling (Carter et al., 2008). African elephants *Loxodonta africana* use antiphonal rumbles to monitor individuals and coordinate group movement, with antiphonal calling increasing as elephants disperse away from one another (Leighty et al., 2008; O'Connell-Rodwell et al., 2012). Moreover, female elephants exchange rumbles more with close social partners (Soltis et al., 2005). Similarly, pulsed "type A" calls of beluga whales *Delphinapterus leucas* enable group cohesion and closely related individuals use similar variants of these calls (Vergara et al., 2010). Other examples of antiphonal calls between related individual include calls between mother and offspring in belugas, as well as in several bat species (bulldog bat *Noctilio albiventris*, sac-winged bat *Saccopteryx bilineata*, Pomona leaf-nosed bat *Hipposideros pomona*) and the Florida manatee *Trichechus manatus latirostris* (Brown et al., 1983; O'Shea and Poché, 2006; Knörnschild and Von Helversen, 2008; Jin et al., 2015).

While antiphonal calls function largely to facilitate social interactions among conspecifics, in Artiodactyls antiphonal calls function as important pursuit deterrent signals to predators. The antiphonal calls of Indian muntjac *Muntiacus muntjak* and Roe deer *Capreolus capreolus* intensify alarm signals, confusing predators as multiple individuals call and respond from different locations (Oli and Jacobson, 1995; Rossi et al., 2002). Klipspringer *Oreotragus oreotragus* elicit short vocal exchanges, females following males swiftly (30 ms intercal interval) in loud and repetitive alarm calls (Tilson and Norton, 1981). While these calls have been traditionally seen as alarm calls, they could also aid in pair and group coordination.

TABLE 1 | Examples of non-primate mammals engaging in temporally coordinated interactive vocal communication.

		Type of interactive vocal communication:								
Scientific name	**Common name**	**Counter-singing/Calling**	**Duetting**	**Antiphonal**	**Habitat**	**Activity period**	**Social system**	**Mating system**	**Function**	**Source**
Rodentia										
*Scotinomys teguina**	Alston's singing mouse	X^{t-t}			t/tr	n	Solitary	Promiscuous	6, 8	Banerjee et al., 2019 Okobi et al., 2019
Heterocephalus glaber	Naked mole-rat			x	t/st	n	Eusocial	Eusocial	3,5	Yosida and Okanoya, 2009, 2012
Chiroptera										
Diaemus youngi	White-winged vampire bat			x	t/tr	n	Colonial	Social polygyny	2,3,4	Carter et al., 2008, 2012
Noctilio albiventris	Bulldog bat			x	t/tr	n	Colonial	Social polygyny	4	Brown et al., 1983
*Saccopteryx bilineata**	Sac-winged bat	x		x	t/tr	n	Colonial	Social polygyny	4,8	Knörnschild and Von Helversen, 2008 Behr et al., 2009
Hipposideros pomona	Pomona leaf-nosed bat			x	t/tr	n	Colonial	Social polygyny	4	Jin et al., 2015
Proboscidea										
Loxodonta africana	African elephant			x	t/sv	d	Fission-fusion	Promiscuous	2,6,7	Soltis et al., 2005 Leighty et al., 2008 O'Connell-Rodwell et al., 2012
Hyracoidea										
*Procavia capensis**	Rock hyrax	x			t	d	Male/multi-female	Social polygyny	5,8	Demartsev et al., 2016a,b, 2017
*Dendrohyrax sp**	Tree hyrax			x	t/tr	n	Solitary	Monogamy	8	Rosti et al., 2020
Artiodactyla										
Oreotragus oreotragus	Klipspringer		x		t/rk	d	Pairs w/offspring	Monogamous	1	Tilson and Norton, 1981
Muntiacus muntjak	Muntjak/Barking deer			x	t/tr	d	Solitary	Polygynous	1	Oli and Jacobson, 1995
Capreolus capreolus	Roe deer			x	t/f	d	Solitary	Monogamous	1	Rossi et al., 2002
Cetacea										
*Physeter macrocephalus**	Sperm whale		x		aq/m	d	Multi-level/matrilineal	Promiscuous	7	Schulz et al., 2008
*Delphinapterus leucas**	Beluga whale			x	aq/m	d	Highly varied	Promiscuous	2,4	Vergara et al., 2010
Orcinus orca	Killer whale			x	aq/m	d	Multi-level/matrilineal	Promiscuous	2,6	Miller et al., 2004
Globicephala melas	Long-finned pilot whale		x		aq/m	d	Fission-fussion	Polygynous	unk	Courts et al., 2020
Sirenia										
	Florida manatee			x	aq/fm	d	Solitary	Promiscuous	3, 4	O'Shea and Poché, 2006

(Continued)

TABLE 1 | (Continued)

Scientific name	Common name	Type of interactive vocal communication:			Habitat	Activity period	Social system	Mating system	Function	Source
		Counter-singing/Calling	Duetting	Antiphonal						
Rodentia										
Carnivora										
Chrysocyon brachyurus	Maned wolf		x	x	t/sv	n/crp	Pairs	Monogamous	2, 3, 5, 6,7, 8	Balieiro and Monticelli, 2019 Paula and Monticelli, 2021; Ferreira et al., 2022
Suricata suricatta	Meerkat	x^{t-t}			t/sv	d	Matriarchy	Monogamous	7	Demartsev et al., 2018
Scandentia										
Tupaia palawanensis	Palawan tree shrew			x	t/tr	d	Pairs w/offspring	Monogamous	2	Williams et al., 1969

Habitat: terrestrial (t) or aquatic (aq)/tropical (tr), subterranean (st), savannah (sv), forest (f), marine (m). Activity period: nocturnal (n), diurnal (d) or crepuscular (crp). Function of vocalizations: (1) alarm signal, (2) contact call/group communication, (3) individual recognition, (4) mother-offspring communication, (5) social dominance/individual quality, (6) group/pair coordination, (7) strengthen social bonds, (8) territorial. Denotations with a x^{t-t} indicate counter-calling or singing with turn-taking behavior. Scientific names with an asterisk highlight singing mammals.

Duets

Unlike the call and response of antiphonal vocalizations, duetting requires two individuals to coordinate vocalizations in a precisely timed manner (Langmore, 2002; Nieder and Mooney, 2020; de Reus et al., 2021). Duets begin with one individual vocalizing, followed by another individual vocalizing, sometimes simultaneously with the first. Duetting by non-primate mammals appears to be rarer in the literature. Within social units of sperm whale. *Physeter macrocephalus*, individual whales will produce codas after another whale begins to sing and in some instances the codas will overlap (Schulz et al., 2008). As with many antiphonal calls, duets in sperm whales help to maintain social bonds. Long finned pilot whales (*Globicephala melas*) also produce vocal duets, yet their function is unknown (Courts et al., 2020). The duets of maned wolves *Chrysocyon brachyurus* also have a social function. Maned wolves live in dispersed pairs that actively defend a territory and their individually identifiable roar-barks produced as loose duets (2 s inter-call intervals on average) may help them facilitate and strengthen pair bonds necessary for this defense (Emmons, 2013; Balieiro and Monticelli, 2019; Paula and Monticelli, 2021). Recent work with maned wolves describe their interactive roar-barks as counter-calling, not duetting. This behavior is most frequent during the mating season and when young are present, suggesting a function in not only pair bonding, but also in care of the young (Ferreira et al., 2022). The function of duetting is still poorly understood and debated for many taxa and as seen in birds it may serve more than one function (Hall, 2009).

Counter-Vocalizations and Turn-Taking

Unlike duetting that involves coordination between members of the same social group, counter-vocalizations involve back and forth vocalizations, often between territorial rivals. Many mammals sing in a territorial fashion and counter-singing has been documented in sac-winged bats *Saccopteryx bilineata*, singing mice *Scotinomys teguina* and rock hyraxes *Procavia capensis* (Behr et al., 2009; Demartsev et al., 2016a,b, Demartsev et al., 2017; Banerjee et al., 2019; Okobi et al., 2019). As with other coordinated vocalizations, they can contain information about caller identity; however, these vocal duels have only been documented in males and function in sexual advertisement and territorial interactions. The singing behavior of rock hyraxes has been well documented, and 25% of all singing is male-male counter singing (Demartsev et al., 2016b). Tree hyraxes *Dendrohyrax* sp., unlike rock hyraxes, are solitary and monogamous, and although they sing, their coordinated communication occurs in antiphonal territorial calls. In a recent study of tree hyrax vocalizations, 75% of *thwack* call sequences involved counter-calling between two or more individuals (Rosti et al., 2020). It is important to note that these two closely related, yet socially and ecologically distinct species, both have coordinated vocalizations, one in song and the other in calls. Future comparative studies of hyrax species, as well as across and within mammalian groups in general, could lead to insights on the evolution of coordinated vocalizations.

Turn-taking behavior does not occur in all instances of counter-singing, but as documented in singing mice, individuals mediate their response to other individuals, regulating their vocalizations and altering their behavior in response (Banerjee et al., 2019; Okobi et al., 2019). Similarly, meerkats engage in turn-taking "sunning vocalizations" when several group members sun individual meerkat calls avoid overlap with other vocalizing meerkats (Demartsev et al., 2018).

In addition to clearer definitions of TCIVs, researchers must identify which mammals engage in this behavior and how it affects each vocalizing individual. While duetting and other coordinated vocalizations have been traditionally viewed and examined as a collective behavior, they occur at both the level of the individual and the collective pair or group (Hall, 2004; Logue and Krupp, 2016; Ravignani et al., 2019; Clink et al., 2020; Clink and Lau, 2020). It is likely that many more mammals engage in TCIVs than is currently recognized. Next, we present observational data of bamboo rats *Dactylomys* spp., a neotropical bamboo specialist, from Ecuador and Bolivia as an example of an understudied mammal that engages in TCIVs.

PERSPECTIVES FROM BAMBOO RAT VOCALIZATIONS

South American bamboo rats *Dactylomys* spp. are nocturnal, arboreal bamboo specialists found in dense bamboo thickets in Bolivia, Peru, Ecuador, Brazil, and Colombia (La Val, 1976; Emmons, 1981; Dunnum and Salazar-Bravo, 2004; Bezerra et al., 2007). *Dactylomys* spp. are rarely seen but can be easily heard due to their loud and distinctive staccato vocalizations that start at dusk (La Val, 1976; Emmons, 1981). Vocalizations are thought to be territorial calls; however, the spatial and social relationship between individuals within and between bamboo patches has not been thoroughly investigated. Little is known about this secretive species due to its nocturnal cryptic behavior and the dense bamboo thickets in which they inhabit, yet *Dactylomys* spp. might provide a good comparative model to study TCIVs. Emmons (1981) first described duetting behavior in the Amazonian bamboo rat *Dactylomys dactylinus* and noted two distinct call types: loud staccato "L calls" given by males, often followed by softer grunting "A calls" given by females. She also noted call and responses between males producing the loud "L calls."

Our preliminary investigations of *D. dactylinus* in Ecuador and *D. boliviensis* in Bolivia provide more evidence for TCIVs. In July 2010 and July 2011, ENV surveyed *D. dactylinus* populations and recorded their vocalizations at Wildsumaco Wildlife Sanctuary in Ecuador, (00° 41.250' S, 77° 36.049' W; ∼1400 m elevation). Between 2015 and 2017, as part of the "Identidad Madidi Project" led by Wildlife Conservation Society in Bolivia, NBH and her team observed and recorded *D. boliviensis* at five sites inside Madidi National Park (14.1892° S, 68.3339° W; 200–1700 m elevation). At both Wildsumaco and Madidi, bamboo rats were heard and seen in bamboo patches (*Guadua* spp). Ten bamboo patches, ranging in size from 25 to 500 m^2, often consisting of several clusters of bamboo within a matrix of other vegetation, were identified at Wildsumaco. Some patches were relatively close to one another (∼15 m), but could be as far apart as 100–200 m. At Madidi, bamboo patches could also have different extensions and were generally localized in wet soils along running water, but their distribution and size was not quantified as it was part of a larger survey. Identification of individuals on most nights was difficult given the dense vegetation and the rats cryptic behavior, as Emmons (1981) notes they move silently one foot at a time making vocalizations the only means to identify if individuals were present. At both Wildsumaco and Madidi loud staccato calls and soft grunting calls, referred to as "L" and "A" calls by Emmons (1981), occurred during the night, between 19:00 and 04:00. Males were visually identified and observed producing the staccato "L calls" twice at Madidi and once at Wildsumaco.

Audio recordings at both sites were made with Marantz PMD 661/671 Digital recorders (sampling rate: 44.1 kHz; resolution: 16-bit) and Sennheiser ME 66 directional microphones. At Madidi several locations were visited and a total of five recordings (one per site) were made. Most recordings were incomplete as they often began once individuals had already started vocalizing. For example, at one location. patches were visited for seven nights in a row and on some nights no bamboo rats could be heard responding to vocalizations of the focal individual and in other instances calling could be heard, but a distance away from it and only on one night was a complete vocalization recorded. Three focal bamboo patches at Wildsumaco were observed for three nights in July 2010 and a total of 19 recordings were made. In line with Emmons' work (1981), our observational data suggest that a single pair resides in each bamboo patch and individuals within patches vocalized approximately every 10 min. At Madidi, vocalizations occurred after longer time intervals, approximately every 45 min. It is unclear what prompts bamboo rats to vocalize; whether it is in response to vocalizations given by rats in other patches is uncertain. Playback of a previous recording was attempted at two sites in Madidi to see if individuals were present in patches, and while at one location a bamboo rat responded, no response was elicited at the other site. Listening to their calls in the forest suggests that the rats are calling and responding to one another in different patches, as seen in other counter-singing mammals, but tests must be conducted to ascertain the true nature of vocalizations to ensure what we perceive as coordinated behavior is not due to random chance. The loud vocalizations of bamboo rats may help them identify and be aware of other rats' location; given their highly specialized low nutrient diet of bamboo, proper spacing and low energy communication networks may be selected for Emmons, 1981.

Overlapping duets were recorded at both site; however, this was a rare occurrence, and because of limited observations and few recordings, it is uncertain if this is a seasonal behavior, occurs year-round or whether it is tied to reproduction. Recordings of duets made at Wildsumaco and Madidi were visualized with Raven bioacoustic software (**Figure 1**). Despite the lack of individual identification, elements of these three vocalizations reveal structural features that are found in other mammals that engage in TCIVs. At both Wildsumaco and Madidi, soft grunting calls in response to the loud staccato L calls–the latter featuring a typical increase in inter-note intervals–could be viewed as interactive duets between members of the same

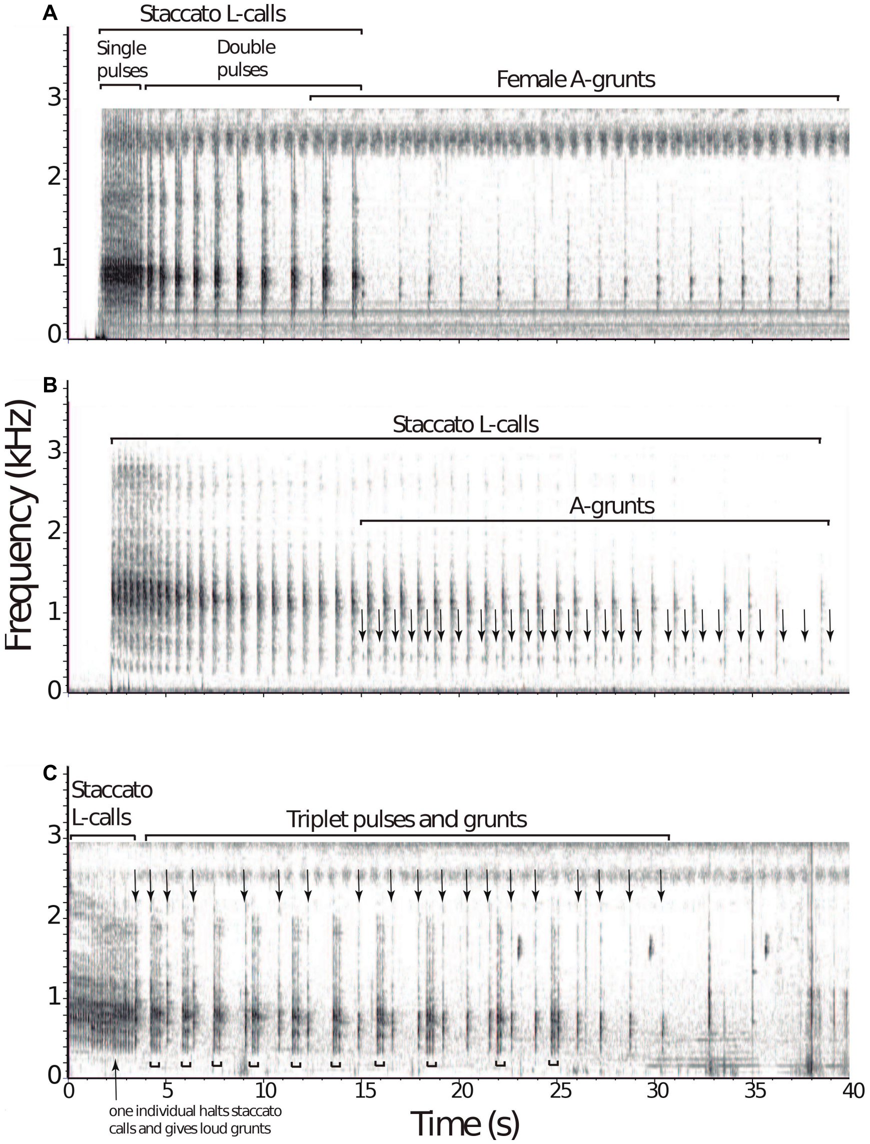


FIGURE 1 | Examples of vocal exchanges in Neotropical bamboo rats. At Wildsumaco, two instances of duetting in the Amazonian bamboo rat, *Dactylomys dactylinus* were recorded. **(A)** The first instance involved loud staccato "L calls" by a male, starting with single pulses followed shortly after by doubles pulses emitted at a decreasing rate. A power spectrum analysis of L-calls (both single and double pulses; n = 23) revealed one peak of energy at 801 Hz. After the eighth double pulse, soft grunting "A calls" by another individual followed, assumed to be a female based on Emmons' previous work. Spectral analysis of A-calls (n = 16) revealed two peaks of energy at 559 and 743 Hz. **(B)** At Madidi, a similar duetting pattern was recorded from the Bolivian bamboo rat, *D. boliviensis*. A male gave 21 loud staccato "L calls," followed by a soft grunting "A calls" from another individual. A power spectrum analysis of L-calls (n = 21) revealed one peak of energy at 1193 Hz. Note the decrease in call rate of the staccato vocalization. **(C)** In the second instance at Wildsumaco, both individuals engaged in the loud staccato calls "L calls." After 10 pulses by the first individual, another individual joined in with loud L calls that ceased shortly after, followed by single loud pulses while the other individual continued staccato L calls in triplets, answered by the grunts. A power spectrum analysis of the first 10 pulses revealed one peak of energy at 754 Hz. Spectrograms were prepared with Raven-Lite software (v. 2.0.3, Cornell Laboratory of Ornithology, Ithaca, NY, United States), using a window size of 2048 points. Spectral analysis of calls was performed using the Audacity software (v. 2.4.2).

patch (**Figures 1A,B**). These intra-patch male-female duets may aid in coordination and social bonding. In one instance two individuals engaged in loud staccato vocalizations (**Figure 1C**). Individuals in different bamboo patches may counter vocalize in a more competitive territorial manner and inter-patch male-male vocalizing may be a form of counter-singing announcing and delimiting borders, especially when territorial rivals come in close proximity to one another. Experimentation is needed to discern the true nature of TCIVs in bamboo rats.

The potential duetting behavior detected in *Dactylomys* spp. might be the result of ecological features unique to these rodents. Duetting and antiphonal vocalizations evolved in several mammals that need to communicate over long distances in dense forest environments, like those inhabited by the bamboo rats. The social structure of bamboo rats is not completely understood; they are thought to live in family groups with a single male (and female and offspring) occupying a single bamboo patch (Eisenberg, 1989). Investigations of the closely related Brazilian bamboo rat *Kannabateomys amblyonyx* have found variation in mating systems (either polygynous or monogamous) depending on resources, including number of females and bamboo patch availability (Silva et al., 2008). It is likely that the genus *Dactylomys* is monogamous and as seen in monogamous primates, (e.g., tarsiers, titi monkeys and hylobatids), duetting may help to strengthen pair bonds, coordinate movement, as well as send territorial information to other conspecifics in the area.

DISCUSSION AND FUTURE DIRECTIONS

Much of the literature on TCIVs, including antiphonal vocalizations and duets, contains confounding terminology. With multiple definitions of antiphonal vocalizations, duets, antiphonal duet, duet calls, songs and duet singing, forming a cohesive framework for discerning patterns and testing hypotheses can be challenging (Filippi et al., 2019; De Gregorio et al., 2022). Regardless of terminology, these TCIVs are rarely investigated in non-primate mammals. The disparity in ecology and sociality of the different mammals herein discussed, makes it difficult to discern if commonalities exist in these important interactive vocalizations. However, two main themes emerge for antiphonal communication and duetting. First, most of the species discussed live in habitats in which visual proximity is restricted, whether it is a dense tropical forest or a vast ocean. The second theme is the highly social nature of these calls. Highly social mammals may require TCIVs to maintain and reinforce social relationships amongst group members, similar to vocal grooming in primates, the more complex mammalian societies become the more complex their vocal repertoire and vocalizations may become (Dunbar, 2012). There is some evidence for this in bats, in which antiphonal vocalizations and counter singing have been identified (Knörnschild, 2014). Knörnschild et al. (2020) showed a positive relationship between the information contained in the contact and isolation calls of bat species and the size of the specie's social group suggesting a link between social and vocal complexity across bat species.

Vocal communication that encodes specific information about individuals and functions in maintaining social relationships may be selected for, regardless of social system, in visually isolated, yet social species. For example, sperm whales are highly social, yet visually restricted from group members in the ocean environment, and duets help to reinforce their social bonds (Schulz et al., 2008). The question then is why do we see so little of this type of interactive communication in mammals? Duetting is common in bonded pairs and the scarcity of monogamy and shared parental responsibilities may also account for less mammalian representation (Lukas and Clutton-Brock, 2013). Many group living mammals may be in close visual proximity to other group members, making these sorts of communicative channels unnecessary. Turn-taking vocalizations may however evolve in social units in which individuals can see and alter their and other's behavior with their vocalizations, as is the case with meerkats (Demartsev et al., 2018). Counter-singing is even rarer than duetting and its rarity might stem from how rare singing generally is in mammals, which has only been investigated in a few taxa. It may be that mammals, relying heavily on scents, are simply less vocal than other groups like birds that rely heavily on vision and sound; or more interactive communication is occurring in mammals, but we have yet to detect it with studies. The human auditory range is limited, and mammals frequently produce and perceive sound at frequencies beyond human auditory abilities (Heffner and Heffner, 2018). Both infrasound and ultrasound are used by mammals in terrestrial and aquatic habitats (Martin et al., 2017) and detection of these vocalizations require specialized bioacoustics monitoring equipment and this fact may help to explain the paucity of data (Ladich and Winkler, 2017; Romero-Mujalli et al., 2021).

Bamboo rats produce loud, audible vocalizations, and they might engage in TCIVs. Evidence for duetting exists, but the frequency, causes and adaptive value of this behavior have yet to be deciphered. Most of what is known about bamboo rat vocalizations comes from anecdotal field recordings by ornithologists and through Emmons (1981) work in Ecuador. The data we collected in both Wildsumaco and Madidi were observational in nature and not intended or designed to test specific hypotheses and in the case of Madidi the records were part of an integrated inventory of wildlife at the park. One difficulty in studying bamboo rats is their nocturnal secretive nature and the dense vegetation in which they reside. At both study sites bamboo rats were seldom seen and often when they were spotted, they froze and stopped vocalizing, making it difficult to follow individuals and gather behavioral data. We expect bamboo rat calls will vary depending on the different ecological and social factors, including vegetation structure, seasonality, population density, reproductive stage, etc. Dialects between and within populations may also exist and be another confounding variable when comparing different species and populations. New passive recording technologies may provide a solution, with arrays of recorders in bamboo patches researchers may be able to answer questions about the timing and frequency of these behaviors (see Szymański et al., 2021). Duetting may be more common during the breeding season, which is unknown, and further analysis of vocalizations could provide information on the temporal elements of the vocalizations themselves as well as inter- and intraspecific and population differences. Duetting may have several functions in bamboo rats and non-primate mammals in general. Playback experiments may shed light on the territorial nature of these vocalizations, whether counter calling exists, and how and if bamboo rats respond differently to individuals within their patch (social unit), in nearby patches (neighbors) and those further away (strangers). Mammals with complex vocal communication should be sought out and investigated to determine their prevalence, as well as to test hypotheses on the ecological and evolutionary pressures leading to TCIVs. Bamboo rats are just one example of a mammal whose conspicuous vocalizations have yet to be fully examined.

Mammalian species that live in visually restrictive habitats and require complex vocalizations to maintain long-term social relationships are likely to utilize TCIVs. Several mammal groups may be ideal targets for future research, including close relatives of mammals known to produce TCIVs. For example, forest dwelling elephant species, including the African forest elephant *Loxodonta cyclotis* and Asian elephants *Elephas maximus* produce rumble vocalizations like African elephants, but reside in slightly different habitats with differing visibilities (Pardo et al., 2019). Canids, including gray wolves *Canis lupus* and jackal species (*C. aureus*, *Lupulella adusta*, and *L. mesomelas*) are also highly vocal and social (Moehlman, 1987; Jenner et al., 2011; Zaccaroni et al., 2012). Like maned wolves, jackals are monogamous, but display variation in social complexity within and across species (Moehlman, 1987). Comparative studies of closely related species may shed light on the evolution of TCIVs. Rock and tree hyraxes mentioned earlier live in very different habitats with different social structures, yet both utilize TCIVs. Spotted hyenas *Crocuta crocuta* living in fission-fusion societies utilize long distance "whoop" calls and comparisons could be made with the solitary and monogamous striped hyena *Hyaena hyaena* (Mills, 1989; Holekamp et al., 2007; Califf et al., 2020). In addition to seeking out new species and conducting comparative studies, examination of the physiological mechanisms underpinning call emission and sound reception should be undertaken. The future of the field of mammal vocalization is promising and insights from diverse taxa will strengthen our understanding of antiphonal calls, duets and counter-singing.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because observational study, review not required by institution.

AUTHOR CONTRIBUTIONS

EV and NH collaborated on the outline and direction of the manuscript. EV gathered information on coordinated vocalizations in mammals, as well as collected and analyzed vocalizations from Wildsumaco, and wrote the manuscript. NH provided vocalizations and observational data from Madidi, as well as reference material and, provided feedback. Both authors read and approved the final version of the manuscript before submission.

FUNDING

Identidad Madidi was an interinstitutional project led by Wildlife Conservation Society, Bolivia Program, with funds from the Gordon and Betty Moore Foundation. EV received funding for work at Wildsumaco from Jacksonville University.

ACKNOWLEDGMENTS

We thank Travis Knowles and Daniela Arenas-Viveros for their feedback on the initial manuscript. Patrice Adret and reviewers made significant suggestion and provided editorial assistance to improve iterations of the manuscript; we thank them for their insight and patience. Additionally, Patrice Adret assisted with the spectrogram analysis and figure creation. NH was grateful for the support of the environmental authorities in Bolivia (DGBAP and SERNAP). The project in Madidi was conducted under the authorization # MMAYA/VMABCCGDF/DGABAP 353/2014 and 130/2019. NH also thanks Rob Wallace, the Director of the Greater Madidi Tambota Program, Wildlife Conservation Society, Bolivia Program for all the support during the Identidad Madidi Project. Ornithologist Victor Hugo García helped NH with recording bamboo rats in the field and the laboratory cleaning of the calls, and Marisol Hidalgo Cossio and the Identidad Madidi team for on the ground support during field work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.906546/full#supplementary-material

REFERENCES

Adret, P., Dingess, K. A., Caselli, C. B., Vermeer, J., Martínez, J., Luna Amancio, J. C., et al. (2018). Duetting Patterns of Titi Monkeys (Primates, Pitheciidae: Callicebinae) and Relationships with Phylogeny. *Animals* 8:178. doi: 10.3390/ani8100178

Balieiro, F., and Monticelli, P. F. (2019). Evidence of individual discrimination in the maned wolf long-distance extended-bark. *Behav. Processes* 158, 219–227. doi: 10.1016/j.beproc.2018.12.004

Banerjee, A., Phelps, S. M., and Long, A. L. (2019). Singing mice. *Curr. Biol.* 29:R190–R191.

Behr, O., Knörnschild, M., and von Helversen, O. (2009). Territorial counter-singing in male sac-winged bats (*Saccopteryx bilineata*): low-frequency songs trigger a stronger response. *Behav. Ecol. Sociobiol.* 63, 433–442.

Bezerra, A. M. R., da Silva, N. J. Jr., and Marinho-Filho, J. (2007). The amazon bamboo rat *Dactylomys dactylinus* (Rodentia: Echimydae: Dactylomyinae) in the cerrado of central Brazil. *Biota Neotrop.* 7, 235–237. doi: 10.1590/S1676-06032007000100030

Boughman, J. W., and Moss, C. F. (2003). "Social sounds: vocal learning and development of mammal and bird calls," in *Acoustic Communication*, eds A. M. Simmons, R. R. Fay, and A. N. Popper (New York, NY: Springer), 138–224.

Brown, P. E., Brown, T. W., and Grinnell, A. D. (1983). Echolocation, Development, and Vocal Communication in the Lesser Bulldog Bat, *Noctilio albiventris*. *Behav. Ecol. Sociobiol.* 13, 287–298.

Califf, K. J., Green, D. S., Wagner, A. P., Scribner, K. T., Beatty, K., Wagner, M. E., et al. (2020). Genetic relatedness and space use in two populations of striped hyenas (*Hyaena hyaena*). *J. Mammal.* 101, 361–372.

Carter, G. G., Logsdon, R., Arnold, B. D., Menchaca, A., and Medellin, R. A. (2012). Adult Vampire Bats Produce Contact Calls When Isolated: Acoustic Variation by Species, Population, Colony, and Individual. *PLoS One* 7:e38791. doi: 10.1371/journal.pone.0038791

Carter, G. G., Skowronski, M. D., Faure, P. A., and Fenton, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Anim. Behav.* 76, 1343–1355.

Clink, D. J., and Lau, A. R. (2020). Adherence to Menzerath's Law is the exception (not the rule) in three duetting primate species. *R. Soc. Open Sci.* 7:201557. doi: 10.1098/rsos.201557

Clink, D. J., Tasirin, J. S., and Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Curr. Zool.* 66, 173–186. doi: 10.1093/cz/zoz035

Courts, R., Erbe, C., Wellard, R., Boisseau, O., Jenner, K. C., and Jenner, M. N. (2020). Australian long-finned pilot whales (*Globicephala melas*) emit stereotypical, variable, biphonic, multi-component, and sequenced vocalisations, similar to those recorded in the northern hemisphere. *Sci. Rep.* 10:20609 . doi: 10.1038/s41598-020-74111-y

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., et al. (2021). Rhythm in dyadic interactions. *Philos. Trans. R. Soc. B* 376:20200337.

Demartsev, V., Bar Ziv, E., Shani, U., Goll, Y., Koren, L., and Geffen, E. (2016a). Harsh vocal elements affect counter-singing dynamics in male rock hyrax. *Behav. Ecol.* 27, 1397–1404.

Demartsev, V., Ilany, A., Barocas, A., Bar Ziv, E., Schnitzer, I., Koren, L., et al. (2016b). A mixed strategy of counter-singing behavior in male rock hyrax vocal competitions. *Behav. Ecol. Sociobiol.* 70, 2185–2193.

Demartsev, V., Ilany, A., Kershenbaum, A., Geva, Y., Margalit, O., Schnitzer, I., et al. (2017). The progression pattern of male hyrax songs and the role of climactic ending. *Sci. Rep.* 7:2794. doi: 10.1038/s41598-017-03035-x

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., and Manser, M. (2018). Vocal Turn-Taking in Meerkat Group Calling Sessions. *Curr. Biol.* 28, 3661–3666.e3. doi: 10.1016/j.cub.2018.09.065

Dunbar, R. I. (2012). Bridging the bonding gap: the transition from primates to humans. *Philos. Trans. R. Soc. B* 367, 1837–1846.

Dunnum, J. L., and Salazar-Bravo, J. (2004). Dactylomys boliviensis. *Mamm. Species.* 745, 1–4.

Eisenberg, J. F. (1989). *Mammals of the Neotropics, The Northern Neotropics Volume 1*. Chicago, IL: University of Chicago Press.

Emmons, L. (1981). Morphological, ecological, and behavioral adaptations for arboreal browsing in *Dactylomys dactylinus (Rodentia. Echimyidae)*. *J. Mammal.* 62, 183–189.

Emmons, L. (2013). *The Maned Wolves of Noel Kempff Mercado National Park*. Washington, DC: Smithsonian Institution Scholarly Press.

Ferreira, L. S., Sábato, V., Pinheiro, T. A., Neto, E., Rocha, L. H., Baumgarten, J., et al. (2022). Long-Distance Counter Calling in Maned Wolves: Friends or Foes? *Animals* 12:1081.

Filippi, P., Hoeschele, M., Spierings, M., and Bowling, D. L. (2019). Temporal modulation in speech, music, and animal vocal communication: evidence of conserved function. *Ann. N.Y. Acad. Sci.* 1453, 99–113.

Fitch, W. T. (2006). Production of vocalizations in mammals. *Vis. Commun.* 3, 115–121.

Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behav. Ecol. Sociobiol.* 55, 415–430.

Hall, M. L. (2009). A Review of Vocal Duetting in Birds. *Adv. Study Behav.* 40, 67–121.

Heffner, H. E., and Heffner, R. S. (2018). "The evolution of mammalian hearing," in *To the Ear and Back – Advances in Auditory Biophysics*, eds C. Bergevin and S. Puria (Melville, NY: American Institute of Physics Publishing), 130001–130008.

Holekamp, K., Benson-Amram, S., Theis, K., and Greene, K. (2007). Sources of variation in the long-distance vocalizations of spotted hyenas. *Behaviour* 144, 557–584.

Jenner, N., Groombridge, J., and Funk, S. M. (2011). Commuting, territoriality and variation in group and territory size in a black-backed jackal population reliant on a clumped, abundant food resource in Namibia. *J. Zool.* 284, 231–238.

Jin, L., Yang, S., Kimball, R. T., Xie, L., Yue, X., Luo, B., et al. (2015). Do pups recognize maternal calls in pomona leaf-nosed bats, *Hipposideros pomona*? *Anim. Behav.* 100, 200–207.

Knörnschild, M. (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* 28, 80–85.

Knörnschild, M., Fernandez, A. A., and Nagy, M. (2020). Vocal information and the navigation of social decisions in bats: is social complexity linked to vocal complexity? *Funct. Ecol.* 34, 322–331.

Knörnschild, M., and Von Helversen, O. (2008). Nonmutual vocal mother–pup recognition in the greater sac-winged bat. *Anim. Behav.* 76, 1001–1009.

Ladich, F., and Winkler, H. (2017). Acoustic communication in terrestrial and aquatic vertebrates. *J. Exp. Biol.* 220, 2306–2317.

Langmore, N. E. (2002). Vocal duetting: definitions, discoveries and directions. *Trends Ecol. Evol.* 17, 451–452.

La Val, R. K. (1976). Voice and Habitat of *Dactylomys dactylinus* (Rodentia: Echimyidae) in Ecuador. *J. Mammal.* 57, 402–404.

Leighty, K. A., Soltis, J., Wesolek, C. M., and Savage, A. (2008). Rumble vocalizations mediate interpartner distance in African elephants, *Loxodonta africana*. *Anim. Behav.* 76, 1601–1608.

Levinson, S. C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Logue, D. M., and Krupp, D. B. (2016). Duetting as a collective behavior. *Front. Ecol. Evol.* 4:7. doi: 10.3389/fevo.2016.00007

Lukas, D., and Clutton-Brock, T. H. (2013). The evolution of social monogamy in mammals. *Science* 341, 526–530. doi: 10.1126/science.1238677

Marler, P. (2004). Bird calls: their potential for behavioral neurobiology. *Ann. N.Y. Acad. Sci.* 1016, 31–44.

Martin, K., Tucker, M. A., and Rogers, T. L. (2017). Does size matter? Examining the drivers of mammalian vocalizations. *Evolution* 7, 249–260. doi: 10.1111/evo.13128

Miller, P. J., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Anim. Behav.* 67, 1099–1107.

Mills, M. G. (1989). "The comparative behavioral ecology of hyenas: the importance of diet and food dispersion," in *Carnivore Behavior, Ecology, and Evolution*, ed. J. L. Gittleman (Boston, MA: Springer), 125–142.

Moehlman, P. D. (1987). Social Organization in Jackals: The complex social system of jackals allows the successful rearing of very dependent young. *Am. Sci.* 75, 366–375.

Nieder, A., and Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos. Trans. R. Soc. B* 375:20190054. doi: 10.1098/rstb.2019.0054

O'Connell-Rodwell, C. E., Wood, J. D., Wyman, M., Redfield, S., Puria, S., and Hart, L. A. (2012). Antiphonal vocal bouts associated with departures in free-ranging African elephant family groups (*Loxodonta africana*). *Bioacoustics* 21, 215–224.

Okobi, D. E., Banerjee, A., Matheson, A. M. M., Phelps, S. M., and Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988. doi: 10.1126/science.aau9480

Oli, M. K., and Jacobson, H. A. (1995). Vocalizations of barking deer (*Muntiacus muntjak*) in Nepal. *Mammalia* 59, 179–186.

O'Shea, T. J., and Poché, L. B. Jr. (2006). Aspects of underwater sound communication in Florida manatees (*Trichechus manatus latirostris*). *J. Mammal.* 87, 1061–1071.

Pardo, M. A., Poole, J. H., Stoeger, A. S., Wrege, P. H., O'Connell-Rodwell, C. E., Padmalal, U. K., et al. (2019). Differences in combinatorial calls among the 3 elephant species cannot be explained by phylogeny. *Behav. Ecol.* 30, 809–820.

Paula, B., and Monticelli, P. (2021). Maned wolf duet? The first record of two maned wolves vocalizing simultaneous. *Acad. Lett.* [Epub ahead of print]. doi: 10.20935/AL1720

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* 285:20180598. doi: 10.1098/rspb.2018.0598

Ravignani, A., Verga, L., and Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Ann. N. Y. Acad. Sci.* 1453, 12–21.

Romero-Mujalli, D., Bergmann, T., Zimmermann, A., and Scheumann, M. (2021). Utilizing DeepSqueak for automatic detection and classification of mammalian vocalizations: a case study on primate vocalizations. *Sci. Rep.* 11:24463. doi: 10.1038/s41598-021-03941-1

Rossi, I., Mauri, L., Laficara, S., and Apollonio, M. (2002). Barking in roe deer (*Capreolus capreolus*): seasonal trends and possibile functions. *Hystrix* 13, 13–18.

Rosti, H., Pihlström, H., Bearder, S., Pellikka, P., and Rikkinen, J. (2020). Vocalization analyses of nocturnal arboreal mammals of the Taita Hills Kenya. *Diversity* 12:473. doi: 10.3390/d12120473

Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Anim. Behav.* 76, 1977–1988.

Silva, R. B., Vieira, E. M., and Izar, P. (2008). Social monogamy and biparental care of the neotropical southern bamboo rat (*Kannabateomys amblyonyx*). *J. Mammal.* 89, 1464–1472.

Soltis, J., Leong, K., and Savage, A. (2005). African elephant vocal communication I: antiphonal calling behavior among affiliated females. *Anim. Behav.* 70, 579–587.

Szymański, P., Olszowiak, K., Wheeldon, A., Budka, M., and Osiejuk, T. S. (2021). Passive acoustic monitoring gives new insight into year-round duetting behaviour of a tropical songbird. *Ecol. Indic.* 122:107271.

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018). Male white-handed gibbons flexibly time duet contributions. *Behav. Ecol. Sociobiol.* 72:16.

Tilson, R. L., and Norton, P. M. (1981). Alarm duetting and pursuit deterrence in an African antelope. *Am. Nat.* 118, 455–462.

Todt, D., and Naguib, M. (2000). Vocal interactions in birds: the use of song as a model in communication. *Adv. Study Behav.* 29, 247–296.

Vergara, V., Michaud, R., and Barrett-Lennard, L. (2010). What can captive whales tell us about their wild counterparts? Identification, usage, and ontogeny of contact calls in belugas (*Delphinapterus leucas*). *Int. J. Comp. Psychol.* 23, 278–309.

Vernes, S. C. (2017). What bats have to say about speech and language. *Psychon. Bull. Rev.* 24, 111–117. doi: 10.3758/s13423-016-1060-3

Williams, H. W., Sorenson, M. W., and Thompson, P. (1969). Antiphonal calling of the tree shrew *Tupaia palawanensis*. *Folia Primatol.* 11, 200–205. doi: 10.1159/000155269

Yoshida, S., and Okanoya, K. (2005). Evolution of turn-taking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165.

Yosida, S., and Okanoya, K. (2009). Naked mole-rat is sensitive to social hierarchy encoded in antiphonal vocalizations. *Ethology* 115, 823–831.

Yosida, S., and Okanoya, K. (2012). Bilateral lesions of the medial frontal cortex disrupt recognition of social hierarchy during antiphonal communication in naked mole-rats (*Heterocephalus glaber*). *J. Comp. Physiol. A.* 198, 109–117. doi: 10.1007/s00359-011-0692-z

Zaccaroni, M., Passilongo, D., Buccianti, A., Dessi-Fulgheri, F., Facchini, C., Gazzola, A., et al. (2012). Group specific vocal signature in free-ranging wolf packs. *Ethol. Ecol. Evol.* 24, 322–331.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 28 July 2022
DOI 10.3389/fevo.2022.906322

EDITED BY
Dena Jane Clink,
Cornell University, United States

REVIEWED BY
Nora Hengst Prior,
University of Maryland, College Park,
United States
Charles T. Snowdon,
University of Wisconsin–Madison,
United States

*CORRESPONDENCE
Chiara De Gregorio
chiara.degregorio@unito.it

SPECIALTY SECTION
This article was submitted to
Behavioral and Evolutionary Ecology,
a section of the journal
Frontiers in Ecology and Evolution

RECEIVED 28 March 2022
ACCEPTED 11 July 2022
PUBLISHED 28 July 2022

CITATION
De Gregorio C, Zanoli A, Carugati F,
Raimondi T, Valente D, Torti V,
Miaretsoa L, Rajaonson A, Gamba M
and Giacoma C (2022)
Parent-offspring turn-taking dynamics
influence parents' song structure
and elaboration in a singing primate.
Front. Ecol. Evol. 10:906322.
doi: 10.3389/fevo.2022.906322

Parent-offspring turn-taking dynamics influence parents' song structure and elaboration in a singing primate

Chiara De Gregorio[1]*, Anna Zanoli[1], Filippo Carugati[1], Teresa Raimondi[1], Daria Valente[1], Valeria Torti[1], Longondraza Miaretsoa[1], Andry Rajaonson[2], Marco Gamba[1] and Cristina Giacoma[1]

[1]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy, [2]Groupe d'Étude et de Recherche sur les Primates de Madagascar, Antananarivo, Madagascar

Parent-offspring interactions are essential to interpret animal social evolution and behavior, but their role in mediating acoustic communication in animals that interact vocally is still unclear. Increasing evidence shows that primate vocal communication is way more flexible than previously assumed, and research on this topic can provide further information on how the social environment shaped vocal plasticity during the evolution of the Primate order. Indris communicate through elaborated vocal emissions, usually termed songs. Songs are interactive vocal displays in which all members of the family group alternate their emissions, taking turns during chorusing events. We aimed to understand whether specific rules regulate the turn-taking of different group members and investigate the flexibility of indris' vocal behavior when co-singing with their offspring. We found that social factors can influence the turn-taking organization in a chorus, as offspring were more likely to drop out from the parents' duet than join in, and we speculate that overlap might signal competition by members of the same-sex. The duet between the reproductive pair was the most common type of singing organization, followed by a duet between mothers and sons and the triadic interaction between mother, father, and son. Interestingly, parents' solo singing seems to stimulate offspring to vocalize, and we also found that mothers and fathers simplify, at least in part, song elaboration when chorusing with offspring. Our results indicate that indris can perform short-time adjustments to the number of co-emitters and their identity: our approach is advantageous in highlighting the multilevel influences on primate vocal flexibility. Moreover, it provides evidence that some aspects of our vocal plasticity were already present in the lemur lineage.

KEYWORDS

chorus, lemur, primate, flexibility, elaboration, duet, rhythm, song

Introduction

Animals of different species interact vocally in their natural environment (Tobias et al., 2016; De Gregorio et al., 2022). Individuals can adapt their vocal behavior to other emitters during these interactions to produce coordinated vocal displays, such as duets or choruses (Gamba et al., 2016). The interplay between emitters is a crucial feature of human conversations, but the level of non-human animals' flexibility in vocal exchanges is still debated (Levinson, 2016). This topic has attracted great interest because of its possible implications for language evolution and similarity with human conversational rules (Chow et al., 2015; Pika et al., 2018).

Increasing evidence shows that the ability to take turns is widespread in different groups of primates. In New World monkeys, for example, squirrel monkeys (*Saimiri sciureus*) adjust the timing of their vocal exchange depending on the co-caller identity (Masataka et al., 1986), while in marmosets (*Callithrix jacchus*), the emission of different call types can be affected by the timing of another individual's vocalization (Liao et al., 2018). Chimpanzees (*Pan troglodytes*) modify their vocalizations to promote chorusing with social partners (Fedurek et al., 2013a), and duetting gibbons can adapt their contribution to that of the other pair-member (*Hylobates pileatus*, Traeholt et al., 2006; *Nomascus leucogenys*, Deputte, 1982). Moreover, gibbons' ability to adapt their vocal behavior to an external factor (e.g., forced partner exchange or predator presence) also emerged in siamangs (*Symphalangus syndactylus*, Geissmann, 1999) and white-handed gibbons (*Hylobates lar*, Clarke et al., 2006).

In particular, vocal interactions can occur between adults and juveniles of many species and are often crucial for developing adult-like vocal communication (humans, Goldstein and Schwade, 2008; birds, Chen et al., 2016; primates, Koda et al., 2013), enhancing vocal production learning (i.e., the ability to change the structure of vocalizations due to hearing others). In birds, for example, this process can occur by listening to a tutor (Mennill et al., 2018) or during direct interactions between older and younger animals (Rivera-Cáceres et al., 2018; Carouso-Peck et al., 2020).

While there is extensive work on birds' juvenile-tutor vocal interactions, these mechanisms have been scarcely investigated in primates. Previous studies examined the antiphonal calling of the common marmoset (*C. jacchus*, Chow et al., 2015; Takahashi et al., 2015) and the co-singing interaction of gibbons (*Hylobates agilis*, Koda et al., 2013). These works suggest that parents could instantaneously influence juvenile/infants' vocalizations. Nevertheless, many of these investigations focused on the offspring side of the interaction (e.g., Takahashi et al., 2016), highlighting infant vocal developmental trajectories shaped by adult feedback, while parents' vocal behavior remained almost unexplored. Koda et al. (2013) provided an interesting case, showing that, in gibbons, mothers had a more stereotyped singing pattern when singing together (co-singing) with daughters than when singing alone. This evidence suggests that the identity of a co-singer can shape individual vocal behavior, but this is not the only feature to consider when investigating individual contributions in collective displays.

Human conversations can occur between more than two people, and the number of people participating can influence turn-taking dynamics (Sacks et al., 1974). As in humans, animal vocal interactions can occur with many participating individuals and varying degrees of overlap between emitters (Passilongo et al., 2015; Torti et al., 2018). In birds, for example, chorusing can often involve two males and one female or two females and one male, and the temporal organization of individuals' contribution may favor or avoid overlap (*Monias benschi*, Seddon, 2002; *Pheugopedius euophrys*, Mann et al., 2006). The composition of the social group can also influence chorus structure and duration (*Dacelo novaeguineae*, Reyer and Schimdl, 1988).

More than two group members' simultaneous emission of utterances occurs in different primate species, such as the pant-hoot chorusing of chimpanzees (*P. troglodytes*, Fedurek et al., 2013b) or the roaring bouts of howler monkeys (*Alouatta pigra*, Horwich and Gebhard, 1983). Despite chorusing occurring quite commonly in many singing primate species, the majority of work on this behavior is still rather descriptive (De Gregorio et al., 2022), and the extent to which the number of conspecifics in a choral display can influence the individual contribution remains pretty much unexplored. Taken altogether, these pieces of evidence suggest that many animals can adjust their utterances to external factors, such as the vocal behavior of co-emitter (e.g., sexual partner, offspring, preferred social partner), but it is unclear whether also the number of those co-emitters can regulate the structure of vocalizations in interacting animals.

We aimed to fill the gap about understanding adult changes during singing with offspring by investigating parent-offspring singing interactions in the indris (*Indri indri*). Besides possessing a rich vocal repertoire (Valente et al., 2019), indris are the only singing lemurs. They live in family groups (Bonadonna et al., 2019; Rolle et al., 2021) in the eastern rainforest of Madagascar, where every member can simultaneously participate in the choral display (Torti et al., 2017, 2018). Units of different types composing indri's songs can be emitted alone (single notes) or organized in phrases of two to six units (Zanoli et al., 2020), with shorter phrases (i.e., including two and three units) more likely to be included in the songs (Valente et al., 2021). Indris emit units and phrases with a precise rhythmic pattern (De Gregorio et al., 2021a). Songs serve different functions, such as inter-and intra-group communication and territory defense (Torti et al., 2013; Bonadonna et al., 2020). Indris' songs are sex-specific duets between males and females (Giacoma et al., 2010), where the calls can be given alternated or simultaneously. One or two additional individuals may participate in the vocal displays

(De Gregorio et al., 2019) so that animals can take turns within the same song. Thus, individuals join in, and others drop out during the same song.

Field observations suggest that sexual competition between parents and offspring of the same-sex can occur (Bonadonna et al., 2014), as observed in birds (Seddon et al., 2002), and it could be of interest to understand if chorusing dynamics can reflect this competition. Therefore, we hypothesize that singing behavior in the indris can be regulated by balancing the competition in singing among group members to advertise their identity or mated/unmated status. Furthermore, avoiding excess overlapping between singers allows for maintaining the communicative function of the vocal display. Therefore, we predict that turn-taking behavior among individuals will not be random. However, it will show specific trajectories as, for example, adult individuals are more likely to sing together than with juveniles. Our second hypothesis is that parent's vocal behavior can enhance offspring's vocal development: as social factors and auditory feedback seems to mediate the development of singing behavior (De Gregorio et al., 2021b), we predict that (I) co-singing interactions would affect the temporal structure of parents' songs, in line with the idea that social influences might shape temporal regulation of utterances (Henry et al., 2015). We also predicted that (II) parents will utter less elaborated songs when co-singing with their offspring, agreeing with previous gibbons' findings (Koda et al., 2013). Our approach allows disentangling different aspects of social influences on parents' contribution, as we will consider not only the identity of co-singers (pair mate, male offspring, and female offspring) but also their numerosity, as previous work showed that the number of singers in a chorus might influence the individual performance (Gamba et al., 2016; De Gregorio et al., 2019).

Materials and methods

Observation and recordings

We collected data in the Maromizaha New Protected Area (Eastern Madagascar: 18° 56′ 49″ S, 48° 27′ 53″ E), with field observations conducted between 6:00 am and 1:00 pm, from 2010 to 2020, for a total of 63 months. We recorded spontaneous songs from a close distance (between 2 and 10 m) of 8 reproductive pairs from 8 habituated groups of indris. We performed the recordings using Sound Devices 702T, Olympus LS-100 and LS05, and Tascam DR-100, DR-40, and DR-05 with semi-directional microphones (ME 67 and AKG CK 98) oriented toward the vocalizing individuals. We set the recorders at a sampling rate of 44.1 kHz and an amplitude resolution of 16 bit during all the recording sessions. Files were saved in wav format. We were able to recognize all animals individually based on their natural marks. Our dataset comprised 440 duets and choruses (of two or more than two individuals, respectively), resulting in 826 individual contributions uttered by 16 individuals within eight reproductive pairs. Indris uttered 260 of the contributions during cosinging interactions with offspring (female offspring, $N_{songs} = 84$; male offspring, $N_{songs} = 176$).

Acoustic analyses

Indris' songs usually start with a series of roars, harsh emissions that are supposed to have an "attention gather" function (e.g., Hopkins et al., 2007). After that, indris emit a variable number of "long notes" (LN), which are longer and less modulated than the subsequent units. After those, we can find isolated units (or "single notes," SN) or units organized in phrases of descending fundamental frequency ("descending phrases," DP) that can include 2–6 units. We analyzed the indris' choruses using Praat 6.0.56 (Boersma and Weenink, 2007).

Rhythmic song features

We identified the contribution to the song of each singer using annotations in Praat TextGrids via visual inspection of the spectrograms and fieldwork notes. Spectrograms had a 0–7,000 dB view range, with a window of 0.006 s and 60.0 dB of dynamic range. First, we annotated the onset and offset of each unit for each contribution and labeled it according to the singer's identity (Mother, Father, Son, Daughter). We labeled each unit according to its type and position (e.g., being part of a phrase or not, position within the phrase). Since the core of the indris' songs relies on descending phrases and single notes, we focused our analysis on these vocal types: SN and DPs (DP2, DP3, DP4, DP5, DP6 based on the number of units forming the phrase). Next, we labeled silent gaps within units according to their position (**Supplementary Figure 1**): "inter" for silent intervals between different DPs, and "intra" for silent intervals between units of the same DP (De Gregorio et al., 2019). We used a custom Praat script to extract each interval duration (Gamba et al., 2015). To evaluate the rhythmic structure of parents' contributions, we imported the duration of the intervals in R (R Core Team, 2020; version 3.4.3) to calculate the inter-onset intervals (IOI) within (WP) and between phrases (BP; De Gregorio et al., 2019, 2021a).

Turn-taking in co-singing dynamics

We then focused on parents' contributions (Mothers and Fathers), and we labeled each overlapping part of the song according to the number and identity of vocalizing individuals. We did it by annotating when an animal would join the song and the exact timing in which it would stop singing. When an offspring started vocalizing during a silent interval between two units of a parent's contribution, we considered the whole interval part of the singing interaction. We did the same when an individual stopped singing in between the silent gap of one of the parents' contributions. We obtained eight types of co-singing associations (**Figure 1**): M-F for mothers' contribution when duetting with fathers, F-M for fathers' contributions when duetting with mothers, M-S for mothers' contribution when duetting with sons, F-D for fathers' contributions when duetting with daughters, M-F-S for mothers' contribution when singing with fathers and sons, F-M-S for fathers' contribution when singing with mothers and sons, and the same for M-F-D and F-M-D. We used the code F for fathers and M for mothers for the portion where parents sang "solo phrases" without overlapping with other family members. We then transformed each contribution into a string of consecutive co-singing types.

Song elaboration

We transformed each parents' song into a string of labels representing the phrases' concatenation within an individual contribution, separated by a break symbol (e.g., SN| DP2| DP3| DP3| DP4). Then, we separated each string into different co-singing types, and we obtained 483 strings for males and 663 for females. To investigate if co-singing with offspring would affect parents' song features, we used two measures of song elaboration: (a) the Levenshtein distance (hereafter, LD): a logic distance expressing the minimum cost to convert a sequence into another one (Kohonen, 1985), which has already been proven to be a robust quantitative approach for investigating animal acoustic sequences (Kershenbaum and Garland, 2015); (b) The Normalized Phrase Diversity: an index indicating the diversity of the individual contribution, calculated as the number of different vocal types emitted during a particular co-singing interaction, normalized for the total number of elements uttered during that interaction.

Statistical Analysis

Rhythmic song features

To investigate if singing with offspring would influence the song's rhythmic structure (between phrase Inter-onset intervals, or bpIOI), we used a linear mixed model (LMM, lmer function of lme4 package, Bates et al., 2015). Before fitting the model, bpIOI was log-transformed since it did not show a normal distribution and then used as a response variable; we used the interaction between parent identity and the type of co-singing as a fixed factor. In addition, we included the singer's identity and the specific song contribution from which we extracted the IOIs as nested random factors. Finally, we used the Tukey test (within the multiple contrast package *multcomp* in R) to perform all pairwise comparisons for all levels of the interaction (Bretz et al., 2010). To investigate if singing with offspring would influence the phrase rhythmic structure (within-phrase IOI, or wpIOI), we used a generalized linear mixed model (GLMM, *glmmTMB* package, Brooks et al., 2017), fitting a beta distribution as suggested by the package *fitdistrplus* (Delignette-Muller and Dutang, 2015) as a suitable theoretical distribution. We used wpIOI as the response variable and the interaction between parent identity and co-singing as a fixed factor. In addition, we included the singer's identity and the specific song contribution from which we extracted the IOIs as nested random factors. We verified the assumptions of normality and homogeneity of residuals for both models via visual inspection of the *qqplot* and the residuals' distribution (a function provided by R. Mundry). We also excluded the presence of collinearity among predictors considering variance inflation factors (*vif* package, Fox and Weisberg, 2011). To test for the significance of our full models (Forstmeier and Schielzeth, 2011), we compared them against null models containing only the random factors, with a likelihood ratio test (Anova with argument test "Chisq", Dobson, 2002). We report estimates, standard error (SE), *z*-, and *p*-values for the Tukey test.

Turn-taking in co-singing dynamics

To understand the mechanisms governing the process of taking turns in indri choruses, we ran strings representing the dynamics of each singing event in Behatrix software (version 0.9.13, Friard and Gamba, 2021). This software independently generates the code for a flowchart representing the transitions between behaviors and performs a permutation test to indicate the statistical significance associated with the different transitions. We considered only the cases in which there is at least an alternation between singers, thus excluding songs consisting of only one type of duet or one type of co-singing with three individuals. Thus, our dataset comprised 203 parents' contributions for this analysis, 135 for mothers and 68 for fathers. We used a 5% cut-off on the total number of transitions. First, we generated a flow diagram with the transitions from one co-singing condition to the next, with the percentage values of transition relative occurrences. Then, we ran a permutation test based on observed counts of the transitions between different co-singing types (*Random permutation test* in Behatrix). We permuted the strings 10,000 times, providing an accuracy of 0.001 of the probability values, and we obtained *p*-values for each transition between different co-singing conditions. Finally, we analyzed mothers' and fathers' co-singing transitions separately to evaluate how chorusing dynamics would influence each parents' singing, and we calculated the frequency of different co-singing types for mothers and fathers.

Song elaboration

Levenshtein distance

To investigate differences in the combinatorics (i.e., the concatenation of phrases) between mothers and fathers in each co-singing type, we calculated the LD for each pair of strings in Behatrix (version 0.9.13, Friard and Gamba, 2021). First, we obtained a squared matrix composed of the distances

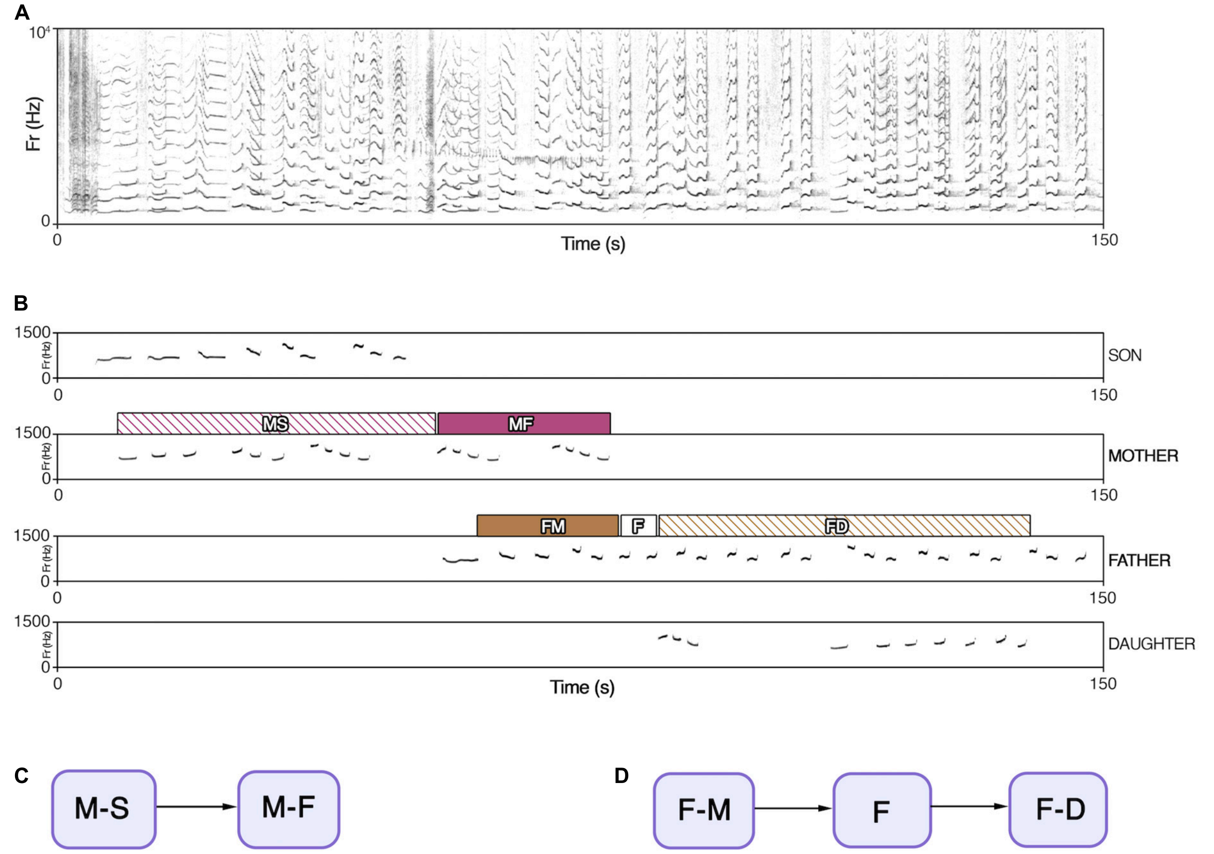


FIGURE 1
(A) A spectrogram of an indri chorus. **(B)** Individual contributions extracted from panel **(A)**; the parents' type of co-singing is highlighted by different color bands. MS, mother duetting with son; MF, mother duetting with father; FM, father duetting with mother; F, fathers' solo units, FD, father duetting with daughter. **(C)** Flow diagram representing mother's co-singing transitions, extracted from panel **(B)**. **(D)** Flow diagram representing fathers' co-singing transition, extracted from panel **(B)**.

between each pair of strings in the dataset. Next, we labeled the files according to the identity of the mother/father and the type of co-singing (e.g., fathers: FM, FD, FMS, MFD; mothers MF, MS, MFS, MFD) during which it was emitted. Then, we investigated whether mothers and fathers differed in their degree of variability depending on the co-singing type. Due to the juvenile singing variability and sample size (De Gregorio et al., 2021b), we did not consider offspring sex (i.e., MFO: the contribution of a mother singing with her pair-mate and an offspring). We averaged the LDs by labels to calculate the within- and between-labels average for each mother and father in each co-singing type using R (R Core Team, 2020). Finally, we performed four Mantel tests (9,999 randomizations) using in each test a reduced matrix with the mean LDs for the labels of interest against a matrix containing zero when the matching labels were of the same co-singing type (Krull et al., 2012), and one when they were of different co-singing type (*vegan* R-package; Oksanen et al., 2013). We checked the admissible number of permutations for our matrices through the function *numPerms* of the same R-Package. We then analysed differences between the mothers/fathers' LDs in different co-singing types and between mothers' and fathers' LDs in the same co-singing type (MF vs. FM, MF vs. MFO, FM vs. FMO, MFO vs. FMO) by using the paired sample *t*-test. Finally, we verified the normal distribution of each label with a Shapiro-Wilk test for normality (built-in R-package *stats*) and computed the test's power with the *pwr.t.test* function (*pwr* R-package; Champely et al., 2018).

Phrase diversity

To investigate differences in the composition of parents' contributions in the different co-singing types, we calculated the Normalized Diversity for each contribution as the sum of each DP type normalized on the total number of DPs composing the string (*stringr* R-package; Wickham, 2019). Then, we ran a linear mixed model (LMM, *lme4* R-package; Bates et al., 2015) to investigate whether the co-singing type influenced the contributions' diversity. The model included log-transformed normalized diversity as the response variable, the co-singing type as the fixed factor and song and individual identity as the nested random factors. The co-singing type was a categorical

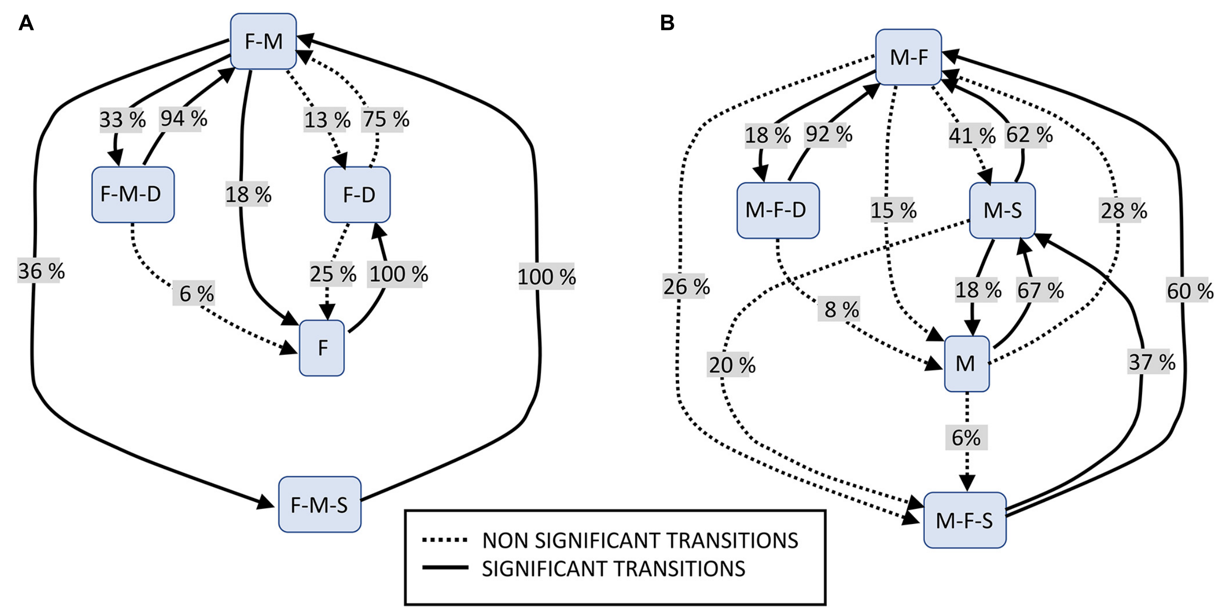


FIGURE 2
Flow diagram showing the occurrence of transitions (%) between different types of co-singing. **(A)** Fathers: FM, fathers' contribution when duetting with mothers; FMS, when duetting with mothers and sons; FD, when duetting with daughters; FMS, when duetting with mothers and sons; F, fathers' phrases with no overlap with any group members. **(B)** Mothers: MF, mothers' contribution when duetting with fathers; MFD, when duetting with fathers and daughters; FMS, when duetting with fathers and sons; MS, when duetting with sons; M, mothers' phrases with no overlap with any group member.

variable indicating in which of the 16 co-singing categories (8 for fathers, 8 for mothers) each individual contribution was emitted. First, we verified the normal distribution and homogeneity of the model residuals through a *qqplot* and residuals against fitted values plot (a function provided by R. Mundry). Next, we compared the full model with a null model comprising only the random factor, we used a likelihood ratio test (Anova test with the "Chisq" argument; Dobson, 2002) and we calculated the *p*-values for predictors using a likelihood ratio test between the full and the null model (Barr et al., 2013). Finally, we performed all pairwise comparisons for the levels of the factor co-singing type with the Tukey test (R-package *multcomp* Bretz et al., 2010).

Results

Turn-taking in co-singing dynamics

The sequential analysis of co-singing types indicated that both fathers and mothers showed non-random turn-taking behavior in co-singing dynamics. In particular, six out of ten possible transitions occurred above chance for fathers (solid lines in **Figure 2A**). Co-singing with mothers only followed co-singing with sons and pair-mates (F-M-S→F-M, $p < 0.001$). After this duetting, fathers' solo phrases (F) take place (F-M→F, $p = 0.002$), followed by duets with the daughters (F→F-D, $p = 0.005$). Daughters can join the duet between the reproductive couple (F-M→F-M-D, $p = 0.001$), and, from that singing organization, they usually drop out from the interaction (F-M-D→F-M, $p < 0.001$), leaving only fathers and mothers singing (F-M). Additionally, after duetting only with mothers (F-M), fathers would duet with mothers and their son (F-M→F-M-S, $p = 0.016$). The other possible transitions did not occur significantly more than chance (dotted lines in **Figure 2**). Moreover, fathers most commonly duetted with mothers (52% of cases, F-M), followed by co-singing with mothers and their sons (F-M-S, 20% of cases) and then by co-singing interaction with mothers and daughters (M-F-D, 15% of cases). On the other hand, duetting with daughters (F-D, 7% of cases) and solo phrases (F, 6% of cases) were less frequent.

For mothers, seven out of 15 possible transitions occurred above chance (solid lines in **Figure 2B**). Co-singing with fathers and their offspring (both sons and daughters) was followed by duetting with fathers only (M-F-S→M-F, $p < 0.001$; M-F-D→M-F, $p < 0.001$), meaning that either daughters or sons ceased singing while their parents kept vocalizing. Still, mothers would also sing with their daughters and partners after duetting only with their partners (M-F→M-F-D, $p < 0.001$), even if this transition was less likely to occur than the opposite one. Moreover, mothers' solo phrases (M) occurred before (M→M-S, $p = 0.002$), but also after duetting with their sons (M-S→M, $p = 0.008$). As for fathers, also for mothers, the most common co-sing type was duetting with their pair mate (42% of cases), followed by duetting with their sons (30% of cases). Singing with both sons and fathers occurred in 14% of co-singing interaction, while the involvement of daughters was more unusual (M-F-D), occurring in 5% of cases, even less than solo phrases (M, 9%).

Influence of co-singing type on parents' rhythmic song features

The comparison between the full and null model for the between-phrases IOI showed that the two models were significantly different ($\chi 2 = 38.877$, df = 13, $p < 0.001$). The Tukey test indicated that mothers had significantly longer inter-onset intervals between phrases (bpIOI) when co-singing with their sons, compared to co-singing with their pair-mates ($p < 0.003$). The same was not true for fathers since we found no differences in the bpIOI values between co-singing with daughters or their pair-mates. Moreover, the Tukey test did not show any differences in bpIOI depending on the number of individuals involved. We reported the detailed results for the models and the Tukey tests in **Supplementary Table 1**.

On the other hand, the comparison between the full and null models for the within-phrase IOI (wpIOI) did not reach statistical significance ($\chi^2 = 8.637$, df = 13, $p = 0.471$) and thus, the fixed factors did not affect the duration of the inter-onset intervals between units given within a particular phrase.

Song elaboration

Levenshtein distances

We analyzed 1,051 parents' contributions composed of 17,326 phrases. We found a significant difference between the LDs calculated for mothers and fathers when duetting between parents (MF vs FM Mantel test: $r = 0.125$, $p = 0.013$). Mothers showed higher average individual means (mean LD = 16 ± 1.92) than fathers (mean LD = 11.8 ± 0.89; Paired t-test: $t = 5.0407$, df = 7, $p = 0.001$). When considering parents' phrase combinations when singing with other two individuals (the other pair-member plus one offspring), we found that mothers and fathers did not differ from each other (**Figure 3B**; MFO vs FMO, Mantel test: r −0.03467, $p = 0.968$). We also found that mothers showed a more stereotyped singing pattern when singing with their partner and one offspring than when singing with their partners only (**Figures 3A**, **4**, MF vs MFO, Mantel test: $r = 0.3478$, $p < 0.001$), with higher average individual means when duetting with fathers (mean LD: 16 ± 1.92) than in the chorus including the offspring (mean LD = 6.1 ± 3.68; t-test: $t = -4.2556$, df = 7, $p = 0.004$; **Figure 3A**). The same was true for fathers, whose LD values were significantly different when duetting with their pair mate than when co-singing with their pair mate and one offspring (**Figure 3B**; FM vs MFO, Mantel test: $r = 0.2303$, $p = 0.005$). As for mothers, fathers had higher average individual means when duetting with their pair (mean LD: 11.8 ± 0.89) than in the chorus with also the offspring (mean LD = 6.1 ± 3.68; t-test: $t = 11.293$, df = 7, $p < 0.001$; **Figure 3A**).

Phrase diversity

When investigating the diversity of phrases forming the individual contributions of mothers and fathers in the different co-singing types, we found that the full model significantly differed from the null model ($\chi 2 = 69.692$, df = 7, $p < 0.001$). We reported estimate, SE, z- and p-values for all the pairwise comparisons of the Tukey test in **Supplementary Table 2**. When considering duets between mothers and fathers, we found that mothers showed less diversity than fathers (FM-MF, **Figure 5**). Moreover, mothers had a higher diversity when singing with their pair and an offspring than in a duet. In other words, we found an effect of the numerosity of individuals singing together, with mothers being more diverse when singing in a chorus of three individuals including the pair and one offspring (whichever its sex) than when in a duet (either with the other member of the reproductive pair or with an offspring, regardless of its sex).

We found a different pattern for fathers, with a partial influence of the number of co-singers on the phrase diversity. In particular, we found a significant difference when comparing fathers singing with the other member of the reproductive pair and fathers singing in a triadic chorus including the pair and a son (FM-FMS, **Figure 5**). We found no difference in phrase diversity when comparing a father singing with the other member of the pair or when a daughter joined the chorus (FM-FMD, **Figure 5**). Lastly, when considering three individuals singing together, we found no effect of the co-singing types on the phrase diversity regardless of the chorus's composition for both mothers and fathers (i.e., MFD vs MFS, FMD vs MFS).

Discussion

We examined turn-taking dynamics in the choruses emitted by the indris' family groups, and we found that the alternation between different singers is not casual but follows specific trajectories. Moreover, we investigated whether co-singing interactions with sons and daughters affected the song structure of adult indris, and we found that co-singing would influence both the rhythmic structure and the song elaboration.

Co-singing dynamics

Our work indicated that, within chorusing dynamics, the duet was the most common type of song organization for parents. However, we also found that duetting with the opposite-sex offspring is quite common for mothers but not fathers. A possible explanation for this difference is that

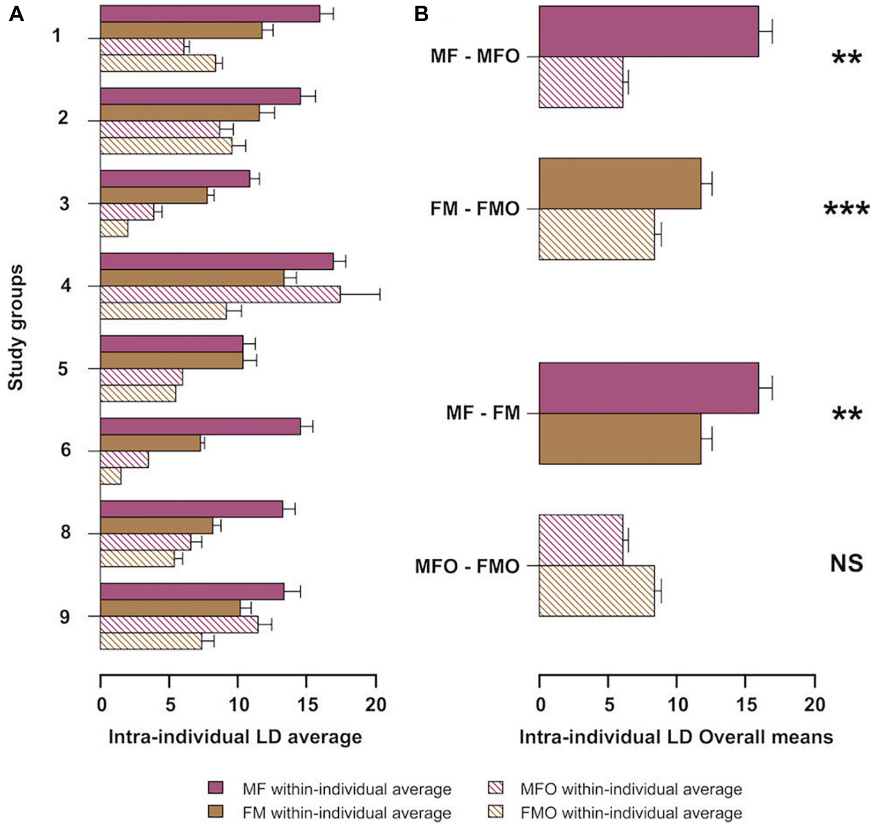


FIGURE 3
The average Levenshtein distance (LD) for mothers and fathers in the two co-singing conditions: singing only with the pair-mate (MF, mothers' contribution when duetting with fathers, FM, fathers' contribution when duetting with mothers; color-filled barplot) and singing with the pair-mate and an offspring (MFO, mothers' contribution when duetting with fathers and offspring; FMO, fathers' contribution when duetting with mothers and offspring; striped barplot). Capped lines represent standard deviation. **(A)** LD within mothers and fathers for the eight studied groups **(B)** Overall LD average between mothers and fathers; *t*-test significance at $p < 0.001$ is denoted by ***, at $p = 0.005$, and $p = 0.004$ is denoted by **.

sons remain in their natal group longer than daughters (De Gregorio et al., 2021b).

In particular, we found that offspring were more likely to drop out from the parents' duet than join in. The inverse dynamic was infrequent and occurred when fathers concluded their singing and left sons duetting with their mothers. Also, the emission of parents' solo phrases was always linked to offspring joining the song. This confirms our first prediction: turn-taking behavior among indris is not random but shows specific trajectories. Our results align with black-crested gibbons' singing dynamics, in which the majority of duet bouts are given by the adult pair, with the adult male initiating the song. Still, occasionally the juvenile male starts calling first and duets with the adult female before giving up the turn to the adult male (*Nomascus concolor*, Fan et al., 2009). Offspring of both sexes are more likely to drop out from triadic singing interactions with parents in the indris. This evidence is in line with the idea that overlap between same-sex singers might involve singing competition, similar to what studies on gibbons suggested. When the juvenile male sang first, the adult male promptly started singing before the adult female (Fan et al., 2009).

Indri duets are composed of sex-specific song contributions (Giacoma et al., 2010). Thus, a new participant joining the song would emit the same unit types of one of the singers while coordinating the emission with the other. In human conversations, the overlap between participants has been considered a troublesome feature (Schegloff, 2000), and it could manifest one person's willingness to take the floor (Sacks et al., 1974). Similarly, the overlap between two birds of the same-sex has been considered a signal of aggression (Naguib and Kipper, 2006) and threat by the animal that starts singing before the other has finished vocalizing (Mennill and Ratcliffe, 2004, Baker et al., 2012). In kookaburras, for example, parent-offspring sexual conflicts would manifest through aggressive interactions linked to the participation of offspring in the adult chorus (Parry, 1973; Reyer and Schimdl, 1988). Therefore, we argue that the overlap of indris' utterances might signal an individual willingness to prevail over the other in singing. In siamangs,

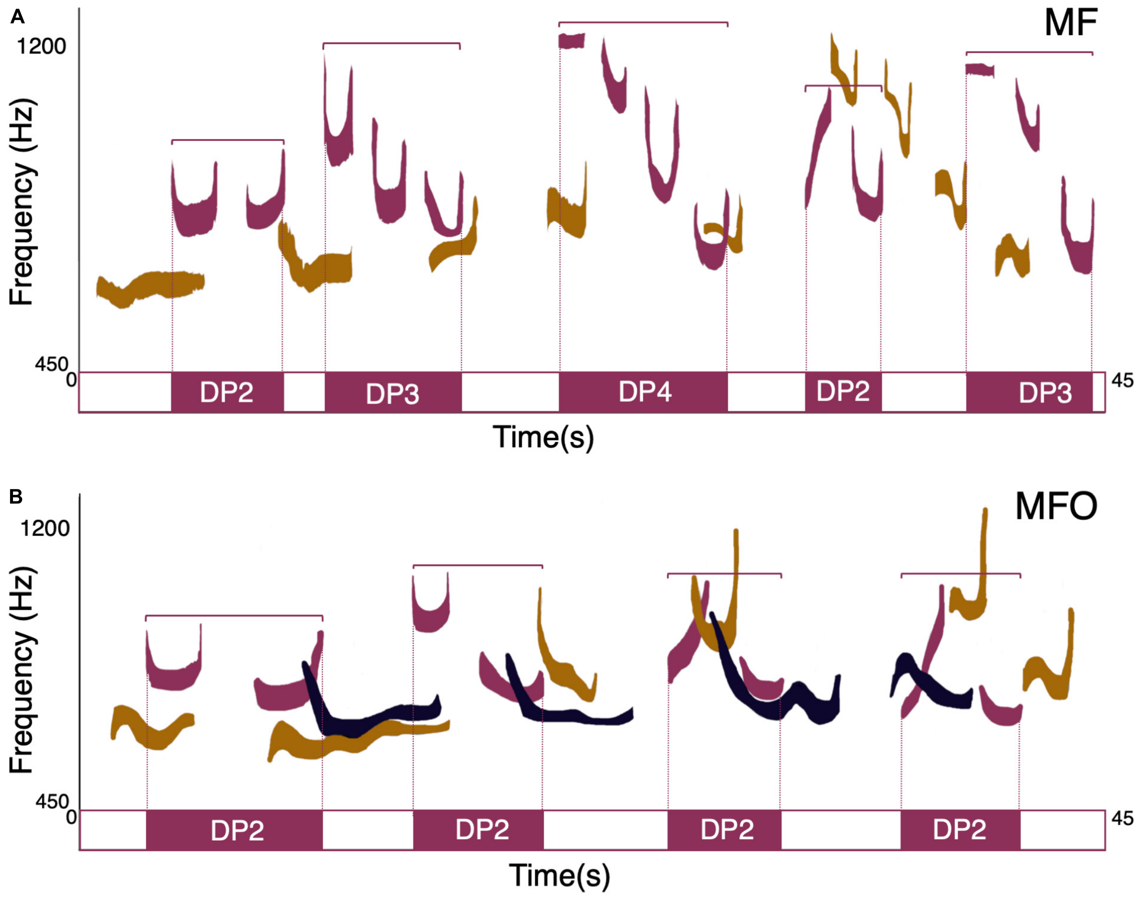


FIGURE 4
Schematic representation of mothers' different phrases' concatenation in two different cosigning types. **(A)** Mother (in purple) duetting with their pair-mate (MF). **(B)** Mother singing with the pair-mate and one offspring (MFO). The fundamental frequency profile of mother's, father's, and offspring's contribution to the song is highlighted in purple, light brown, and blue, respectively.

young individuals join the parents' chorus more frequently as they mature (*S. syndactylus*, Chivers, 1976) and, although these mechanisms have not been investigated in indris yet, we can suppose that the older the offspring gets, the more competition in singing with parents of the same-sex can take place before the offspring disperse. In the case of the replacement of the dominant female, the dominant male and his mature son would compete for mating with the new female, by singing or by physical aggressions (De Gregorio, pers. obs.).

Parents' solo singing can also be critical in regulating and motivating offspring to sing. Our results suggest that both mothers and fathers sang alone before duetting with the offspring of the opposite-sex. This finding aligns with the idea that adults might provide offspring with vocal stimulation for their song development, as auditory feedback and practicing might be essential for song maturation (De Gregorio et al., 2021b). Similarly, Merker and Cox (1999, *Nomascus gabriellae*) reported that, in gibbons, the mother-juvenile co-singing interactions were always initiated by mothers. Hence, if there might be some competition in singing, why do parents encourage sons and daughters to sing? For birds, it has been suggested that group singing might be involved in territorial defense (Mann et al., 2006), as chorusing can inform how many individuals are present in a given territory and, implicitly, their willingness to defend it.

Duetting birds may communicate their ability to engage in aggressive interactions through the degree of vocal coordination (Diniz et al., 2021). Furthermore, during territorial encounters between different indris' groups, both parents and offspring participate in the territorial song (Torti et al., 2013). Thus, maintaining the vocal coordination in the chorus could be essential to indicate the ability or motivation of the family group to defend their resources cooperatively. Also, chorusing behavior might be favored in a territorial context as the overlap of vocalizations can enhance their transmission (Rehberg-Besler et al., 2017). Finally, indris can discriminate between neighboring and non-neighboring singing family groups (Spezie et al., 2022), suggesting that they might vocally recognize

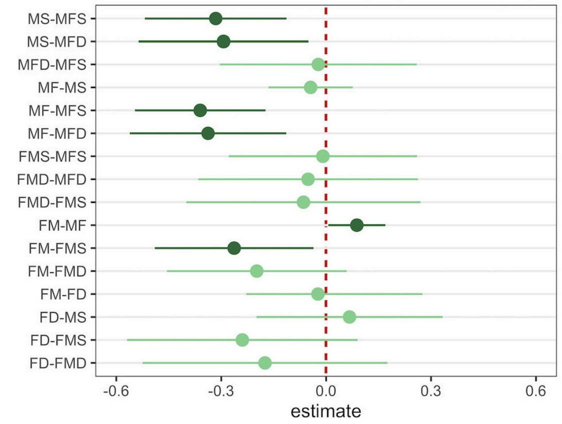


FIGURE 5
Plot showing all pairwise comparisons resulting from the Tukey test. Dark green points and lines represent the estimate and 95% confidence interval of significant comparisons. Light green points and lines represent the estimate and 95% confidence interval of non-significant comparisons. M, mothers; F, fathers; D, daughters; S, sons.

members of the nearby groups. Thus, the offspring's presence, identity, and status should be regularly broadcast to other indris' families.

Summarizing, turn-taking dynamics in the indris' choruses consist in a trade-off between the need for young animals to participate in the chorus in order to practice and to broadcast their unmated status (Gamba et al., 2016; De Gregorio et al., 2021b), and the willingness of their same-sex-parents to renounce to their possibility to advertise their mated status and their presence. Therefore, we can conclude that the probability of an animal singing seems to be influenced by family's social dynamics.

Rhythmic features

We found that mothers, but not fathers, had longer bpIOI when duetting with sons than when duetting with their pair-mate, thus partially confirming our second prediction. Indeed, mothers slowed down the rate of phrase emissions only when duetting with their sons might indicate that variation in song temporal structure can result from a vocal adjustment performed to facilitate offspring vocal development. This interpretation agrees with what was suggested for gibbons' mother-daughter vocal interactions (*H. agilis*, Koda et al., 2013) and marmosets, whose mothers offered a vocal reinforcement to offspring exhibiting appropriate turn-taking (*C. jacchus*, Chow et al., 2015). It is interesting to notice that in humans, slowed infant-directed speech benefits child language learning (Raneri et al., 2020), and human caregivers can improve vocal articulation in stuttering children by slowing down their child-directed speech (Sawyer et al., 2017).

If this evidence suggests that this behavior might enhance sons' vocal development, it remains unclear why there are such differences with father-daughter co-singing. An alternative explanation would be that, given that indri females are notably more flexible than males (Torti et al., 2017; De Gregorio et al., 2019; Zanoli et al., 2020), mothers are simply adjusting their timing to sons' utterances to improve synchronization, even with a less experienced singer. Still, we found that duets between fathers and daughters occur more rarely than mother-son duets. Moreover, in line with previous work (De Gregorio et al., 2019), for both parents, the duration of the inter-onset intervals between two different phrases was not affected by the number of individuals (either one or two) singing simultaneously, independently of the identity and sex of co-singers. Considering the inter-onset intervals between units, the type of co-singing did not have a statistical effect on its duration. This result is in line with previous work on the indris' song evidencing how notes within a phrase are more constrained than the organization of phrases within a song, as this trait does not change during ontogeny (Gamba et al., 2016; De Gregorio et al., 2021b).

Song elaboration

We examined differences in parents' song elaboration during different co-singing interactions, and we found that both the number and identity of co-singers can influence the sequential organization of phrases and their diversity. In particular, our results confirm what was previously found by Zanoli et al. (2020), namely that adult females are more variable in their phrases' combination than adult males when singing together. We also found that fathers and mothers did not differ when a third individual joined the chorus, but their contributions became most stereotyped. Moreover, mothers had lower phrase diversity than fathers during duetting interactions. Nevertheless, mothers uttered more diverse contributions when offspring of both sexes joined the chorus. The same was true for fathers only when the male offspring joined the pair's duet.

According to changes in adults' phrase concatenation, the increase in chorus size is in line with previous work showing that the number of singers can influence the duration of both contribution (the total time spent singing) and phonation (the cumulative duration of the emitted notes, without considering silent gaps) of a female individual in a chorus (De Gregorio et al., 2019). Still, in our case, a variation of phrase combination occurs in both sexes. We suggest two possible, non-mutually exclusive, explanations: first, that it could be an effect of adults adjusting their singing behavior to maintain the coordination of utterances when a third animal joins the chorus (De Gregorio et al., 2019); second, mothers and fathers emit more stereotyped contributions when co-singing with offspring to facilitate them (Koda et al., 2013). Since the third animal joining the pair is always an offspring in our dataset, this analysis did not allow us to separate the influence of the number of co-singers and their identity and/or sex on parents' phrase organization. On the other hand, the number of co-singers seems to have a stronger impact on the diversity of phrases uttered by adult indris. Indeed, mother and father did not vary this feature during duets with the pair-mate compared to duets with the opposite-sex offspring. This is in contrast to findings on humans, where parents simplify their speech during vocal interactions with children by using fewer unique words (Elmlinger et al., 2019). On the one hand, parents increased their phrase's diversity when co-singing with two individuals. On the other hand, fathers' contributions were less influenced by chorus size: they showed a more extensive phrase repertoire only when co-singing with their pair-mate and their sons, in agreement with previous studies (Torti et al., 2017). This result is interesting as it suggests that fathers could face major pressure when singing with an individual of the same-sex so that they might differentiate their singing to better broadcast their individuality. Thus, we can only partially confirm our third prediction: parents uttered less elaborated songs when co-singing with offspring. Parents' phrases were more stereotyped in terms of combination but were more diverse in terms of phrases type during vocal interactions with their pair-mate and offspring.

Future works might consider the longitudinal development of co-singing with parents to understand whether these interactions influence vocal development. A focus on the acoustic resemblance between parents and offspring over time would also be beneficial to understanding whether parent-offspring similarity increases or decreases during ontogeny and whether the sexes show similar trajectories. These findings would be useful for further investigating co-singing conditions and understanding which particular traits are typical of juvenile phases.

In conclusion, our work indicates that indris might regulate parents-offspring turn-taking dynamics in family choruses by different degrees of motivation for conflict or cooperation between parents and offspring. If overlap might signal competition by members of the same-sex, parents also seem to stimulate offspring singing behavior. Moreover, indris perform rapid adaptation not only to the number of co-emitters, as it strongly influences the elaboration of parents' songs, but also to their identity, as it affected mothers' rhythmic structure and fathers' phrase diversity. This mechanism is similar to what happens in human language, where speakers can adapt their speech to their interlocutors (Lee et al., 2021). Our work demonstrates that the interplay between different emitters is a fundamental aspect to consider when investigating short-term flexibility in animals' vocal behavior, and that the social environment is among the major determinants of indris' song structure. Finally, we provide strong evidence that some of the traits that characterize human vocal plasticity were already in place in the lemur lineage, possibly providing a foundation for further evolutionary paths leading to the emergence of human language.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The non-invasive methods used for data collections of wild indris adhere to the International Primatological Society (IPS) 'Principles for the Ethical Treatment of Non-Human Primates'. Field data collection protocols were reviewed and approved by Madagascar's Ministere de l'Environnement, de l'Ecologie et des Forets. Field data collection protocols were also approved by GERP (Groupe d'Etude et de Recherche sur les Primates de Madagascar), the association governing research in the Maromizaha New Protected Area.

Author contributions

CDG: conceptualization. CDG, DV, AZ, TR, FC, AR, and MG: methodology. CDG, DV, VT, TR, and LM: investigation. CDG and MG: writing–original draft. VT, AZ, DV, LM, AR, and CG: writing–review and editing. FC, TR, AZ, and DV: visualization. MG and CG: supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the University of Torino and the Parco Natura Viva—Centro Tutela Specie Minacciate, with the financial assistance of the European Union, through the Project BIRD (ACP SandT Program, contract FED/2009/217077).

Acknowledgments

We thank the local field guides and the field assistants helping during the data collection. We are also grateful to GERP (Groupe d'Etude et des Recherche sur les Primates de Madagascar) for their support during the field activities and to Cesare Avesani Zaborra and Caterina Spiezio for the financial and technical support. We also thank the editors and reviewers for their helpful suggestions during the revision process. We have received permits for this research, each year,

from "Direction des Eaux et Forêts" and "Madagascar National Parks" (formerly ANGAP) [2010 (N°118/10/MEF/SG/DGF/DCB.SAP/SCBSE and N°293/10/MEF/SG/DGF/DCB.SAP/SCB), 2011 (N°274/11/MEF/SG/DGF/DCB.SAP/SCB), 2012 (N° 245/12/MEF/SG/DGF/DCB.SAP/SCB), 2014 (N°066/14/MEF/SG/DGF/DCB.SAP/SCB), 2015 (N°180/15/MEEMF/SG/DGF/DAPT/SCBT), 2016 (N°98/16/MEEMF/SG/DGF/DAPT/SCB.Re and N° 217/16/MEEMF/SG/ DGF/DSAP/SCB.Re)], 2017 (73/17/MEEF/SG/DGF/DSAP/SCB.RE). 2018: 91/18/MEEF/SG/DGF/DSAP/SCB.Re; 2019: 118/19/MEDD/SG/DGEF/DSAP/DGRNE and 284/19/MEDD/SG/DGEF/DSAP/ DGRNE; 2019/2020: 338/19/MEDD/SG/DGEF/DSAP/DGRNE]. Data collection did not require a permit for 2013 because it has been performed by Malagasy citizens only.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.906322/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Schematic representation of the spectrogram of the isolated fundamental frequencies of three Descending Prhases. The sound spectrogram displays time (s) on the x-axis, frequency (Hz) on the vertical axis. We describe acoustic parameter collection of bpIOI and wpIOI.

References

Baker, T. M., Wilson, D. R., and Mennill, D. J. (2012). Vocal signals predict attack during aggressive interactions in black-capped chickadees. *Anim. Behav.* 84, 965–974. doi: 10.1016/j.anbehav.2012.07.022

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., and Weenink, D. (2007). *PRAAT: Doing Phonetics by Computer (Version 5.3.51).*

Bonadonna, G., Torti, V., De Gregorio, C., Valente, D., Randrianarison, R. M., Pozzi, L., et al. (2019). Evidence of genetic monogamy in the lemur Indri (*Indri indri*). *Am. J. Primatol.* 81:e22993. doi: 10.1002/ajp.22993

Bonadonna, G., Torti, V., Randrianarison, R. M., Martinet, N., Gamba, M., and Giacoma, C. (2014). Behavioral correlates of extra-pair copulation in *Indri indri*. *Primates* 55, 119–123. doi: 10.1007/s10329-013-0376-0

Bonadonna, G., Zaccagno, M., Torti, V., Valente, D., De Gregorio, C., Randrianarison, R. M., et al. (2020). Intra- and intergroup spatial dynamics of a pair-living singing primate, Indri indri: a multiannual study of three indri groups in Maromizaha forest, Madagascar. *Int. J. Primatol.* 41, 224–245. doi: 10.1007/s10764-019-00127-5

Bretz, F., Hothorn, T., and Westfall, P. (2010). *Multiple Comparisons Using R.* Boca Raton, FL: Chapman & Hall/CRC Press.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., et al. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* 9, 378–400. doi: 10.3929/ethz-b-000240890

Carouso-Peck, S., Menyhart, O., DeVoogd, T. J., and Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. *Anim. Behav.* 165, 123–132. doi: 10.1016/j.anbehav.2020.04.019

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., et al. (2018). *Package 'pwr'*. Available online at: https://cran.r-project.org/web/packages/pwr/pwr.pdf (accessed February 10, 2022).

Chen, Y., Matheson, L. E., and Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6641–6646. doi: 10.1073/pnas.1522306113

Chivers, D. J. (1976). Communication within and between family groups of siamang (*Symphalangus syndactylus*). *Behaviour* 57, 116–135. doi: 10.1163/156853976X00136

Chow, C. P., Mitchell, J. F., and Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B* 282:20150069. doi: 10.1098/rspb.2015.0069

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One* 1:e73. doi: 10.1371/journal.pone.0000073

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., et al. (2021a). Categorical rhythms in a singing primate. *Curr. Biol.* 31, R1379–R1380. doi: 10.1016/j.cub.2021.09.032

De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., et al. (2021b). Born to sing! Song development in a singing primate. *Curr. Zool.* 67, 585–596. doi: 10.1093/cz/zoab018

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 1:15. doi: 10.1080/03949370.2021.2015451

De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., et al. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Curr. Zool.* 65, 89–97. doi: 10.1093/cz/zoy058

Delignette-Muller, M. L., and Dutang, C. (2015). fitdistrplus: an R package for fitting distributions. *J. Stat. Soft.* 64, 1–34. doi: 10.18637/jss.v064.i04

Deputte, B. L. (1982). "Duetting in male and female songs of the white-cheeked gibbon (*Hylobates concolor leucogenys*)," in *Primate Communication*, eds C. H. Snowdon and M. R. Petersen (Cambridge: Cambridge University Press), 67–93.

Diniz, P., Ramos, D. M., Webster, M. S., and Macedo, R. H. (2021). Rufous horneros perceive and alter temporal coordination of duets during territorial interactions. *Anim. Behav.* 174, 175–185. doi: 10.1016/j.anbehav.2021.02.007

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*, 2nd Edn. Boca Raton, FL: Chapman and Hall.

Elmlinger, S. L., Schwade, J. A., and Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: parents simplify the structure of their speech in response to babbling. *J. Child Lang.* 46, 998–1011. doi: 10.1017/S0305000919000291

Fan, P. F., Xiao, W. E. N., Huo, S., and Jiang, X. L. (2009). Singing behavior and singing functions of black-crested gibbons (*Nomascus concolor jingdongensis*) at Mt. Wuliang, Central Yunnan, China. *Am. J Primatol.* 71, 539–547. doi: 10.1002/ajp.20686

Fedurek, P., Machanda, Z. P., Schel, A. M., and Slocombe, K. E. (2013b). Pant hoot chorusing and social bonds in male chimpanzees. *Anim. Behav.* 86, 189–196. doi: 10.1016/j.anbehav.2013.05.010

Fedurek, P., Schel, A. M., and Slocombe, K. E. (2013a). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behav. Ecol. Sociobiol.* 67, 1781–1789. doi: 10.1007/s00265-013-1585-7

Forstmeier, W., and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav. Ecol. Sociobiol.* 65, 47–55. doi: 10.1007/s00265-010-1038-5

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*, 2nd Edn. Thousand Oaks, CA: SAGE Publications, Inc.

Friard, O., and Gamba, M. (2021). *Behatrix: Behavioral Sequences Analysis with Per-Mutations Test*. Available online at: http://www.boris.unito.it/pages/behatrix (accessed February 04, 2022).

Gamba, M., Friard, O., Riondato, I., Righini, R., Colombo, C., Miaretsoa, L., et al. (2015). Comparative analysis of the vocal repertoire of *Eulemur*: a dynamic time warping approach. *Int. J. Primatol.* 36, 894–910. doi: 10.1007/s10764-015-9861-1

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., et al. (2016). The *Indris* have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Front. Neurosci.* 10:249. doi: 10.3389/fnins.2016.00249

Geissmann, T. (1999). Duet songs of the siamang, *Hylobates syndactylus*: II. Testing the pair-bonding hypothesis during a partner exchange. *Behaviour* 136, 1005–1039. doi: 10.1163/156853999501694

Giacoma, C., Sorrentino, V., Rabarivola, C., and Gamba, M. (2010). Sex differences in the song of *Indri indri*. *Int. J. Primatol.* 31, 539–551. doi: 10.1007/s10764-010-9412-8

Goldstein, M. H., and Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523. doi: 10.1111/j.1467-9280.2008.02117.x

Henry, L., Craig, A. J. F. K., Lemasson, A., and Hausberger, M. (2015). Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. *Front. Psychol.* 6:1416. doi: 10.3389/fpsyg.2015.01416

Hopkins, W. D., Taglialatela, J., and Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Anim. Behav.* 73, 281–286. doi: 10.1016/j.anbehav.2006.08.004

Horwich, R. H., and Gebhard, K. (1983). Roaring rhythms in black howler monkeys (*Alouatta pigra*) of Belize. *Primates* 24, 290–296. doi: 10.1007/BF02381093

Kershenbaum, A., and Garland, E. C. (2015). Quantifying similarity in animal vocal sequences: which metric performs best? *Methods Ecol. Evol.* 6, 1452–1461. doi: 10.1111/2041-210X.12433

Koda, H., Lemasson, A., Oyakawa, C., Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS One* 8:e71432. doi: 10.1371/journal.pone.0071432

Kohonen, T. (1985). Median strings. *Pattern Recognit. Lett.* 3, 309–313. doi: 10.1016/0167-8655(85)90061-3

Krull, C. R., Ranjard, L., Landers, T. J., Ismar, S. M., Matthews, J. L., and Hauber, M. E. (2012). Analyses of sex and individual differences in vocalizations of Australasian gannets using a dynamic time warping algorithm. *J. Acoust. Soc. Am.* 32, 1189–1198. doi: 10.1121/1.4734237

Lee, Y., Goldstein, L., Parrell, B., and Byrd, D. (2021). Who converges? Variation reveals individual speaker adaptability. *Speech Commun.* 131, 23–34. doi: 10.1016/j.specom.2021.05.001

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Liao, D. A., Zhang, Y. S., Cai, L. X., and Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3978–3983. doi: 10.1073/pnas.1722426115

Mann, N. I., Dingess, K. A., and Slater, P. J. B. (2006). Antiphonal four-part synchronized chorusing in a Neotropical wren. *Biol. Lett.* 2, 1–4. doi: 10.1098/rsbl.2005.0373

Masataka, N., Biben, M., and Symmes, D. (1986). Temporal and structural analysis of affiliative vocal exchanges in squirrel monkeys (*Saimiri sciureus*). *Behaviour* 98, 259–273. doi: 10.1163/156853986X00991

Mennill, D. J., and Ratcliffe, L. M. (2004). Overlapping and matching in the song contests of black-capped chickadees. *Anim. Behav.* 67, 441–450. doi: 10.1016/j.anbehav.2003.04.010

Mennill, D. J., Doucet, S. M., Newman, A. E., Williams, H., Moran, I. G., Thomas, I. P., et al. (2018). Wild birds learn songs from experimental vocal tutors. *Curr. Biol.* 28, 3273–3278. doi: 10.1016/j.cub.2018.08.011

Merker, B., and Cox, C. (1999). Development of the female great call in *Hylobates gabriellae*: a case study. *Folia Primatol.* 70, 97–106. doi: 10.1159/000021680

Naguib, M., and Kipper, S. (2006). Effects of different levels of song overlapping on singing behaviour in male territorial nightingales (*Luscinia megarhynchos*). *Behav. Ecol. Sociobiol.* 59, 419–426. doi: 10.1007/s00265-005-0066-z

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., et al. (2013). *Package 'vegan'. Community Ecology Package, Version 2*. Available online at: http://CRAN.R-project.org/package=vegan (accessed February 15, 2022).

Parry, V. (1973). The auxiliary social system and its effect on territory and breeding in kookaburras. *Emu* 73, 81–100. doi: 10.1071/MU973081

Passilongo, D., Mattioli, L., Bassi, E., Szabó, L., and Apollonio, M. (2015). Visualizing sound: counting wolves by using a spectral view of the chorus howling. *Front. Zool.* 12:22. doi: 10.1186/s12983-015-0114-0

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. Lond. B* 285:20180598. doi: 10.1098/rspb.2018.0598

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raneri, D., Von Holzen, K., Newman, R., and Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *J. Child Lang.* 47, 1263–1275. doi: 10.1017/S030500091900093X

Rehberg-Besler, N., Doucet, S. M., and Mennill, D. J. (2017). Overlapping vocalizations produce far-reaching choruses: a test of the signal enhancement hypothesis. *Behav. Ecol.* 28, 494–499. doi: 10.1093/beheco/arw176

Reyer, H. U., and Schimdl, D. (1988). Helpers have little to laugh about: group structure and vocalization in the laughing kookaburra *Dacelo novaeguineae*. *Emu* 88, 150–160. doi: 10.1071/MU9880150

Rivera-Cáceres, K. D., Quirós-Guerrero, E., Araya-Salas, M., Templeton, C. N., and Searcy, W. A. (2018). Early development of vocal interaction rules in a duetting songbird. *R. Soc. Open Sci.* 5, 171791. doi: 10.1098/rsos.171791

Rolle, F., Torti, V., Valente, D., De Gregorio, C., Giacoma, C., and Von Hardenberg, A. (2021). Sex and age-specific survival and life expectancy in a free ranging population of *Indri indri* (Gmelin, 1788). *Eur. Zool. J.* 88, 796–806. doi: 10.1080/24750263.2021.1947398

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). The simplest systematics for the organization of turn taking for conversations. *Language* 50, 696–735.

Sawyer, J., Matteson, C., Ou, H., and Nagase, T. (2017). The effects of parent-focused slow relaxed speech intervention on articulation rate, response time latency, and fluency in preschool children who stutter. *J. Speech Lang. Hear. Res.* 60, 794–809. doi: 10.1044/2016_JSLHR-S-16-0002

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Lang. Soc.* 29, 1–63. doi: 10.1017/S0047404500001019

Seddon, N. (2002). The structure, context and possible functions of solos, duets and choruses in the subdesert mesite (*Monias benschi*). *Behaviour* 139, 645–676.

Seddon, N., Butchart, S. H., and Odling-Smee, L. (2002). Duetting in the subdesert mesite *Monias benschi*: Evidence for acoustic mate defence? *Behav. Ecol. Sociobiol.* 52, 7–16. doi: 10.1007/s00265-002-0488-9

Spezie, G., Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Giacoma, C., et al. (2022). Evidence for acoustic discrimination in lemurs: a playback study on wild indris (*Indri indri*). *Curr. Zool.* zoac009. doi: 10.1093/cz/zoac009

Takahashi, D. Y., Fenley, A. R., and Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* 371:20150370. doi: 10.1098/rstb.2015.0370

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., et al. (2015). The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734–738. doi: 10.1126/science.aab1058

Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., and Nakagawa, S. (2016). Territoriality, social bonds, and the evolution of communal signaling in birds. *Front. Ecol. Evol.* 4:74. doi: 10.3389/fevo.2016.00074

Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Randrianarison, R. M., Friard, O., et al. (2017). An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. *Sci. Rep.* 7:10140. doi: 10.1038/s41598-017-10656-9

Torti, V., Gamba, M., Rabemananjara, Z. H., and Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Ital. J. Zool.* 80, 596–607. doi: 10.1080/11250003.2013.845261

Torti, V., Valente, D., De Gregorio, C., Comazzi, C., Miaretsoa, L., Ratsimbazafy, J., et al. (2018). Call and be counted! Can we reliably estimate the number of callers in the indri's (Indri indri) song? *PLoS One* 13:e0201664. doi: 10.1371/journal.pone.0201664

Traeholt, C., Bonthoeun, R., Virak, C., Samuth, M., and Vutthin, S. (2006). Song activity of the pileated gibbon, *Hylobates pileatus*, in Cambodia. *Primate Conserv.* 2006, 139–144. doi: 10.1896/0898-6207.21.1.139

Valente, D., De Gregorio, C., Favaro, L., Friard, O., Miaretsoa, L., Raimondi, T., et al. (2021). Linguistic laws of brevity: conformity in *Indri indri*. *Anim. Cogn.* 24, 897–906. doi: 10.1007/s10071-021-01495-3

Valente, D., De Gregorio, C., Torti, V., Miaretsoa, L., Friard, O., Randrianarison, R. M., et al. (2019). Finding meanings in low dimensional structures: stochastic neighbor embedding applied to the analysis of *Indri indri* vocal repertoire. *Animals* 9:243. doi: 10.3390/ani9050243

Wickham, H. R. (2019). *Package 'stringr.' CRAN*. Available online at: https://cran.rproject.org/web/packages/stringr/stringr.pdf (accessed February 9, 2022).

Zanoli, A., De Gregorio, C., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., et al. (2020). Sexually dimorphic phrase organization in the song of the indris (*Indri indri*). *Am. J. Primatol.* 82:e23132. doi: 10.1002/ajp.23132

frontiers | Frontiers in Ecology and Evolution

PERSPECTIVE
published: 29 July 2022
doi: 10.3389/fevo.2022.916956

Interhemispheric Brain Communication and the Evolution of Turn-Taking in Mammals

Andrea Ravignani[1,2]†, Massimo Lumaca[2]*† and Sonja A. Kotz[3,4]**

[1] Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Department of Clinical Medicine, Center for Music in the Brain, Aarhus University and The Royal Academy of Music Aarhus, Aalborg, Denmark, [3] Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands, [4] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

In the last 20 years, research on turn-taking and duetting has flourished in at least three, historically separate disciplines: animal behavior, language sciences, and music cognition. While different in scope and methods, all three ultimately share one goal—namely the understanding of timed interactions among conspecifics. In this perspective, we aim at connecting turn-taking and duetting across species from a neural perspective. While we are still far from a defined neuroethology of turn-taking, we argue that the human neuroscience of turn-taking and duetting can inform animal bioacoustics. For this, we focus on a particular concept, interhemispheric connectivity, and its main white-matter substrate, the corpus callosum. We provide an overview of the role of corpus callosum in human neuroscience and interactive music and speech. We hypothesize its mechanistic connection to turn-taking and duetting in our species, and a potential translational link to mammalian research. We conclude by illustrating empirical venues for neuroethological research of turn-taking and duetting in mammals.

Keywords: bioacoustics, brain connectivity, turn-taking, time, music cognition, speech science

Edited by:
Charles T. Snowdon,
University of Wisconsin-Madison,
United States

Reviewed by:
Giorgio Vallortigara,
University of Trento, Italy
Sebastian Ocklenburg,
Ruhr University Bochum, Germany

***Correspondence:**
Andrea Ravignani
andrea.ravignani@mpi.nl
Massimo Lumaca
massimo.lumaca@clin.au.dk
Sonja A. Kotz
sonja.kotz@maastrichtuniversity.nl

† These authors share first authorship

Specialty section:
This article was submitted to
Behavioral and Evolutionary Ecology,
a section of the journal
Frontiers in Ecology and Evolution

***Received:** 10 April 2022*
***Accepted:** 23 June 2022*
***Published:** 29 July 2022*

Citation:
Ravignani A, Lumaca M and Kotz SA (2022) Interhemispheric Brain Communication and the Evolution of Turn-Taking in Mammals.
Front. Ecol. Evol. 10:916956.
doi: 10.3389/fevo.2022.916956

TURN-TAKING AND DUETTING: FROM BEHAVIOR TO COGNITION AND NEUROSCIENCE

No matter the discipline, research on turn-taking and duetting (TTD) is currently a hot topic (e.g., Demartsev et al., 2018; Pika et al., 2018; Benichov and Vallentin, 2020). In the last 20 years (see e.g., **Figure 1**; Ravignani et al., 2019), there has been an increased and converging interest in TTD in at least three separate disciplines: Animal behavior, language sciences and music cognition (Greenfield et al., 2021). Research into language has explored the nuances of both the semantics and precise timing of turn-taking; because of the short time scales involved in turn-taking in conversation and the comparably slower reactivity of the human nervous system, turn-taking in human speech must be predictive, rather than reactive (Stivers et al., 2009). Some have even argued that turn-taking is at the core of human linguistic abilities (Levinson, 2016). Human music research has experienced a strong empirical turn and likewise explored the behavioral bases of interaction, with prime examples including the dynamics of jazz improvised duetting or string quartet synchronization (Wing et al., 2014; Zeng et al., 2017). Finally, animal behavior has also moved on—at least in bioacoustics—from the exclusive study of behavioral patterns in isolation to a balanced mélange of individuals, duets, and choruses. Catching up with avian

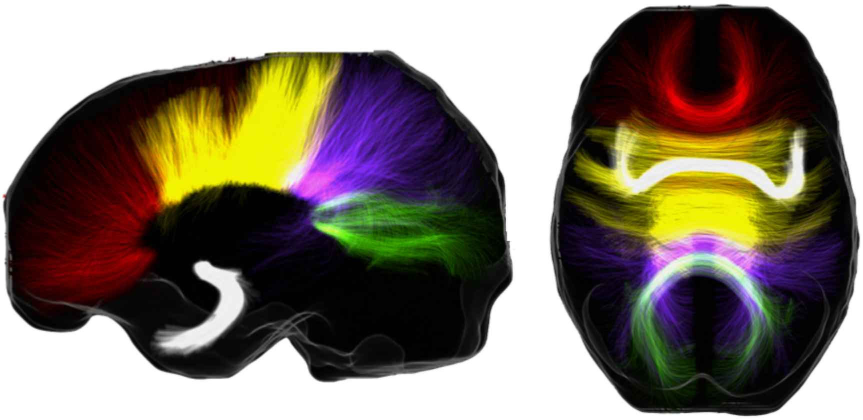


FIGURE 1 | Sagittal (left figure) and axial view (right figure) of the 2 main interhemispheric tracts in humans overlaid on a population template. The tracts included are the corpus callosum with its four geometrical subdivisions (rostrum and genu, in red; truncus or body, in yellow; isthmus, in violet; splenium, in green) and the anterior commissure (in white).

bioacoustics (e.g., Benichov and Vallentin, 2020), findings on TTD patterns in mammals are continuously accumulating (Takahashi et al., 2016; Demartsev et al., 2018; Pika et al., 2018; Ravignani, 2019; Ravignani et al., 2019; de Reus et al., 2021).

Sometimes in all three disciplines, but especially in animal behavior of TTD, the brain is "black-boxed." Behavioral patterns are considered and are sometimes interpreted in cognitive terms. However, the mapping of these behaviors to neural processes is still in its infancy. Here we propose that while we are still far from a clear picture in the neuroethology of turn-taking, there are still many potential low-hanging fruits in this research area. In particular, we offer a perspective on a particular structure in the mammalian brain, the corpus callosum (CC, **Figure 1**), which, we hypothesize, might form basic building blocks of mammalian TTD. We know about the potential neural bases of TTD comparatively much more in humans than any other mammal. Therefore, in the following, we provide an overview of the CC, and its role in human music and spoken language and propose a potential connection between CC and TTD. We discuss the empirical venues its study can open for mammalian comparative TTD research. Of note is that even the human evidence we present is sometimes indirect and speculative; however, we consider building cross-disciplinary bridges to advance valuable knowledge of TTD that has lately received so much attention.

INTERHEMISPHERIC COMMUNICATION AND CORPUS CALLOSUM

In both music and language, an interplay between interhemispheric specialization and communication plays an important role. A key structure enabling and regulating this interplay is the CC, connecting the two hemispheres (Clarke and Zaidel, 1994; Schlaug et al., 1995; Paul et al., 2003; Friederici et al., 2007) and consisting of about 160–190 million fibers (Aboitiz et al., 1992). In music performance, where among others the timing of interactions is key, CC size correlates with musical training (Ozturk et al., 2002). Interhemispheric connectivity has also been linked to musical improvisation: CC may act as a support structure in the lateral perception-action network (Loui, 2018). In language, the CC regulates among others the interplay between syntax/semantics (what to say) and suprasegmental prosody (how to say it) (Friederici et al., 2007). Throughout development, we see that individuals with impaired functionality of the CC may have, among others, impaired TTD-relevant traits (Beens, 1995; Stickles et al., 2002). So, what are the (neuro) biological functions of the CC?

Since the seminal split-brain studies by Sperry (1961) in the fifties and sixties of the last century, we have gained a significant understanding of the role of the CC in interhemispheric brain communication. Specifically, this evidence confirmed that the CC not only transfers information between the two hemispheres of the human brain, but significantly contributes to the development of lateralized function as well as the upkeep of functional integration across the hemispheres (Gazzaniga, 2000; Güntürkün et al., 2020). Despite the lack of general agreement on the functional significance of callosal morphology (Aboitiz et al., 1992), most authors associate a larger callosal area with a better capacity for interhemispheric transfer. A larger diameter and number density of myelinated fibers increase the conduction speed between the two hemispheres, leading to faster and more efficient cross-hemispheric integration and communication (Westerhausen et al., 2006; Horowitz et al., 2015). It is not yet fully understood whether callosal fibers exert an inhibitory or excitatory influence on interhemispheric communication and integration (Bloom and Hynd, 2005); we make no claims here as the particular case of TTD could be explained by both influences. One hypothesis is that the CC provided placental mammals with shorter and more direct pathways between bilateral cortical regions than the anterior commissure, thus speeding up the interhemispheric transfer of information (Buzsáki et al., 2013; Aboitiz, 2017). This could have been even more advantageous in larger brains, such as those of mammals (Aboitiz, 2017; Font et al., 2019).

HUMAN MUSIC AND TURN-TAKING

Some of the characteristics predicted by the general function of CC have been observed in musicians—individuals with more than 6 years of formal musical training (Zhang et al., 2020)—due to training-related neuroplastic effects on callosal fiber composition and volume (Lee et al., 2003). Musicians exhibit a larger midsagittal callosal size (e.g., Schlaug et al., 1995, 2005) and more organized callosal bundles (e.g., Elmer et al., 2016; Habibi et al., 2018) than musically naive individuals. These findings suggest a positive association between the amount of musical training and the strength of interhemispheric connectivity (Schlaug et al., 1995; Bengtsson et al., 2005; Hyde et al., 2009; Steele et al., 2013; Elmer et al., 2016; Habibi et al., 2018). Stronger cross-hemispheric anatomical connections may explain the enhanced capabilities observed in musicians, as well as in non-musicians (Lumaca et al., 2021) for music perception and performance that rely on high-speed interhemispheric conduction, such as binaural temporal integration, visuo-motor integration, and bimanual motor planning, execution, and control (Gooijers and Swinnen, 2014).

In music, these perceptuo-motor skills are critical for different forms of interpersonal synchrony, including playing jazz in ensembles. Thus, one may hypothesize that the ability of individuals to coordinate their actions in time is associated, and can be predicted by, the microstructural characteristics and the size of their CC. Jazz improvisers, who show a large flexibility and precision in their coordination and joint action, exhibit higher callosal integrity and larger tract volume than classical musicians and non-musicians (Zeng et al., 2017). Conversely, deficits in spatial and temporal interpersonal synchrony have been observed in populations with reduced size of the CC such as Autism Spectrum Disorder (Casanova et al., 2009; Fitzpatrick et al., 2016; Kaur et al., 2018). A recent study with fNIRS on Autism Spectrum Disorder children shows that these deficits are accompanied by reduced symmetrical activations in superior and middle temporal regions compared to typically developed children (Su et al., 2020). This research further supports a key role of interhemispheric brain communication in socially embedded actions.

HUMAN SPOKEN LANGUAGE AND TURN-TAKING

A prime example of interactive human communication, with some potential parallels to mammalian TTD, is language use. Successful communication depends on at least two aspects: the rapid integration of lateralized verbal (segmental) and non-verbal (suprasegmental) information in speech/language comprehension and production (Kotz and Schwartze, 2010, 2016) and swift turn-taking in interpersonal interaction (Holler et al., 2015; Levinson, 2016).

The integration of left-hemispheric segmental and right-hemispheric suprasegmental information necessitates rapid information flow between the two hemispheres (Friederici and Alter, 2004). Such transfer likely engages the commissural fibers crossing through the CC, but the structural and functional differentiation of these fibers indicates that they are topographically organized based on their cortical origin. Diffusion-weighted imaging revealed that while the anterior portions of the CC (genu and truncus) connect the orbital and frontal lobes, the posterior third (isthmus and splenium) link the temporal, parietal, and occipital lobes of the two hemispheres (e.g., Huang et al., 2005; Hofer and Frahm, 2006; Zarei et al., 2006; Dougherty et al., 2007; Park et al., 2008). Empirical evidence in support of such topographical connectivity comes from studies exploring the rapid integration of segmental (syntax) and suprasegmental (prosody) information in patients with CC lesions. Results confirmed a reciprocal speech processing interface between the two hemispheres via the posterior CC that seems to underlie speech/language comprehension (Friederici et al., 2007; Sammler et al., 2010). Recent work with children and adolescents, suffering from developmental agenesis of the corpus callosum, a congenital brain disorder where the axons of the CC are either completely or partially absent (Rauch and Jinkins, 1994; Laìbadi and Beke, 2017), have further confirmed the integrative function of the posterior CC within the language network (Bartha-Doering et al., 2020) and its relation to language abilities (Bartha-Doering et al., 2021). Further evidence also points to interhemispheric information flow for other functions. For example, lesions in the middle and posterior troncus of the CC mainly affect interhemispheric transfer of motor and somatosensory information (Risse et al., 1989; Fabri et al., 2001), while isthmus and splenium lesions mostly affect auditory (Pollmann et al., 2002) and visual transfer (Corballis et al., 2004).

Whether or not turn-taking, a communicative act between two or more speakers, relies on the integrative function of the CC *per se* or may only relate to the swift integration between segmental and suprasegmental information that speakers and listeners use in communication is currently an open scientific question. Levinson (2016) proposed that turn-taking is at the root of human communication and might derive from three factors. First, turn-taking is one of many means that form human interaction, including non-verbal information such as facial expression, gaze, and gestures. Second, turn-talking might reflect a prototypical form of human interaction as evidenced in pre-lingual infants, and third turn-taking is also found in non-human primates as part of their vocal communicative repertoire. Recently, the cognitive and neural processes underlying turn-taking in humans have also been explored, highlighting the human capacity to anticipate one's turn to communicate (Holler et al., 2015). Thus, the timely integration of segmental and suprasegmental information while anticipating one's turn might facilitate swift information flow between a speaker and a listener.

To bridge the gap on how turn-taking might rely on CC interhemispheric connectivity might be best informed by studies on developmental agenesis of the CC. Next to altered language abilities (Bartha-Doering et al., 2021), there are reports on communication deficits that entail pragmatic skills (understanding jokes and non-verbal cues such as emotional vocal and facial expressions; see for example, Brown et al., 2005;

Tu et al., 2009). Further, younger children with CC agenesis seem to struggle with initiating and sustaining conversations and lack a basic understanding of social reciprocity and non-verbal communication (Badaruddin et al., 2007). Adults with CC agenesis further exhibit difficulties in social cognition and social behavior (Laìbadi and Beke, 2017). Of note is their atypical facial scanning while observing emotional expressions of others (Bridgman et al., 2014). A speculative conclusion could therefore be that missing out on rapidly detecting and integrating non-verbal information in social interaction affects how children and adults, suffering from CC agenesis, understand and implement turn-taking.

PERSPECTIVES FOR TURN-TAKING RESEARCH ACROSS MAMMALS LINKING BRAIN TO BEHAVIOR

Some building blocks of human TTD may have mechanistic bases in the CC. These building blocks may be potentially analogous or homologous to TTD in other mammals (Fröhlich, 2017; Pika et al., 2018; Anichini et al., 2020; Greenfield et al., 2021). What do we know about TTD mechanisms in other species? Some work has found links between rodent turn-taking behavior and underlying neural mechanisms, highlighting hierarchical, cortical control of this behavior (Okobi et al., 2019). If the interhemispheric neuroscience of TTD in humans is little explored, the non-human animal counterpart is even less so. While, to our knowledge, no mechanistic link has been sought between TTD and their interhemispheric neural bases in other mammals, we consider that combined behavioral evidence in mammals and neural evidence in humans can provide fruitful research directions and predictions for experiments. In other words, we suggest to (1) look for mammalian TTD traits which show strong behavioral similarities to either human music or spoken language; (2) capitalize on human neuroscientific evidence while translating it to other species for which at least some (callosal) brain anatomy is known, and (3) within these species, target and behaviorally test those with maximally divergent predictions in terms of how interhemispheric connectivity should affect TTD features.

Let us consider a concrete example. A recent important finding is that, in mammals, two possible routes of interhemispheric communication exist. While eutherian mammals evolved a corpus callosum for interhemispheric communication, non-eutherian mammals, i.e., monotremes and marsupials (e.g., platypus and kangaroo) do not have this—they use other, potential slower routes for interhemispheric communication (Suárez et al., 2018). In other words, the CC evolved within the last 120 million years, so that non-eutherian mammals and birds use non-callosal structures for interhemispheric communication (Rogers et al., 2013). For instance, monotremes and marsupials mostly use the anterior commissure (see **Figure 1**; Aboitiz and Montiel, 2003).

This "discontinuity" in mammals provides a powerful testbed for our proposed TTD-CC link; in particular, even in behavioral experiments, we expect to find an evolutionary jump in the TTD phenotype. Within mammals, one often finds pairs of species that have convergently evolved similar behavioral traits or anatomy from two separate ancestors, one marsupial the other placental. Examples of this are, respectively: sugar gliders vs. flying squirrels, marsupial vs. non-marsupial moles, echidnas vs. porcupines. Based on our hypothesis, one could compare TTD abilities between pairs of species, expecting more developed ones in eutherian mammals. Even within eutherians there may be pairs of closely related species one of which duets while the other does not, featuring a variety of social and parental care systems. One could study variation in CC size between pairs of duetting and phylogenetically close non-duetting placental mammals. More generally and going beyond this admittedly simple dichotomy, one could study TTD acoustic phenotypes in species for which tractography data or measures of CC thickness exist, potentially expecting a positive correlation between the two.

In addition, non-invasive cognitive neuroscience techniques could be used to either measure or disrupt interhemispheric connectivity and relate this to behavioral TTD markers. For instance, electroencephalography (EEG) can be used to measure brain responses to sounds or phonatory actions. It appears that interhemispheric EEG coherence correlates with the size of CC and strongly decreases when the CC is damaged (Nielsen et al., 1993; Pogarell et al., 2005). Non-invasive EEG could then be employed in TTD experiments to test whether, as we hypothesize, interhemispheric EEG coherence will positively correlate with behavioral metrics of well-coordinated TTD. In addition, and to obtain more causal inference, techniques such as magnetic stimulation could be used to disrupt TTD patterns and therefore to obtain mechanistic explanations (Fitzgerald et al., 2002; Voineskos et al., 2010).

While we focus on mammals here, it is important to mention that several avian species have very developed TTD capacities (Mann et al., 2006; Pika et al., 2018; Benichov and Vallentin, 2020; Coleman et al., 2021; Kishimoto and Seki, 2022; Norton et al., 2022). Birds achieve this fine timing via developed intra-hemispheric connectivity, often resulting in motor inhibition to avoid overlap (Norton et al., 2022). However, similarly to non-eutherian mammals, birds do not have a CC. How do birds achieve impressive TTD without a CC? We hypothesize that TTD have different neuroanatomical bases in birds and eutherian mammals. While in humans the CC is responsible, and grants high-speed transmission, for most interhemispheric excitation and inhibition, birds rely on non-callosal structures (Rogers et al., 2013), which may still grant fast communication given the relatively smaller brain sizes and hence long distances (Ringo et al., 1994). In other words, we hypothesize that fast interhemispheric connection is achieved in birds via smaller brains and in eutherian mammals via the CC granting similar cross-brain speed irrespective of size (Phillips et al., 2015). This indirectly predicts lack of, or slow, TTD in non-eutherian mammals (and speculatively in large dinosaurs) which have brains on average larger than extant birds but also no CC (Weisbecker and Goswami, 2010; Naumann, 2015; Font et al., 2019).

CONCLUSION

Our perspective aims at spurring more mammalian TTD research at the behavior-cognition-neuroscience interface. We highlight the value of back-forth human-animal translational approaches, especially important in TTD because the three disciplines have partly solved three parts of the puzzle. The path we propose might become easier now that the field of diffusion analyses is rapidly advancing (e.g., Berns et al., 2015). Compared to a few years ago, diffusion analyses are more sophisticated and their results are getting close to what is observed *in vivo*. White matter diffusion analysis research may help pinpoint TTD behaviors to specific aspects of micro- (density) or microstructure (volume) and to biologically plausible metrics of connectivity (e.g., fiber bundle capacity). We predict that humans will end up being one example within a plethora of mammalian instances of TTD: The "neural phylogenies" approach we suggest will hopefully provide a roadmap for future integrative work.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117). The Comparative Bioacoustics Group is funded by Max Planck Group Leader funding to AR.

REFERENCES

Aboitiz, F. (2017). *A Brain for Speech: A View From Evolutionary Neuroanatomy*. Berlin: Springer.

Aboitiz, F., and Montiel, J. (2003). One hundred million years of interhemispheric communication: the history of the corpus callosum. *Braz. J. Med. Biol. Res.* 36, 409–420. doi: 10.1590/s0100-879x2003000400002

Aboitiz, F., Scheibel, A. B., Fisher, R. S., and Zaidel, E. (1992). Fiber composition of the human corpus callosum. *Brain Res.* 598, 143–153. doi: 10.1016/0006-8993(92)90178-C

Anichini, M., de Heer Kloots, M., and Ravignani, A. (2020). Interactive rhythms in the wild, in the brain, and in silico. *Can. J. Exp. Psychol.* 74:170. doi: 10.1037/cep0000224

Badaruddin, D. H., Andrews, G. L., Bolte, S., Schilmoeller, K. J., Schilmoeller, G., Paul, L. K., et al. (2007). Social and behavioral problems of children with agenesis of the corpus callosum. *Child Psychiatry Hum. Dev.* 38, 287–302. doi: 10.1007/s10578-007-0065-6

Bartha-Doering, L., Kollndorfer, K., Schwartz, E., Fischmeister, F. P. S., Alexopoulos, J., Langs, G., et al. (2020). The role of the corpus callosum in language network connectivity in children. *Dev. Sci.* 24:e13031. doi: 10.1111/desc.13031

Bartha-Doering, L., Schwartz, E., Kollndorfer, K., Fischmeiser, F. P. S., Novak, A., Langs, G., et al. (2021). Effect of corpus callosum agenesis on the language network in children and adolescents. *Brain Struc. Funct.* 226, 701–713. doi: 10.1007/s00429-020-02203-6

Beens, K. D. (1995). The Effects of Interactive Songs on Syllable Production in a Child with Agenesis of the Corpus Callosum.

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., and Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* 8, 1148–1150. doi: 10.1038/nn1516

Benichov, J. I., and Vallentin, D. (2020). Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-019-13938-0

Berns, G. S., Cook, P. F., Foxley, S., Jbabdi, S., Miller, K. L., and Marino, L. (2015). Diffusion tensor imaging of dolphin brains reveals direct auditory pathway to temporal lobe. *Proc. R. Soc. B Biol. Sci.* 282:20151203. doi: 10.1098/rspb.2015.1203

Bloom, J. S., and Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: excitation or inhibition? *Neuropsychol. Rev.* 15, 59–71. doi: 10.1007/s11065-005-6252-y

Bridgman, M. W., Brown, W. S., Spezio, M. L., Leonard, M. K., Adolphs, R., Paul, L. P., et al. (2014). Facial emotion recognition in agenesis of the corpus callosum. *J. Neurodev. Disord.* 6:32. doi: 10.1186/1866-1955-6-32

Brown, W. S., Paul, L. K., Symington, M., and Dietrich, R. (2005). Comprehension of humor in primary agenesis of the corpus callosum. *Neuropsychologia* 43, 906–916. doi: 10.1016/j.neuropsychologia.2004.09.008

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron* 80, 751–764. doi: 10.1016/j.neuron.2013.10.002

Casanova, M. F., El-Baz, A., Mott, M., Mannheim, G., Hassan, H., Fahmi, R., et al. (2009). Reduced gyral window and corpus callosum size in autism: possible macroscopic correlates of a minicolumnopathy. *J. Autism Dev. Disord.* 39, 751–764. doi: 10.1007/s10803-008-0681-4

Clarke, J. M., and Zaidel, E. (1994). Anatomical-behavioral relationships: corpus callosum morphometry and hemispheric specialization. *Behav. Brain Res.* 64, 185–202. doi: 10.1016/0166-4328(94)90131-7

Coleman, M. J., Day, N. F., Rivera-Parra, P., and Fortune, E. S. (2021). Neurophysiological coordination of duet singing. *Proc. Natl. Acad. Sci.* 118:e2018188118. doi: 10.1073/pnas.2018188118

Corballis, M. C., Corballis, P. M., and Fabri, M. (2004). Redundancy gain in simple reaction time following partial and complete callosotomy. *Neuropsychologia* 42, 71–81. doi: 10.1016/s0028-3932(03)00152-0

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., et al. (2021). Rhythm in dyadic interactions. *Philos. Trans. R. Soc. B* 376:20200337. doi: 10.1098/rstb.2020.0337

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., and Manser, M. (2018). Vocal turn-taking in meerkat group calling sessions. *Curr. Biol.* 28, 3661–3666. doi: 10.1016/j.cub.2018.09.065

Dougherty, R. F., Ben-Shachar, M., Deutsch, G. K., Hernandez, A., Fox, G. R., and Wandell, B. A. (2007). Temporal-callosal pathway diffusivity predicts phonological skills in children. *Pro. Natl. Acad. Sci.* 104, 8556–8561. doi: 10.1073/pnas.0608961104

Elmer, S., Hänggi, J., and Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Struc. Funct.* 221, 331–344. doi: 10.1007/s00429-014-0910-x

Fabri, M., Polonara, G., Pesce, M. D., Quattrini, A., Salvolini, U., and Manzoni, T. (2001). Posterior corpus callosum and interhemispheric transfer of somatosensory information: an fMRI and neuropsychological study of a partially callosotomized patient. *J. Cogn. Neurosci.* 13, 1071–1079. doi: 10.1162/089892901753294365

Fitzgerald, P. B., Brown, T. L., Daskalakis, Z. J., DeCastella, A., and Kulkarni, J. (2002). A study of transcallosal inhibition in schizophrenia using transcranial magnetic stimulation. *Schizophr. Res.* 56, 199–209. doi: 10.1016/S0920-9964(01)00222-5

Fitzpatrick, P., Frazier, J. A., Cochran, D. M., Mitchell, T., Coleman, C., and Schmidt, R. C. (2016). Impairments of social motor synchrony evident in autism spectrum disorder. *Front. Psychol.* 7:1323. doi: 10.3389/fpsyg.2016.01323

Font, E., García-Roa, R., Pincheira-Donoso, D., and Carazo, P. (2019). Rethinking the effects of body size on the study of brain size evolution. *Brain Behav. Evol.* 93, 182–195. doi: 10.1159/000501161

Friederici, A. D., and Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang.* 89, 267–276. doi: 10.1016/S0093-934X(03)00351-1

Friederici, A. D., Cramon, D. Y., and Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron* 53, 135–145. doi: 10.1016/j.neuron.2006.11.020

Fröhlich, M. (2017). Taking turns across channels: conversation-analytic tools in animal communication. *Neurosci. Biobehav. Rev.* 80, 201–209. doi: 10.1016/j.neubiorev.2017.05.005

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication. Does the corpus callosum enable the human condition? *Brain* 123, 1293–1326. doi: 10.1093/brain/123.7.1293

Gooijers, J., and Swinnen, S. P. (2014). Interactions between brain structure and behavior: the corpus callosum and bimanual coordination. *Neurosci. Biobehav. Rev.s* 43, 1–19. doi: 10.1016/j.neubiorev.2014.03.008

Greenfield, M. D., Honing, H., Kotz, S. A., and Ravignani, A. (2021). Synchrony and rhythm interaction: from the brain to behavioural ecology. *Philos. Trans. R. Soc. B* 376:20200324. doi: 10.1098/rstb.2020.0324

Güntürkün, O., Ströckens, F., and Ocklenburg, S. (2020). Brain lateralization: a comparative perspective. *Physiol. Rev.* 100, 1019–1063. doi: 10.1152/physrev.00006.2019

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., et al. (2018). Childhood music training induces change in micro and macroscopic brain structure: results from a longitudinal study. *Cereb. Cortex* 28, 4336–4347. doi: 10.1093/cercor/bhx286

Hofer, S., and Frahm, J. (2006). Topography of the human corpus callosum revisited– comprehensive fibre tractography using diffusion tensor magnetic resonance imaging. *NeuroImage* 32, 989–994. doi: 10.1016/j.neuroimage.2006.05.044

Holler, J., Kendrick, K. H., Casillas, M., and Levinson, S. C. (2015). Editorial: turn-taking in human communicative interaction. *Front. Psychol.* 6:1919. doi: 10.3389/fpsyg.2015.01919

Horowitz, A., Barazany, D., Tavor, I., Bernstein, M., Yovel, G., and Assaf, Y. (2015). In vivo correlation between axon diameter and conduction velocity in the human brain. *Brain Struc. Funct.* 220, 1777–1788. doi: 10.1007/s00429-014-0871-0

Huang, H., Zhang, J., Jiang, H., Wakana, S., Poetscher, L., Miller, M. I., et al. (2005). DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of corpus callosum. *NeuroImage* 26, 195–205. doi: 10.1016/j.neuroimage.2005.01.019

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025. doi: 10.1523/JNEUROSCI.5118-08.2009

Kaur, M., Srinivasan, S. M., and Bhat, A. N. (2018). Comparing motor performance, praxis, coordination, and interpersonal synchrony between children with and without Autism *Spectrum* disorder (ASD). *Res. Dev. Disabil.* 72, 79–95. doi: 10.1016/j.ridd.2017.10.025

Kishimoto, R., and Seki, Y. (2022). Response timing of budgerigars in a turn-taking task under operant conditioning. *Behav Process.* 198:104638. doi: 10.1016/j.beproc.2022.104638

Kotz, S. A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi: 10.1016/j.tics.2010.06.005

Kotz, S. A., and Schwartze, M. (2016). "Motor timing and sequencing in speech production: a general purpose framework," in *Neurobiology of Language*, eds G. S. Hickock and S. L. Small (London: Elsevier), 717–723. doi: 10.1016/B978-0-12-407794-2.00057-2

Laìbadi, B., and Beke, A. M. (2017). Mental state understanding in children with agenesis of the corpus callosum. *Front. Psychol.* 8:94. doi: 10.3389/fpsyg.2017.00094

Lee, D. J., Chen, Y., and Schlaug, G. (2003). Corpus callosum: musician and gender effects. *Neuroreport* 14, 205–209. doi: 10.1097/00001756-200302100-00009

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Loui, P. (2018). Rapid and flexible creativity in musical improvisation: review and a model. *Ann. N. Y. Acad. Sci.* 1423, 138–145. doi: 10.1111/nyas.13628

Lumaca, M., Baggio, G., and Vuust, P. (2021). White matter variability in auditory callosal pathways contributes to variation in the cultural transmission of auditory symbolic systems. *Brain Struct. Funct.* 226, 1943–1959. doi: 10.1007/s00429-021-02302-y

Mann, N. I., Dingess, K. A., and Slater, P. J. B. (2006). Antiphonal four-part synchronized chorusing in a neotropical wren. *Biol. Lett.* 2, 1–4. doi: 10.1098/rsbl.2005.0373

Naumann, R. K. (2015). Even the smallest mammalian brain has yet to reveal its secrets. *Brain Behav. Evol.* 85, 1–3. doi: 10.1159/000375438

Nielsen, T., Montplaisir, J., and Lassonde, M. (1993). Decreased interhemispheric EEG coherence during sleep in agenesis of the corpus callosum. *Eur. Neurol.* 33, 173–176. doi: 10.1159/000116928

Norton, P., Benichov, J. I., Pexirra, M., Schreiber, S., and Vallentin, D. (2022). A feedforward inhibitory premotor circuit for auditory–vocal interactions in zebra finches. *Proc. Natl. Acad. Sci.* 119:e2118448119. doi: 10.1073/pnas.2118448119

Okobi, D. E. Jr., Banerjee, A., Matheson, A. M., Phelps, S. M., and Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988. doi: 10.1126/science.aau9480

Ozturk, A. H., Tascioglu, B., Aktekin, M., Kurtoglu, Z., and Erden, I. (2002). Morphometric comparison of the human corpus callosum in professional musicians and non-musicians by using in vivo magnetic resonance imaging. *J. Neuroradiol.* 29, 29–34.

Park, H. J., Kim, J. J., Lee, S. K., Seok, J. H., Chun, J., Kim, D. I., et al. (2008). Corpus callosal connection mapping using cortical gray matter parcellation and DT-MRI. *Hum. Brain Mapp.* 29, 503–516. doi: 10.1002/hbm.20314

Paul, L. K., Van Lancker-Sidtis, D., Schieffer, B., Dietrich, R., and Brown, W. S. (2003). Communicative deficits in agenesis of the corpus callosum: nonliteral language and affective prosody. *Brain Lang.* 85, 313–324. doi: 10.1016/s0093-934x(03)00062-2

Phillips, K. A., Stimpson, C. D., Smaers, J. B., Raghanti, M. A., Jacobs, B., Popratiloff, A., et al. (2015). The corpus callosum in primates: processing speed of axons and the evolution of hemispheric asymmetry. *Proc. R. Soc. B Biol. Sci.* 282:20151535. doi: 10.1098/rspb.2015.1535

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* 285:20180598. doi: 10.1098/rspb.2018.0598

Pogarell, O., Teipel, S. J., Juckel, G., Gootjes, L., Möller, T., Bürger, K., et al. (2005). EEG coherence reflects regional corpus callosum area in Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 76, 109–111. doi: 10.1136/jnnp.2004.036566

Pollmann, S., Maertens, M., von Cramon, D. Y., Lepsien, J., and Hugdahl, K. (2002). Dichotic listening in patients with splenial and nonsplenial callosal lesions. *Neuropsychology* 16:56. doi: 10.1037//0894-4105.16.1.56

Rauch, R. A., and Jinkins, J. R. (1994). "Magnetic resonance imaging of corpus callosum dysgensis," in *Callosal Agenesis: a Natural Split brain?*, eds M. Lassonde and M. A. Jeeves (New York, NY: Plenum Press), 83–95. doi: 10.1007/978-1-4613-0487-6_11

Ravignani, A. (2019). Timing of antisynchronous calling: a case study in a harbor seal pup (*Phoca vitulina*). *J. Comp. Psychol.* 133:272. doi: 10.1037/com0000160

Ravignani, A., Verga, L., and Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Ann. N. Y. Acad. Sci.* 1453, 12–21.

Ringo, J. L., Doty, R. W., Demeter, S., and Simard, P. Y. (1994). Time is of the essence: a conjecture that hemispheric specialization arises from interhemispheric conduction delay. *Cereb. Cortex* 4, 331–343. doi: 10.1093/cercor/4.4.331

Risse, G. L., Gates, J., Lund, G., Maxwell, R., and Rubens, A. (1989). Interhemispheric transfer in patients with incomplete section of the corpus callosum: anatomic verification with magnetic resonance imaging. *Arch. Neurol.* 46, 437–443. doi: 10.1001/archneur.1989.00520400097026

Rogers, L. J., Vallortigara, G., and Andrew, R. J. (2013). *Divided Brains: the Biology and Behaviour of Brain Asymmetries*. Cambridge, MA: Cambridge University Press.

Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V. M., and Friederici, A. D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain* 133, 2643–2655. doi: 10.1093/brain/awq231

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., and Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia* 33, 1047–1055. doi: 10.1016/0028-3932(95)00045-5

Schlaug, G., Norton, A., Overy, K., and Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Ann. N. Y. Acad. Sci.* 1060, 219–230. doi: 10.1196/annals.1360.015

Sperry, R. W. (1961). Cerebral organization and behavior. *Science* 133, 1749–1757. doi: 10.1126/science.133.3466.1749

Steele, C. J., Bailey, J. A., Zatorre, R. J., and Penhune, V. B. (2013). Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *J. Neurosci.* 33, 1282–1290. doi: 10.1523/JNEUROSCI.3578-12.2013

Stickles, J. L., Schilmoeller, G. L., and Schilmoeller, K. J. (2002). A 23-year review of communication development in an individual with agenesis of the corpus callosum. *Int. J. Disabil. Dev. Educ.* 49, 367–383. doi: 10.1080/1034912022000028349

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Su, W.-C., Culotta, M., Mueller, J., Tsuzuki, D., Pelphrey, K., and Bhat, A. (2020). Differences in cortical activation patterns during action observation, action execution, and interpersonal synchrony between children with or without autism spectrum disorder (ASD): an fNIRS pilot study. *PLoS One* 15:e0240301. doi: 10.1371/journal.pone.0240301

Suárez, R., Paolino, A., Fenlon, L. R., Morcom, L. R., Kozulin, P., Kurniawan, N. D., et al. (2018). A pan-mammalian map of interhemispheric brain connections predates the evolution of the corpus callosum. *Proc. Natl. Acad. Sci.* 115, 9622–9627. doi: 10.1073/pnas.1808262115

Takahashi, D. Y., Fenley, A. R., and Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150370. doi: 10.1098/rstb.2015.0370

Tu, S., Doherty, D., Schilmoeller, K. J., and Schilmoeller, G. L. (2009). "Agenesis of the corpus callosum: a literature review," in *International Review of Research in Mental Retardation*, Vol. 38, eds S. Tu, D. Doherty, K. J. Schilmoeller, and G. L. Schilmoe (Amsterdam: Elsevier Inc), 171–193. doi: 10.1016/S0074-7750(08)38007-0

Voineskos, A. N., Farzan, F., Barr, M. S., Lobaugh, N. J., Mulsant, B. H., Chen, R., et al. (2010). The role of the corpus callosum in transcranial magnetic stimulation induced interhemispheric signal propagation. *Biol. Psychiatry* 68, 825–831.

Weisbecker, V., and Goswami, A. (2010). Brain size, life history, and metabolism at the marsupial/placental dichotomy. *Proc. Natl. Acad. Sci.* 107, 16216–16221.

Westerhausen, R., Kreuder, F., Woerner, W., Huster, R. J., Smit, C. M., Schweiger, E., et al. (2006). Interhemispheric transfer time and structural properties of the corpus callosum. *Neurosci. Lett.* 409, 140–145.

Wing, A. M., Endo, S., Bradbury, A., and Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *J. R. Soc. Interface* 11:20131125.

Zarei, M., Johansen-Berg, H., Smith, S., Ciccarelli, O., Thompson, A. J., and Matthews, P. M. (2006). Functional anatomy of interhemispheric cortical connections in the human brain. *J. Anat.* 209, 311–320.

Zeng, T., Przysinda, E., Pfeifer, C., Arkin, C., and Loui, P. (2017). White matter connectivity reflects success in musical improvisation. *BioRxiv* 218024. doi: 10.1101/218024

Zhang, J. D., Susino, M., McPherson, G. E., and Schubert, E. (2020). The definition of a musician in music psychology: a literature review and the six-year rule. *Psychol. Music* 48, 389–409. doi: 10.1177/0305735618804038

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 02 August 2022
DOI 10.3389/fevo.2022.910121

OPEN ACCESS

EDITED BY
Mingbo Yin,
Fudan University, China

REVIEWED BY
Marco Gamba,
University of Turin, Italy
Oliver M. Beckers,
Murray State University, United States

*CORRESPONDENCE
Isabel Comella
ic265@cornell.edu

SPECIALTY SECTION
This article was submitted to Behavioral and Evolutionary Ecology, a section of the journal Frontiers in Ecology and Evolution

RECEIVED 31 March 2022
ACCEPTED 27 June 2022
PUBLISHED 02 August 2022

CITATION
Comella I, Tasirin JS, Klinck H, Johnson LM and Clink DJ (2022) Investigating note repertoires and acoustic tradeoffs in the duet contributions of a basal haplorrhine primate.
Front. Ecol. Evol. 10:910121.
doi: 10.3389/fevo.2022.910121

Investigating note repertoires and acoustic tradeoffs in the duet contributions of a basal haplorrhine primate

Isabel Comella[1]*, Johny S. Tasirin[2], Holger Klinck[1], Lynn M. Johnson[3] and Dena J. Clink[1]

[1]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States, [2]Faculty of Agriculture, Sam Ratulangi University, Manado, Indonesia, [3]Cornell Statistical Consulting Unit, Cornell University, Ithaca, NY, United States

Acoustic communication serves a crucial role in the social interactions of vocal animals. Duetting—the coordinated singing among pairs of animals—has evolved independently multiple times across diverse taxonomic groups including insects, frogs, birds, and mammals. A crucial first step for understanding how information is encoded and transferred in duets is through quantifying the acoustic repertoire, which can reveal differences and similarities on multiple levels of analysis and provides the groundwork necessary for further studies of the vocal communication patterns of the focal species. Investigating acoustic tradeoffs, such as the tradeoff between the rate of syllable repetition and note bandwidth, can also provide important insights into the evolution of duets, as these tradeoffs may represent the physical and mechanical limits on signal design. In addition, identifying which sex initiates the duet can provide insights into the function of the duets. We have three main goals in the current study: (1) provide a descriptive, fine-scale analysis of Gursky's spectral tarsier (*Tarsius spectrumgurskyae*) duets; (2) use unsupervised approaches to investigate sex-specific note repertoires; and (3) test for evidence of acoustic tradeoffs in the rate of note repetition and bandwidth of tarsier duet contributions. We found that both sexes were equally likely to initiate the duets and that pairs differed substantially in the duration of their duets. Our unsupervised clustering analyses indicate that both sexes have highly graded note repertoires. We also found evidence for acoustic tradeoffs in both male and female duet contributions, but the relationship in females was much more pronounced. The prevalence of this tradeoff across diverse taxonomic groups including birds, bats, and primates indicates the constraints that limit the production of rapidly repeating broadband notes may be one of the few 'universals' in vocal communication. Future carefully designed playback studies that investigate the behavioral response, and therefore potential information transmitted in duets to conspecifics, will be highly informative.

KEYWORDS

primate vocalization, duetting, acoustic tradeoff, Gursky's spectral tarsier, universals

Introduction

Animal vocal communication serves several social functions, including mate attraction, species recognition, territory and mate defense, and threat notification (Wilkins et al., 2013; Price et al., 2015). Natural selection, sexual selection, and neutral evolution are the three primary processes by which observed patterns of differentiation in acoustic signals form across populations. These processes can ultimately lead to the formation of new species, a process known as speciation (Jones, 1997; Wilkins et al., 2013; Blute, 2019; Shuker and Kvarnemo, 2021). The study of animal vocalization systems sets the groundwork for isolating and defining common patterns of phenotypic variation that link species across vast evolutionary distances, providing us with a better understanding of common ancestral constraints that have guided evolution (Derryberry et al., 2012).

The various evolutionary processes mentioned above can drive differentiation across populations, but the selection for certain traits is not without limits. Eventually a trait, such as the loudness of a call or the frequency bandwidth of a note, can no longer be shaped by natural or sexual selection. These limits may be imposed by neural or biomechanical constraints, or a combination of the two (Wilkins et al., 2013). Acoustic neural constraints are limits imposed on vocal production by the capacity of the neural pathways in the brain to produce acoustic signals (Römer, 1993; DeVoogd, 2004; Fitch et al., 2016), while biomechanical constraints are imposed by the physical and morphological composition of the vocal production structure(s), including lung capacity (Fedurek et al., 2017), mouth or beak size and shape (Derryberry et al., 2012), and laryngeal configuration and motor control (Lieberman et al., 1969; Podos, 1996; Fedurek et al., 2017). Such constraints limit the physical abilities of the individual in such a way that reaching these constraints presumably conveys information about the individual's fitness. It follows that if these constraints are honest indicators of caller quality, individuals who are capable of nearing or reaching the evolutionary limit of a certain trait will be more attractive to potential mates, as successfully exhibiting costly traits can be an indicator of a high-quality individual (Clutton-Brock and Albon, 1979; Reby and McComb, 2003; Terleph et al., 2016; Sun et al., 2021).

One constraint of interest in acoustic signals is the tradeoff between the rate of trill notes and the bandwidth of those notes. In order to produce high frequency trills with wide bandwidths, individuals must make rapid and comprehensive vocal modifications, which may be physically demanding (Podos, 1996; Ballentine et al., 2004). The presumed energetic and/or morphological constraints on modifications of the vocal tract result in high frequency trills with relatively narrower bandwidths. This results in a triangular distribution on a graph, where at low trill frequencies, both wide and narrow bandwidths are possible, while at higher trill frequencies, only narrower bandwidths are exhibited (Derryberry et al., 2012; Wilson et al., 2014). Thus, it appears there are limits to the properties (such as frequency, bandwidth, or amplitude) of acoustic communication. While the limits themselves may vary based on species and vocalization type, the presence of some kind of acoustic tradeoff is thought to be near universal (Podos, 1997).

Acoustic tradeoffs such as this one have been studied in birds, mice, bats, and primates (Podos, 1996; Ballentine et al., 2004; Pasch et al., 2011; Derryberry et al., 2012; Wilson et al., 2014; Clink et al., 2018; Sun et al., 2021). One study examined the ability of a species of sparrow to learn a song containing high frequency trills with artificially broad bandwidths (Podos, 1996). While numerous studies have shown that songbirds are able to learn species-specific songs with high accuracy (Thorpe, 1961; Marler, 1970; Brainard and Doupe, 2002), the individuals exposed to the artificial song were unable to learn it with high fidelity, indicating that morphological acoustic tradeoffs are likely responsible for this inability to replicate an artificial song (Podos, 1996). In Himalayan leaf-nosed bats (*Hipposideros armiger*), the tradeoff between trill note frequency and individual note bandwidth was found to reflect the quality of the caller; higher quality callers (as indicated using body mass as a proxy) were able to produce higher frequency trills with broader bandwidths (Sun et al., 2021). One of the few studies examining this tradeoff in primates found that female Northern gray gibbons (*Hylobates funereus*) exhibited vocal patterns consistent with this constraint: an increase in trill rate was correlated with a decrease in note bandwidth (Clink et al., 2018). More studies on a wide variety of taxa are needed to determine the extent to which this acoustic tradeoff is present in animal vocalizations, and if this tradeoff represents one of the few documented universals in vertebrate vocal communication.

A note repertoire is an itemization of different note types produced by a species, while a vocal repertoire is an expansion of a note repertoire by the addition of combinations of the individual note types (Clarke et al., 2006). Repertoire descriptions are often made more robust by the addition of descriptions of the various contexts in which each vocalization type is made, but a crucial first step is analyzing the repertoire across individuals in the same context (Clarke et al., 2006; Blue, 2020). The compilation and thorough definition of a species' repertoire is a straight-forward yet powerful mode of communication system analysis. While interesting in their own right, vocal and note repertoires can reveal differences and similarities on multiple levels of analysis, including species, sex, and individual, and can reveal how species transmit and receive information about external states such as the presence and type of predator (Clarke et al., 2006; Price et al., 2015; Segbroeck et al., 2017; Sainburg et al., 2020). Large repertoires are presumed to be costly to develop, and in some cases repertoire size reflects aspects of caller quality including age, condition, and parasites (Balsby and Hansen, 2010). Additionally, the establishment of note and vocal repertoires provides the groundwork necessary for further studies and analyses of the vocal communication patterns of the focal species (Blue, 2020; Sainburg et al., 2020). The accumulation of comprehensive vocal repertoires for many taxa is vital to our complete understanding of the vocal communication patterns, functions, and contexts of those taxa, and will be invaluable in informing future studies, especially in vocal but otherwise cryptic or elusive species.

Comprehensive note and vocal repertoires have been compiled for many taxa, including many species of birds (Ficken et al., 1978) and non-human primates (Winter et al., 1966; Gros-Louis et al., 2008; Blue, 2020). Commonly, vocalizations are classified into groups based on physical characteristics observable in a spectrogram, including duration, volume, note frequency, and note shape. While the vocal repertoire size (number of vocalization categories) is often used as an indicator of vocal complexity, repertoire size alone does not provide any information about the functions of the call types or associated contexts (Blue, 2020). Studies that rely solely on acoustic data provide valuable analyses of note types and classifications, but the addition of behavioral observations can allow for the inference of vocalization function by providing social and environmental context (Winter et al., 1966; Ficken et al., 1978; Gros-Louis et al., 2008). Important information can also be gained by understanding the distribution of call types based on individual maturity level, social affiliation, and sex (Clarke et al., 2006; Nousek et al., 2006; Clink et al., 2017; Andrieu et al., 2020).

Gibbons (Hylobatidae), indris (Indriidae), titi monkeys (Callicebinae), and some tarsier species (Tarsiidae) are pair-living primates that produce species- and sex-specific coordinated vocalizations between mated pairs (Haimoff, 1986; Geissmann, 2002; De Gregorio et al., 2022). Reproductive pairs of *Lepilemur edwardsi* also show coordinated vocal exchanges, but these are not considered proper songs (Méndez-Cárdenas and Zimmermann, 2009). The precise functions of duetting in these various species are, for the most part, yet unknown, although there are a number of hypothesized duet functions, most notably the advertisement and strengthening of the pairbond (Smith, 1994; Geissmann, 1999; Geissmann and Orgeldinger, 2000), territorial communications with extrapair individuals (Clink et al., 2020), and reunion of the mated pair after a period of separation, such as occurs with individual foraging (Méndez-Cárdenas and Zimmermann, 2009). In a review on 59 duetting avian bird species, it was shown that duets used solely for extra-pair communication were more likely to consist of sex-specific notes. The authors noted that the sample size for sex-specific number of notes was too small for statistical analysis, but the median number of notes for males and females in those species with available data was similar (Dahlin and Benedict, 2014). It is unclear whether sex-specificity (or lack thereof) in primate duets is related to differences in function of the duets, as most gibbon species (Geissmann, 2002), tarsiers (MacKinnon and MacKinnon, 1980), and indris (Giacoma et al., 2010) exhibit varying degrees of sex-specificity, whereas titi monkey duets do not (Clink et al., 2022).

In this study, we have three specific aims: (1) provide a descriptive, fine-scale analysis of tarsier duets, including information on which sex initiates the duet, duration of the duets, and total number of notes; (2) use unsupervised approaches to investigate sex-specific note repertoires; and (3) test for evidence of acoustic tradeoffs on the rate of note repetition and bandwidth of tarsier duet contributions. The purpose of aim (1) is purely descriptive, so we do not have any associated hypotheses or predictions. For aim (2), we hypothesize that due to the sex-specificity in duet contributions and the presumed extra-pair communication function of tarsier duets, the number of notes in the note repertoires will be sex-specific, following the trends seen in the other duetting primate species, such as gibbons and indris (Geissmann, 2002; Giacoma et al., 2010). For aim (3), we hypothesize that there are constraints in vocal production that make it difficult to produce broadband notes at a relatively fast rate, and therefore predict that in light of the evidence for acoustic tradeoffs in multiple taxa, including a species of non-human primate, tarsiers will also conform to these vocal patterns.

Materials and methods

Gursky's spectral tarsier

Gursky's spectral tarsier (*Tarsius spectrumgurskyae*; hereafter tarsiers) is a species of small, nocturnal primate endemic to the northern part of the island of Sulawesi in Indonesia (MacKinnon and MacKinnon, 1980; Gursky, 2000, 2002). They are the only faunivorous primate and survive on a diet of insects (Gursky, 2002). Tarsiers live in social groups generally consisting of one adult mated pair and two to four of their juvenile offspring (MacKinnon and MacKinnon, 1980). They are known to be highly territorial and occupy semi-overlapping home ranges (MacKinnon and MacKinnon, 1980). After a night of hunting, tarsiers will return to the same sleep tree or trees each morning, and the mated pair will perform a series of duets around sunrise, roughly between the hours of 0500 and 0600 local time (MacKinnon and MacKinnon, 1980; Gursky, 2000, 2002). Occasionally, juveniles will join in these coordinated vocal displays, in which cases the duets become choruses (Voigt et al., 2006; De Gregorio et al., 2022).

Study system and data collection

We collected focal and autonomous acoustic recordings of tarsiers during July and August of 2018 in Tangkoko National Park on the northeastern tip of the island of Sulawesi, Indonesia (**Figure 1**). We did not tag or label individual animals for identification in any way, so their reliable territoriality, fidelity to sleep trees, and minimally overlapping ranges allowed specific pairs and individuals to be distinguished. Tarsiers' general lack of fear of humans means that habituation was unnecessary and that alterations in behavior due to observer presence were minimal (MacKinnon and MacKinnon, 1980).

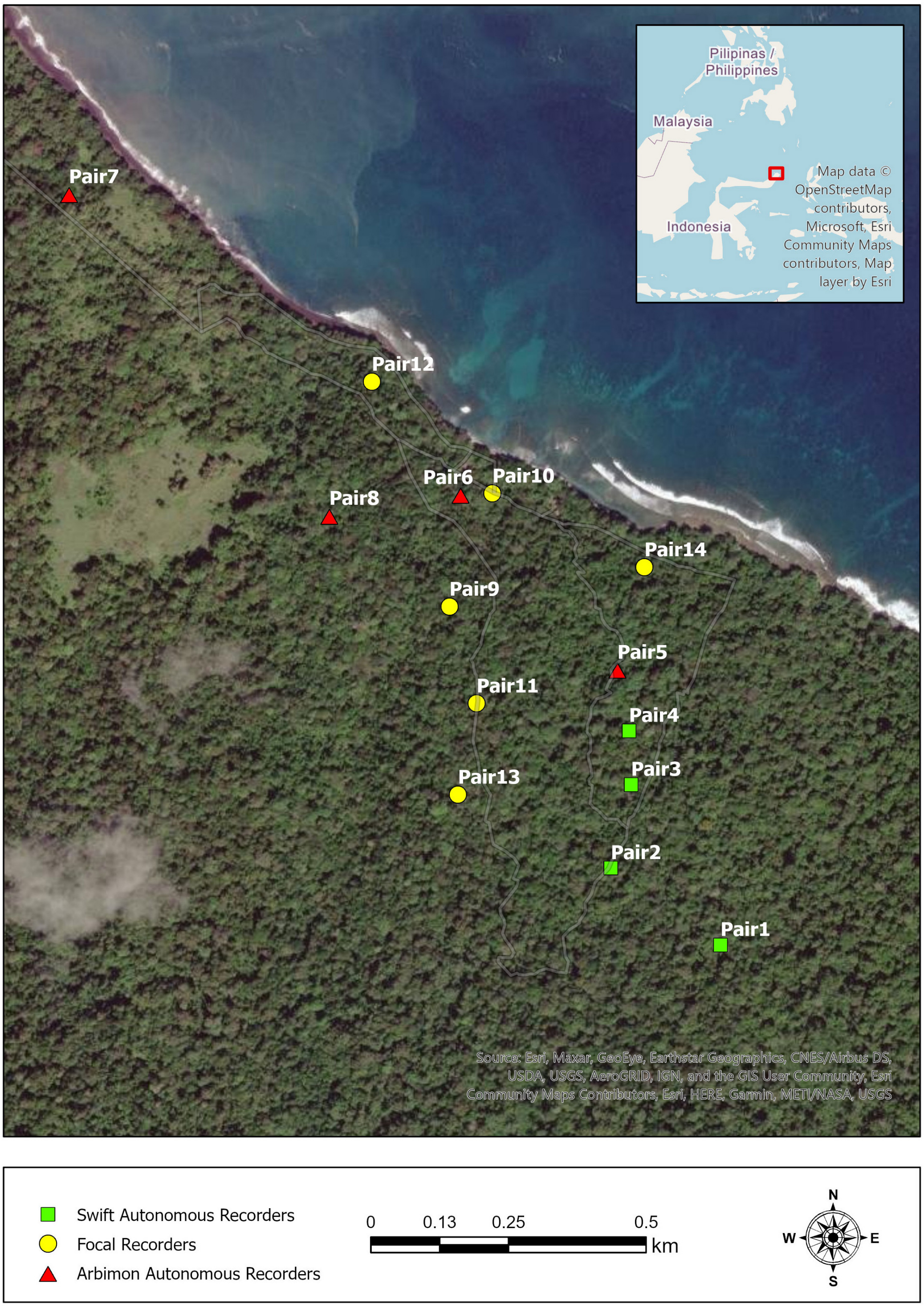
We used a RØDE NT-USB Condenser Microphone (Røde Microphones, Sydney, Australia) in conjunction with a 32 GB Apple iPad Air (Apple Inc., Cupertino, CA, United States) and the Voice Record Pro application (sampling rate of 44.1 kHz and 16 bits) for focal recordings. DC and a research assistant opportunistically took focal recordings in the early mornings. We took autonomous recordings *via* either an ARBIMON (Aide et al., 2013) portable recorder (44.1 kHz and 16 bits) or a Swift recorder (Koch et al., 2016) (48 kHz and 16 bits). ARBIMON units recorded daily from 1800 to 0600, while Swift units recorded 24 h continuously. The ARBIMON units had substantially reduced storage capabilities compared to the Swift units, which is why we recorded using different settings and recording schedules. Since tarsiers limit their duetting to a time window of approximately 1 h per day, the variable recording schedules had limited impact on our data collection capabilities. Different autonomous recording units may have different detection ranges due to variation in microphone sensitivities; we aimed to minimize these potential differences by only using high signal-to-noise ratio (SNR) recordings for analyses. Indeed distance from the animal to the recording device can influence the spectral feature estimates, so we limited our analyses to high-quality calls (>12 dB SNR), using high SNR as a proxy for recording distance (Zollinger et al., 2012). Each autonomous unit recorded from 2 to 7 days' worth of data. Differences in recording durations were due to battery and/or unit malfunction. It has been reported that tarsier duets can be heard up to 500 m by a human observer, but early field tests indicated that the detection range of the recording units for high-quality recordings was much less than that, and generally restricted to animals calling <50 meters from the recording unit (Gursky, 2015; Clink et al., 2019). We saved all recordings as Waveform Audio Files. ARBIMON recorders saved 1-h long files at a size of 317.5 MB, Swift recorders saved 40-min files at a size of 230.4 MB, and focal recordings files were of variable duration and size. We downsampled the 48 kHz to 44.1 kHz sound files using the open-source program Audacity (version 3.1.3) before further analysis. Full details of acoustic data collection methods can be found in Clink et al. (2019). We used only duets (as opposed to choruses or solos) and only those that showed a completed song progression in a spectrogram.

Acoustic analysis

We imported each sound file into Raven Pro v. 1.6 (K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Ithaca, NY, United States) and converted them into spectrograms using the following settings: a 1,600-sample Hann window, 3 dB filter bandwidth of 39.6 Hz, with a 2048-point discrete Fourier transform and 50% overlap, resulting in a time grid hop size of 18.1 ms and a frequency grid spacing of 21.5 Hz. IAC annotated all duets by hand using the selection table functionality in Raven Pro. For each note within the duet, we documented the begin time, end time, minimum frequency, maximum frequency, and sex of the individual. We were able to easily distinguish between male and female duet contributions given the sex-specific differences (MacKinnon and MacKinnon, 1980). Although generally the robust features in Raven are preferable as they reduce variability in intra-observer reliability in terms of how the annotation boxes are selected, they calculate the features based on the energy within the selection and are not appropriate when there is a substantial amount of overlap between signals from different individuals (Charif et al., 2010). Therefore, we calculated bandwidth based on the minimum and maximum frequency bounds of the annotation boxes. This required us to maintain the same brightness, contrast, and focus settings (brightness: 50; contrast: 50) to minimize variation in how annotation boxes were drawn. To calculate the rate of note output, we counted the number of notes emitted per 3-s. Previous analyses calculated note rate using 1-s (Clink et al., 2018), but we found that the rate of note output for tarsiers was relatively slow compared to previous studies (∼1 note per 1-s). Therefore, using a longer duration time bin allowed us to capture more variation in the rate of note output for tarsier duets. In order to allow our results to be compared to other results in this field, we have also standardized these rates into 1-s rates; so, although we used a longer time interval (3-s) to measure the rate, we divided these values by 3 so that our reported values could be used for cross-taxa comparisons. See **Figure 2** for representative spectrograms of tarsier duets and phrases and **Figure 3** indicating male and female contributions to the duet and a schematic of how we estimated note features for the present analysis.

Unsupervised analysis of note types

We aimed to identify the number of unique clusters or note types in the male and female tarsier duet contributions. To identify the number of unique clusters we used an unsupervised random forests framework that can be used to identify patterns in an unlabeled dataset (Breiman, 2001). We analyzed male and female notes separately from each other due to structural differences in their respective duet contributions.



Map data © OpenStreetMap contributors, Microsoft, Esri Community Maps contributors, Map layer by Esri

Source: Esri, Maxar, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community, Esri Community Maps Contributors, Esri, HERE, Garmin, METI/NASA, USGS

FIGURE 1
Map of recording locations of tarsier pairs in Tangkoko National Park, North Sulawesi, Indonesia. Each point denotes the recording location of a tarsier pair, and the shape of the points reflects the type of recorder used (see section "Materials and methods" for details).

We used the R programming environment to implement the random forest network using the "randomForest" package (Liaw and Wiener, 2002); we specified the number of trees equal to 10,000 and otherwise used the default settings. As input for the "randomForest" algorithm, we used four features computed from each note–minimum frequency (Hz),

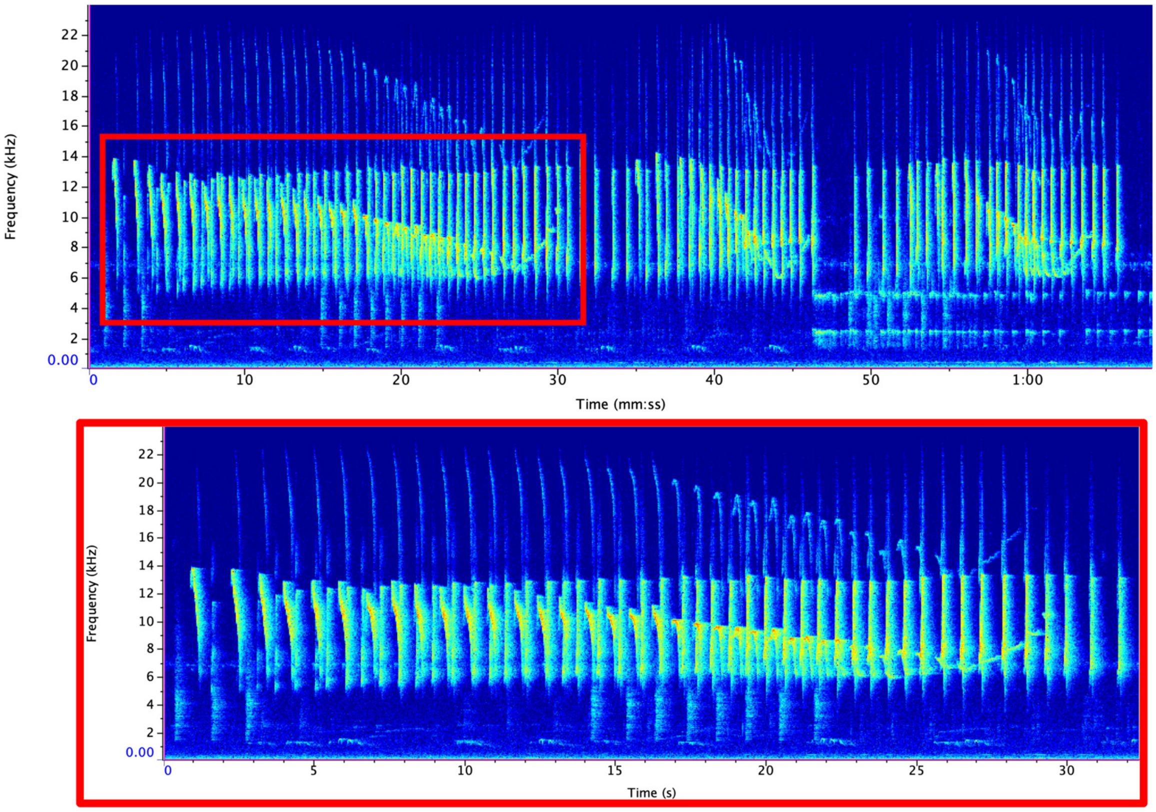


FIGURE 2
Representative spectrogram of a tarsier duet **(top)** and phrase **(bottom)**. A single duet **(top)** can be comprised of many phrases **(bottom)**. Phrases can vary in length but generally follow the structure shown above. Spectrograms were created using RavenPro with the same settings that were used for analysis (see text for details).

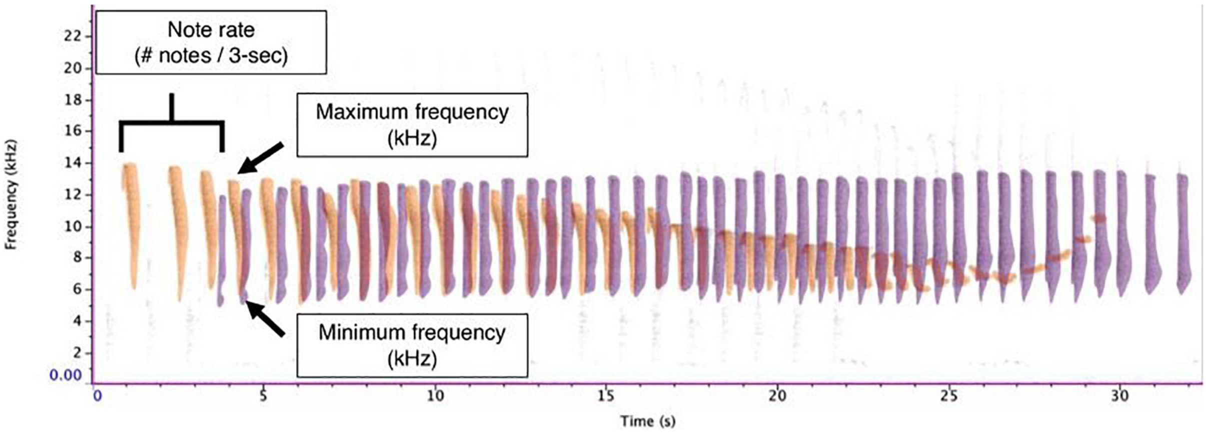


FIGURE 3
Exemplar of male and female duet contributions and analyzed features. The male duet contribution is shown in purple, and the female duet contribution is shown in orange. Note rate was calculated as the number of notes per 3-s interval. Note bandwidth was determined by subtracting the minimum frequency from the maximum frequency.

maximum frequency (Hz), bandwidth (Hz), and duration (s). This algorithm returns a dissimilarity metric for each observation which can be used to identify groupings within the data. To identify the optimal number of clusters in our dataset we applied k-medoids clustering (Kaufman and Rousseeuw, 2009) to the distance matrix output of the random

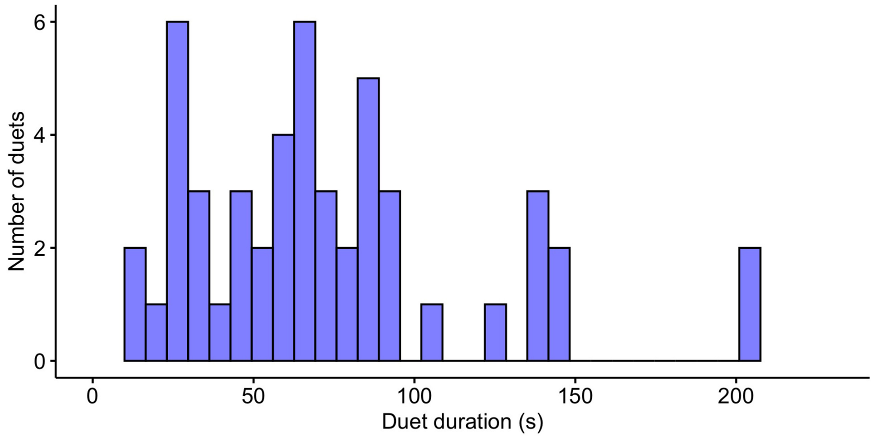


FIGURE 4
Histogram indicating the durations of tarsier duets (*N* = 50). Duets ranged in duration from 12.88 to 203.96 s and show a distribution with a slight right skew.

forest algorithm using the “pam” function in the “cluster” R package (Maechler et al., 2012). K-medoids is more robust version of K-means (Madhulatha, 2011). K-medoids requires the input of the number of clusters (k) so we ran the algorithm iteratively for values of k from 2 to 10 and then calculated a silhouette coefficient for each cluster solution. Silhouette coefficients range from −1 to 1 and provide a measure of how similar an object is relative to the established clusters; a higher silhouette coefficient indicates a more appropriate cluster solution (Rousseeuw, 1987). To identify the optimum number of clusters in our dataset we chose the cluster number with the highest silhouette coefficient. We used a uniform manifold learning technique (McInnes et al., 2018) for visualization of the results using the R package “umap” (Konopka, 2020). UMAP is a dimensionality reduction technique that has been used to effectively visualize differences in acoustic signals of multiple bird taxa (Parra-Hernández et al., 2020), forest soundscapes (Sethi et al., 2020) and female gibbon vocalizations (Clink and Klinck, 2021). We input a feature vector consisting of the four features estimated for each note into the UMAP algorithm, which returned two coordinates or two-dimensional embeddings that can be used to visualize clustering of note types within our dataset. In our study, we define gradation as the degree to which clusters are separated from each other–low gradation means high cluster separation (Wadewitz et al., 2015).

Acoustic tradeoffs statistical analysis

To investigate the relationship between note bandwidth and note rate we used a series of Bayesian multilevel models implemented using the R package “brms” (Bürkner, 2017a,b). The “brms” package provides an interface to the probabilistic programming language STAN (Carpenter et al., 2017). Due to the structural differences in male and female duet contributions we analyzed males and females separately. For both males and females, we created a series of three models. The first model, which we considered the null model, included note bandwidth as the outcome and a random effect for pair identity. The second model, which we used to test for evidence of acoustic tradeoffs, included note bandwidth as an outcome, note rate as a predictor, and pair identity as a random effect. The third model contained a random intercept and slope, with note rate as a predictor. The third model allowed correlation between the random intercepts and slopes.

We used a model selection approach to compare model fit between two models fit to the same data using leave-one-out cross-validation (LOO) (Vehtari et al., 2017) implemented in the “brms” package. The “loo_compare” function returns

TABLE 1 Sample size along with mean, standard deviation and range of features included in the present analysis.

Features	Female	Male
N individuals	14	14
N duets	50	49
Note rate mean ± SD	0.24 ± 0.12	0.37 ± 0.11
Note rate range	0.11–0.67	0.11–0.78
Bandwidth (Hz) mean ± SD	6096.11 ± 2573.75	7341.59 ± 1035.45
Bandwidth (Hz) range	360.11–11052.7	3502.06–10569.19
Low frequency (Hz) mean ± SD	5843.54 ± 1019.13	5052.23 ± 555.65
Low frequency (Hz) range	2825.5–11412.81	3478.26–9894.74
High frequency (Hz) mean ± SD	11076.25 ± 2314.99	12341.48 ± 1000.78
High frequency (Hz) range	5480.26–15961.45	5996.05–14596.49
Note duration (s) mean ± SD	0.36 ± 0.17	0.21 ± 0.06
Note duration (s) range	0.04–2.88	0.04–0.56

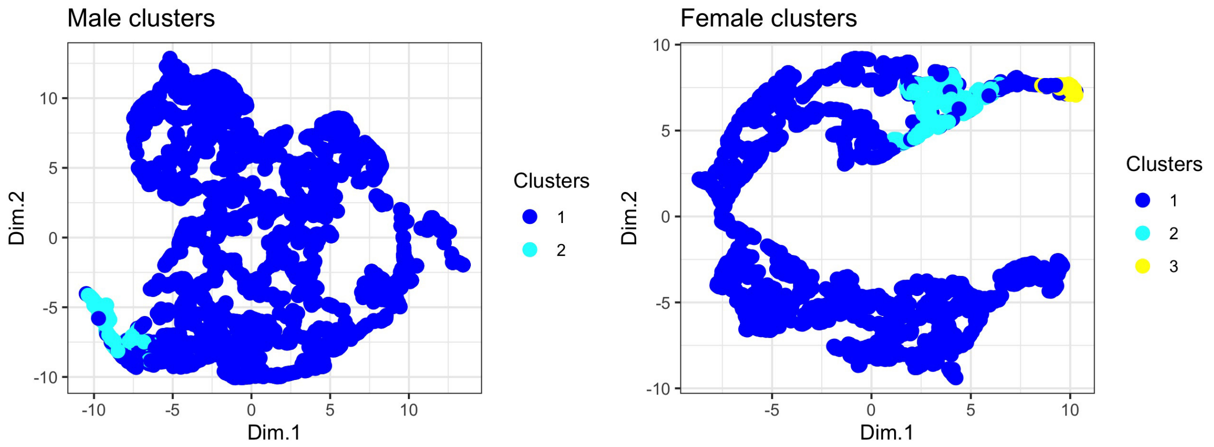


FIGURE 5
UMAP projections for male and female tarsier duet contributions. For each note in the tarsier duet, we input a feature vector consisting of the four features estimated into the UMAP algorithm, which returned two-dimensional embeddings (Dim. 1 and Dim. 2) used to visualize clustering of note types. Each point represents a single note from the tarsier duet. The color indicates cluster assignment by the random forest algorithm (see text for details).

the difference between the expected log-predictive density (ELPD) of all models relative to the model with the highest ELPD (elpd_diff). The function also returns an estimate of the uncertainty (se_diff), which can be used to assess if the differences among models are reliable (Bürkner, 2017a). We simulated a total of 8,000 samples for inference from four chains, with each chain utilizing 2,000 samples for warmup. We specified weakly informative normal priors for the slope and intercept parameters, and weakly informative half-t priors for the variance components. To further assess fit of the top models we used the posterior predictive check function in "bayesplot" that simulates data from the posterior predictive distribution; if the model is a good fit, then data simulated from the posterior predicted distribution should be similar to the observed data (Gabry et al., 2019). To ensure proper mixing and convergence we inspected trace plots.

Results

Descriptive analysis

We report the results of 6,681 notes from 28 tarsier individuals (14 males and 14 females). We initially analyzed 50 duets, but we omitted one highly irregular male contribution from our final analyses, bringing our total to 50 female duet contributions and 49 male duet contributions. On average, female notes had lower maximum and higher minimum frequencies covering a narrower bandwidth than male notes. Female notes also had a longer average duration than male notes. Duets ranged in duration from 12.88 to 203.96 s, and the median duration of the duets was 64.80 s ± 44.29 standard deviation (**Figure 4**). Males initiated 25 of the duets and females initiated 25 of the duets. See **Table 1** for a summary of sample size along with mean, standard deviation and range of features included in the present analysis.

Unsupervised analysis of note types

Using the unsupervised random forest analysis, we found evidence for two clusters in male note types and three clusters in female note types. Visual inspection of the UMAP biplots does not show a strong tendency to cluster in the dataset, although female notes do show stronger clustering than males (**Figure 5**). The lack of many discrete clusters in both the unsupervised analysis and UMAP biplots is indicative of strong gradation in male and female tarsier note types.

Acoustic tradeoffs statistical analysis

For both females and males, we found that the model with note rate as a predictor was ranked higher than the null model, providing evidence for acoustic tradeoffs in these two variables (**Figure 6**). The estimate for the influence of note rate on note bandwidth for the female model was substantially lower (estimate = −13312.17, 95% CI = −15451.09 to −11323.88]) than the estimate for the male note rate (estimate = −679.78, 95% CI = [−1836.83 to 508.25]; **Figure 7** and **Table 2**). Although the male estimates were negative the 95% confidence interval did overlap zero. The top model for females included a random intercept, and slope for pair and performed substantially better than the null model (elp_diff = −163.4; se_diff = 19.4). The top model for males also performed substantially better than the null model (elp_diff = −36.5; se_diff = 11.7) and included a random intercept and slope.

Discussion

Summary of results

Our results show that both males and females were equally likely to begin a duet bout. Both male and female note repertoires show highly graded notes rather than discrete note categories, with male notes having a higher degree of gradation. Additionally, our results provide evidence for an acoustic tradeoff between the rate of note repetition and the frequency bandwidth of those notes for both male and female tarsier duet contributions, with a much stronger effect for female contributions.

The degree of note gradation can indicate different functions for various notes. For instance, discrete vocalizations are associated with predator notification and identification in meerkats (*Suricata suricatta*), Japanese great tits (*Parus major minor*), and vervet monkeys (*Cercopithecus aethiops*) (Seyfarth et al., 1980; Suzuki, 2014; Rauber and Manser, 2017), while in black-capped chickadees (*Parus atricapillus*), discrete notes were associated with courtship behaviors and graded note types were associated with escape and conflict behaviors (Ficken et al., 1978). Given that both tarsier sexes had highly graded note types and our acoustic data is not accompanied by contextual data, we are unable to draw any definite conclusions about the various purposes of each note type; however, future studies that consider the complete note repertoire of the species along with contextual observations and/or playback studies may be able to discern functionality differences in notes based on degree of gradation.

Given the limitations of the data used for our analyses and considering we did not have information regarding animal age, pair-length or other demographic parameters, we can only speculate as to why we observed a stronger pattern of this acoustic tradeoff in female duet contributions compared to male duet contributions. It is possible that this difference is due to the inherent differences in the sex-specific contributions of the tarsier duets, as there is greater variability in female duet contribution note bandwidth. Our results also lead to bigger questions about the function of the male and female contributions and why the female contribution is more complex than the male. In duetting birds, it has been proposed that sex-specificity in duets is due to an extra-pair communication function (Dahlin and Benedict, 2014), and it seems likely that this is also the case with tarsier duets. In addition, males also had a larger range of note rates and in some cases individual males showed patterns opposite that predicted by the acoustic tradeoff we examined. Therefore, it appears that other pressures apart from those consistent with the acoustic tradeoff shaped male tarsier duet contributions.

Previous research on the acoustic tradeoff between rate of note repetition and note bandwidth showed the existence of this tradeoff in a multitude of taxa, including birds, mice, bats, and non-human primates (Podos, 1996; Ballentine et al., 2004; Pasch et al., 2011; Derryberry et al., 2012; Wilson et al., 2014; Clink et al., 2018; Sun et al., 2021). Our findings add another species to this growing list, contributing to the literature that suggests that the existence of this acoustic tradeoff may be effectively universal. This is significant, as universals are relatively rare in animal behavior research (Ferrer-i-Cancho et al., 2013), and can serve to guide our understanding of how certain traits and behaviors evolved over time. However, the evolutionary causes of this acoustic tradeoff are not yet fully understood and may vary between species. Our study does not definitively rule out either morphological or neurological causes of this acoustic performance constraint but adds more literature to the discourse on the ubiquity of this acoustic tradeoff.

Potential limitations

Our study had a few limitations which warrant discussion. First, we examined only notes included in the duets of this tarsier species. This excludes all other vocalizations, including those emitted during hunting, mating, aggressive, parent-offspring, and feeding contexts. Inclusion of these vocalizations may result in different outcomes relating to repertoire size. In addition, different call types may have different constraints, so it is unclear if the acoustic tradeoff between the rate of note repetition and the bandwidth of those notes is prevalent across call types. Additionally, we did not collect nor present data on the non-acoustic behaviors of the individuals at the time of recording (i.e., height in a tree, proximity to conspecifics, maturity, age, reproductive status, presence of predators). It is possible that, like gibbons, tarsier duets vary across different contexts (Clarke et al., 2006; Andrieu et al., 2020). Future studies that compare duets emitted under different contexts (e.g., territorial encounters vs. reuniting at the sleep tree) will be informative. Due to significant temporal and spectral overlap between male and female duet contributions, we were unable to use the robust features in Raven for our unsupervised analysis. These features are calculated based on the energy in the selection and given the substantial overlap in male and female notes the values would have been skewed. This means we were restricted to the four aforementioned features of note duration, minimum frequency, maximum frequency, and bandwidth. If we were able to use the robust features and include a larger number of features that

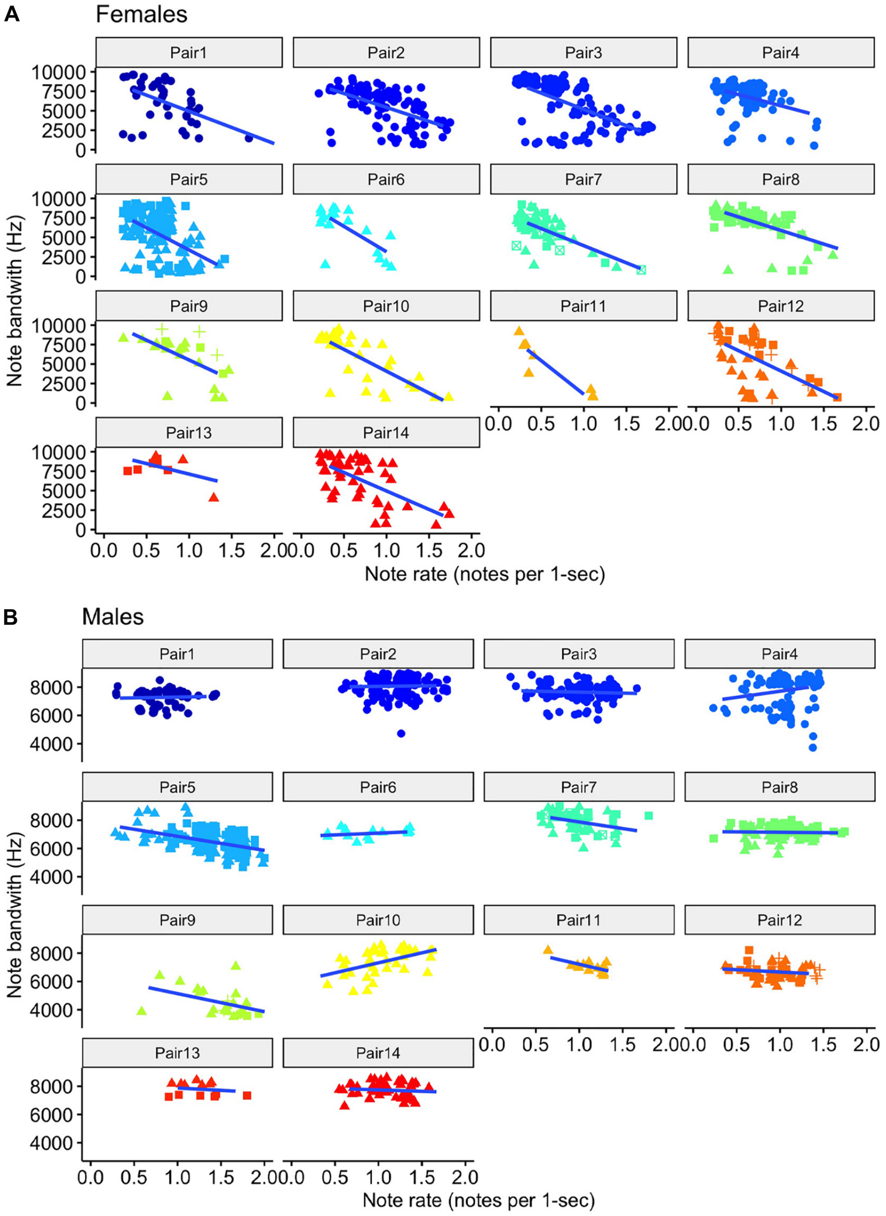


FIGURE 6
Scatter plots of note bandwidth as a function of the rate of note output (number of notes per 1-s) for female **(A)** and male **(B)** tarsier duets. Note that the female **(A)** scatterplots use a broader bandwidth scale than the male **(B)** scatterplots, reflecting the generally wider note bandwidth exhibited by the females. The female scatterplots also show much stronger negative slopes than the male scatterplots. The shape of the points indicates which duet the notes came from. Trend lines were added using the "geom_smooth" function in "ggplot2" to visualize differences across pairs (Wickham, 2016).

described the notes, then our unsupervised clustering results may have been different.

Future directions

The evidence for the acoustic tradeoff between note rate and bandwidth may be due to either morphological or neurological constraints, or a combination of the two. More research is needed to examine the extent of the existing morphological limitations on the vocal production system, as well as the existence of neurological constraints. Future studies compiling a more exhaustive vocal repertoire of the tarsier, as well as the contexts in which each call type is produced, would be extremely valuable and inform many subsequent studies in tarsier

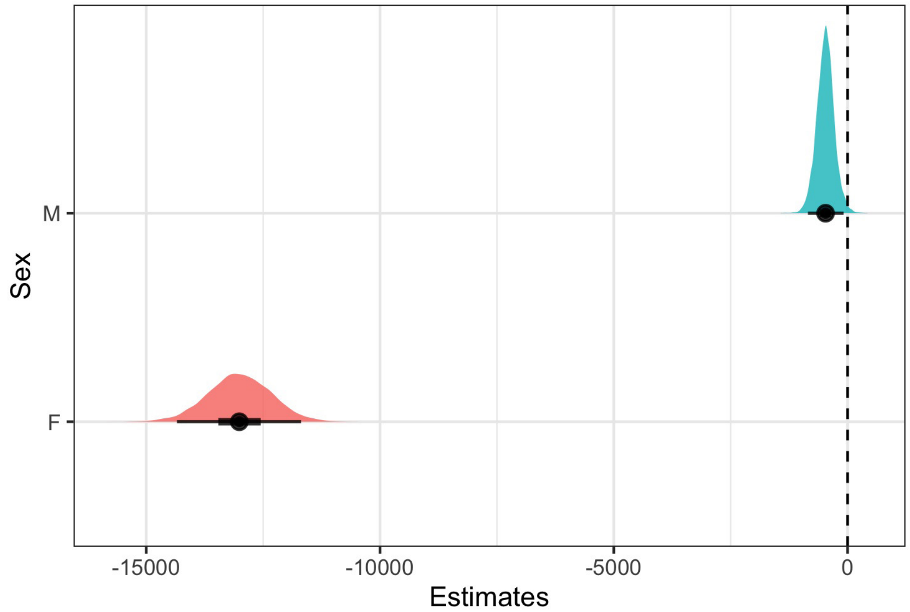


FIGURE 7
Coefficient estimates ± 95% credible intervals for male and female models with bandwidth as an outcome and note rate as a predictor. We considered the predictors reliable if the confidence intervals (indicated in black) did not overlap zero. Each point indicates the median posterior density credible interval value, the inner black bars represent the 50% credible interval, and the outer black bars represent the 95% credible intervals. The colored distribution plots indicate the associated uncertainty in the point estimates.

TABLE 2 Model summary of the top models for male and female note rate along with the null models.

	Female top model		**Female null model**		**Male top model**		**Male null model**	
Predictors	***Estimates***	***CI (95%)***	***Estimates***	***CI (95%)***	***Estimates***	***CI (95%)***	***Estimates***	***CI (95%)***
Intercept	9192.73	8800.32 to 9604.21	6141.29	5795.31 to 6506.23	7550.73	7133.45 to 7946.99	7259.46	6732.94 to 7764.83
Note rate	−13312.17	−15451.09 to −11323.88			−679.78	−1836.83 to 508.25		
Random Effects								
σ^2	4365651.87		6426636.93		495856.24		535599.34	
τ_{00}	155685.42 $_{\text{pair}}$		266227.99 $_{\text{pair}}$		407679.79 $_{\text{pair}}$		990274.99 $_{\text{pair}}$	
τ_{11}	7380961.94 $_{\text{pair.noterate}}$				3976751.18 $_{\text{pair.noterate}}$			
ICC	0.14		0.04		0.63		0.65	
N	14 $_{\text{pair}}$		14 $_{\text{pair}}$		14 $_{\text{pair}}$		14 $_{\text{pair}}$	
Observations	856		856		1079		1079	
Marginal R^2/Conditional R^2	0.331/0.343		0.000/0.031		0.005/0.539		0.000/0.501	

For each model we included note bandwidth as the outcome. The top models for both males and females included note rate (notes per 1-s) as a predictor, and the 95% confidence intervals did not overlap zero for females. In the table above the number of observations refers to the number of time bins for all males or females wherein we measured note rate.

acoustics. Additionally, future studies that take into account variables such as age, weight, time since pairing, and number of offspring will likely make significant contributions toward determining if these tradeoffs are honest reflections of caller quality.

Data availability statement

All R code and data needed to recreate the analyses are available on GitHub: https://github.com/DenaJGibbon/Tarsier-acoustic-tradeoffs.

Ethics statement

The animal study was reviewed and approved by the Institutional approval was provided by the Cornell University (IACUC 2017-0098).

Author contributions

DC and JT performed field work and recorded acoustic data. IC annotated all spectrograms. DC ran analyses in the R programming environment. IC and DC collaborated to develop, revise, and prepare the manuscript for publication. LJ provided input on statistical methods. All authors reviewed and approved of the final version.

Funding

Funding for the field work portion of this project was provided to DC by the Fulbright ASEAN U.S. Scholar Grant (no award number given).

Acknowledgments

We are highly appreciative of early and informative conversations with Russ Charif regarding acoustic tradeoffs. We thank the Fulbright US Scholar Program for providing funding for the data collection for this project. We are incredibly grateful to Vandem Tundu for his assistance with data collection. We also acknowledge the staff at the American Indonesian Exchange Foundation for their help obtaining the necessary permits. We also thank Rob Raguso for his indispensable mentorship, advice, and encouragement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aide, T. M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., and Vega, G. (2013). Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1:e103.

Andrieu, J., Penny, S. G., Bouchet, H., Malaivijitnond, S., Reichard, U. H., and Zuberbühler, K. (2020). White-handed gibbons discriminate context-specific song compositions. *PeerJ* 8:e9477. doi: 10.7717/peerj.9477

Ballentine, B., Hyman, J., and Nowicki, S. (2004). Vocal performance influences female response to male bird song: an experimental test. *Behav. Ecol.* 15, 163–168. doi: 10.1093/beheco/arg090

Balsby, T. J. S., and Hansen, P. (2010). Element repertoire: change and development with age in whitethroat sylvia communis song. *J. Ornithol.* 151, 469–476. doi: 10.1007/s10336-009-0481-4

Blue, S. K. (2020). "The vocal repertoire of tibetan macaques (*Macaca thibetana*) and congeneric comparisons," in *The Behavioral Ecology of the Tibetan Macaque*, ed J. H. Li, L. Sun, and P. M. Kappeler (Cham: Springer), 119–139. doi: 10.1007/978-3-030-27920-2_7

Blute, M. (2019). A new, new definition of evolution by natural selection. *Biol. Theory* 14, 280–281. doi: 10.1007/s13752-019-00328-4

Brainard, M. S., and Doupe, A. J. (2002). What songbirds teach us about learning. *Nature* 417, 351–358. doi: 10.1038/417351a

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bürkner, P.-C. (2017a). Advanced Bayesian multilevel modeling with the R package brms. *arXiv* [Preprint]. 1705.11123.,

Bürkner, P.-C. (2017b). Brms: an R package for Bayesian multilevel models using Stan. *J. Statist. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Statist. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01

Charif, R. A., Waack, A. M., and Strickman, L. M. (2010). *Raven Pro 1.4 User's Manual.* Ithaca, NY: Cornell Lab of Ornithology, 25506974.

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One* 1:e73. doi: 10.1371/journal.pone.0000073

Clink, D. J., and Klinck, H. (2021). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods Ecol. Evol.* 12, 328–341. doi: 10.1111/2041-210X.13520

Clink, D. J., Bernard, H., Crofoot, M. C., and Marshall, A. J. (2017). Investigating individual vocal signatures and small-scale patterns of geographic variation in female Bornean gibbon (*Hylobates muelleri*) great calls. *Int. J. Primatol.* 38, 656–671. doi: 10.1007/s10764-017-9972-y

Clink, D. J., Charif, R. A., Crofoot, M. C., and Marshall, A. J. (2018). Evidence for vocal performance constraints in a female nonhuman primate. *Anim. Behav.* 141, 85–94. doi: 10.1016/j.anbehav.2018.05.002

Clink, D. J., Lau, A. R., Kanthaswamy, S., Johnson, L. M., and Bales, K. L. (2022). Moderate evidence for heritability in the duet contributions of a South American primate. *J. Evol. Biol.* 35, 51–63. doi: 10.1111/jeb.13962

Clink, D. J., Tasirin, J. S., and Klinck, H. (2019). Vocal individuality and rhythm in male and female duet contributions of Gursky's spectral tarsier. *Curr. Zool.* 66, 173–186. doi: 10.1093/cz/zoz035

Clink, D. J., Tasirin, J. S., and Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Curr. Zool.* 66, 173–186.

Clutton-Brock, T. H., and Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour* 69, 145–170. doi: 10.1163/156853979X00449

Dahlin, C. R., and Benedict, L. (2014). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 120, 1–10. doi: 10.1111/eth.12182

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219 . doi: 10.1080/03949370.2021.2015451

Derryberry, E. P., Seddon, N., Claramunt, S., Tobias, J. A., Baker, A., Aleixo, A., et al. (2012). Correlated evolution of beak morphology and song in the neotropical woodcreeper radiation. *Evolution* 66, 2784–2797. doi: 10.1111/j.1558-5646.2012.01642.x

DeVoogd, T. J. (2004). Neural constraints on the complexity of avian song. *Brain Behav. Evol.* 63, 221–232. doi: 10.1159/000076783

Fedurek, P., Zuberbühler, K., and Semple, S. (2017). Trade-offs in the production of animal vocal sequences: insights from the structure of wild chimpanzee pant hoots. *Front. Zool.* 14, 1–10. doi: 10.1186/s12983-017-0235-8

Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., and Semple, S. (2013). Compression as a universal principle of animal behavior. *Cogn. Sci.* 37, 1565–1578. doi: 10.1111/cogs.12061

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (1978). Vocal repertoire of the black-capped chickadee. *Auk* 95, 34–48. doi: 10.2307/4085493

Fitch, W. T., De Boer, B., Mathur, N., and Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Sci. Adv.* 2:e1600723. doi: 10.1126/sciadv.1600723

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in Bayesian workflow. *J. R. Statist. Soc.* 182, 389–402. doi: 10.1111/rssa.12378

Geissmann, T. (1999). Duet songs of the siamang, Hylobates syndactylus: II. Testing the pair-bonding hypothesis during a partner exchange. *Behaviour* 136, 1005–1039. doi: 10.1163/156853999501694

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/S1464793101005826

Geissmann, T., and Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Anim. Behav.* 60, 805–809. doi: 10.1006/anbe.2000.1540

Giacoma, C., Sorrentino, V., Rabarivola, C., and Gamba, M. (2010). Sex differences in the song of *Indri indri*. *Int. J. Primatol.* 31, 539–551. doi: 10.1007/s10764-010-9412-8

Gros-Louis, J. J., Perry, S. E., Fichtel, C., Wikberg, E., Gilkenson, H., Wofsy, S., et al. (2008). Vocal repertoire of *Cebus capucinus*: acoustic structure, context, and usage. *Int. J. Primatol.* 29, 641–670. doi: 10.1007/s10764-008-9263-8

Gursky, S. (2000). Sociality in the spectral tarsier, *Tarsius spectrum*. *Am. J. Primatol.* 51, 89–101. doi: 10.1002/(SICI)1098-2345(200005)51:1<89::AID-AJP7>3.0.CO;2-7

Gursky, S. (2002). Determinants of gregariousness in the spectral tarsier (Prosimian: *Tarsius spectrum*). *J. Zool.* 256, 401–410. doi: 10.1017/S0952836902000444

Gursky, S. (2015). Ultrasonic vocalizations by the spectral tarsier, *Tarsius spectrum*. *Folia Primatol.* 86, 153–163. doi: 10.1159/000371885

Haimoff, E. H. (1986). Convergence in the duetting of monogamous Old World primates. *J. Hum. Evol.* 15, 51–59. doi: 10.1016/S0047-2484(86)80065-3

Jones, G. (1997). Acoustic signals and speciation: the roles of natural and sexual selection in the evolution of cryptic species. *Adv. Study Behav.* 26, 317–354. doi: 10.1016/S0065-3454(08)60383-6

Kaufman, L., and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons.

Koch, R., Raymond, M., Wrege, P., and Klinck, H. (2016). SWIFT: a small, low-cost acoustic recorder for terrestrial wildlife monitoring applications. in *North American Ornithological Conference*. Washington, DC.

Konopka, T. (2020). *UMAP: Uniform Manifold Approximation and* Projection. R Package Version 0.2. 5.0.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

Lieberman, P. H., Klatt, D. H., and Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164, 1185–1187. doi: 10.1126/science.164.3884.1185

MacKinnon, J., and MacKinnon, K. (1980). The behavior of wild spectral tarsiers. *Int. J. Primatol.* 1, 361–379. doi: 10.1007/BF02692280

Madhulatha, T. S. (2011). "Comparison between k-means and k-medoids clustering algorithms," in *International Conference on Advances in Computing and Information Technology*, eds D. C. Wyld, M. Wozniak, N. Chaki, N. Meghanathan, and D. Nagamalai (Berlin: Springer), 472–481. doi: 10.1007/978-3-642-22555-0_48

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2012). *Cluster: Cluster Analysis Basics and Extensions.* R Package Version 1.56.

Marler, P. (1970). A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* 71:1. doi: 10.1037/h0029144

McInnes, L., Healy, J., and Melville, J. (2018). Umap: uniform manifold approximation and projection for dimension reduction. *arXiv* [Preprint]. 1802.03426 doi: 10.21105/joss.00861

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting-A mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi)*? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Nousek, A. E., Slater, P. J. B., Wang, C., and Miller, P. J. O. (2006). The influence of social affiliation on individual vocal signatures of northern resident killer whales (*Orcinus orca*). *Biol. Lett.* 2, 481–484. doi: 10.1098/rsbl.2006.0517

Parra-Hernández, R. M., Posada-Quintero, J. I., Acevedo-Charry, O., and Posada-Quintero, H. F. (2020). Uniform manifold approximation and projection for clustering taxa through vocalizations in a Neotropical passerine (rough-legged tyrannulet. *Phyllomyias burmeisteri*). *Animals* 10:1406. doi: 10.3390/ani10081406

Pasch, B., George, A. S., Campbell, P., and Phelps, S. M. (2011). Androgen-dependent male vocal performance influences female preference in Neotropical singing mice. *Anim. Behav.* 82, 177–183. doi: 10.1016/j.anbehav.2011.04.018

Podos, J. (1996). Motor constraints on vocal development in a songbird. *Anim. Behav.* 51, 1061–1070. doi: 10.1006/anbe.1996.0107

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (*Passeriformes: Emberizidae*). *Evolution* 51, 537–551. doi: 10.1111/j.1558-5646.1997.tb02441.x

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., and Fischer, J. (2015). Vervets revisited: a quantitative analysis of alarm call structure and context specificity. *Sci. Rep.* 5:13220. doi: 10.1038/srep13220

Rauber, R., and Manser, M. B. (2017). Discrete call types referring to predation risk enhance the efficiency of the meerkat sentinel system. *Sci. Rep.* 7:44436. doi: 10.1038/srep44436

Reby, D., and McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Anim. Behav.* 65, 519–530. doi: 10.1006/anbe.2003.2078

Römer, H. (1993). Environmental and biological constraints for the evolution of long-range signalling and hearing in acoustic insects. *Philos. Trans. R. Soc. Lond. Biol. Sci.* 340, 179–185. doi: 10.1098/rstb.1993.0056

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi: 10.1016/0377-0427(87)90125-7

Sainburg, T., Thielk, M., and Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* 16:e1008228. doi: 10.1371/journal.pcbi.1008228

Segbroeck, M. V., Knoll, A. T., Levitt, P., and Narayanan, S. (2017). MUPET-mouse ultrasonic profile extraction: a signal processing tool for rapid and unsupervised analysis of ultrasonic vocalizations. *Neuron* 94, 465.e5–485.e5. doi: 10.1016/j.neuron.2017.04.005

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., et al. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proc. Natl. Acad. Sci. U.S.A.* 117, 17049–17055. doi: 10.1073/pnas.2004702117

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094. doi: 10.1016/S0003-3472(80)80097-2

Shuker, D. M., and Kvarnemo, C. (2021). The definition of sexual selection. *Behav. Ecol.* 32, 781–794. doi: 10.1093/beheco/arab055

Smith, W. J. (1994). Animal duets: forcing a mate to be attentive. *J. Theor. Biol.* 166, 221–223. doi: 10.1006/jtbi.1994.1019

Sun, C., Zhang, C., Lucas, J. R., Gu, H., Feng, J., and Jiang, T. (2021). Vocal performance reflects individual quality in male Great Himalayan leaf-nosed bats (*Hipposideros armiger). Integr. Zool.* 1–10. doi: 10.1111/1749-4877.12545

Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim. Behav.* 87, 59–65. doi: 10.1016/j.anbehav.2013.10.009

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2016). Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *BMC Evol. Biol.* 16:4. doi: 10.1186/s12862-015-0578-8

Thorpe, W. H. (1961). *Bird-Song: The Biology of Vocal Communication and Expression in Birds*. New York, NY: Cambridge University Press.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statist. Comput.* 27, 1413–1432. doi: 10.1007/s11222-016-9696-4

Voigt, C., Leitner, S., and Gahr, M. (2006). Repertoire and structure of duet and solo songs in cooperatively breeding white-browed sparrow weavers. *Behaviour* 159–182. doi: 10.1163/156853906775900739

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., and Fischer, J. (2015). Characterizing vocal repertoires-Hard vs. Soft classification approaches. *PLoS One* 10:e0125785. doi: 10.1371/journal.pone.0125785

Wickham, H. (2016). *Ggplot2: Elegant Graphics For Data Analysis*. Berlin: Springer. doi: 10.1007/978-3-319-24277-4

Wilkins, M. R., Seddon, N., and Safran, R. J. (2013). Evolutionary divergence in acoustic signals: causes and consequences. *Trends Ecol. Evol.* 28, 156–166. doi: 10.1016/j.tree.2012.10.002

Wilson, D. R., Bitton, P.-P., Podos, J., and Mennill, D. J. (2014). Uneven sampling and the analysis of vocal performance constraints. *Am. Natural.* 183, 214–228. doi: 10.1086/674379

Winter, P., Ploog, D., and Latta, J. (1966). Vocal repertoire of the squirrel monkey *(Saimiri sciureus)*, its analysis and significance. *Exp. Brain Res.* 1, 359–384. doi: 10.1007/BF00237707

Zollinger, S. A., Podos, J., Nemeth, E., Goller, F., and Brumm, H. (2012). On the relationship between, and measurement of, amplitude and frequency in birdsong. *Anim. Behav.* 84, e1–e9. doi: 10.1016/j.anbehav.2012.04.026

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fevo.2022.910260

Transient co-singing of offspring and mothers in non-duetting Javan gibbons (*Hylobates moloch*)

Yoonjung Yi[1,2,3*†‡], Ahyun Choi[3,4†], Saein Lee[3,4], Soojung Ham[2], Haneul Jang[5], Rahayu Oktaviani[3,6], Ani Mardiastuti[7] and Jae C. Choe[2*‡]

[1]Laboratory of Animal Behavior and Conservation, College of Biology and the Environment, Nanjing Forestry University, Nanjing, China, [2]Division of Ecoscience, Ewha Womans University, Seoul, South Korea, [3]Javan Gibbon Research and Conservation Project, Bogor, Indonesia, [4]Interdisciplinary Program of Ecocreative, Ewha Womans University, Seoul, South Korea, [5]Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [6]Yayasan Konservasi Ekosistem Alam Nusantara (KIARA), Bogor, Indonesia, [7]Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment, IPB University, Bogor, Indonesia

Check for updates

OPEN ACCESS

EDITED BY
Dena Jane Clink,
Cornell University, United States

REVIEWED BY
Pengfei Fan,
Sun Yat-sen University, China
Michal Hradec,
Czech University of Life Sciences Prague, Czechia
Valeria Torti,
University of Turin, Italy

*CORRESPONDENCE
Yoonjung Yi
yi.yoonjung@gmail.com
Jae C. Choe
jaechoe9@gmail.com

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Behavioral and Evolutionary Ecology, a section of the journal Frontiers in Ecology and Evolution

RECEIVED 01 April 2022
ACCEPTED 18 July 2022
PUBLISHED 17 August 2022

CITATION
Yi Y, Choi A, Lee S, Ham S, Jang H, Oktaviani R, Mardiastuti A and Choe JC (2022) Transient co-singing of offspring and mothers in non-duetting Javan gibbons (*Hylobates moloch*). *Front. Ecol. Evol.* 10:910260. doi: 10.3389/fevo.2022.910260

While the vocalizations of non-human primates were thought to be innate, recent studies have revealed highly flexible vocalizations in immatures. This behavior suggests that social influences have an important role in developing vocalizations. Yet not much is known about how non-human primate vocalization develop and how the vocalizations of immature animals differ between sexes. Here, we analyzed 95 cases of co-singing between mothers and offspring out of 240 female songs from three groups of wild Javan gibbon (*Hylobates moloch*) in Gunung Halimun-Salak National Park, Indonesia, between 2009 and 2021. *Hylobates moloch* is one of only two gibbon species with pairs that do not duet. Instead, they produce sex-specific solo songs. We found that both offspring female and male *H. moloch* follow their mothers' female-specific songs, similar to other duetting gibbon species. Immatures started co-singing with their mothers from 7 months old, but with an average starting age of about 24 months. As female offspring grew older, they co-sung with mothers more often while male offspring did not. After 7 years of age, both sexes stopped co-singing with their mothers and started singing alone, following their own sex-specific vocalizations. We did not find any relation between male offspring co-singing and territorial functions (e.g., co-singing more during intergroup encounters or closer to home range borders). Our results suggest that mothers' songs may trigger male offspring and females to practice singing, but not specifically for males to defend territories. We highlight that despite the absence of duets, *H. moloch* develop their vocalizations from early infancy and throughout their maturation while co-singing with mothers. However, the level of co-singing varies depending on the sexes. Our study is the first to elucidate the sex-specific trajectories of vocal development in *H. moloch* across years, indicating that offspring in non-duetting gibbons co-sing with mothers like in duetting species.

KEYWORDS

co-singing, immature, duet acoustics, Javan gibbon, vocal development

Introduction

Humans socially learn speech coordination, vocalization usage, and comprehension early in their development (Bruner, 1975). In contrast, the vocalizations of non-human primates were thought to be primarily innate (Winter et al., 1973; Pistorio et al., 2006; Tomasello, 2010). However, recent studies have revealed highly variable and flexible patterns of immature vocalizations in non-human primates. Those developmental changes in vocalization can be influenced by physical maturation, experience, and social feedback on vocal development (Lemasson et al., 2011; Chow et al., 2015; Gultekin and Hage, 2017; Takahashi et al., 2017; Zhang and Ghazanfar, 2018; Gultekin et al., 2021). Many non-human primates exhibit sex differences in their vocalizing behavior. Those sex differences are call types (gibbons: Geissmann, 2002; indris: Giacoma et al., 2010), acoustic features (baboons: Ey et al., 2007; owl monkeys: Garcia de la Chica et al., 2020; titi monkeys: Clink et al., 2019; tamarins: Masataka, 1987), and vocalization usage such as occurrence rates and contexts (colobus monkeys: Bene and Zuberbueler, 2009; macaques: Bernstein and Ehardt, 1985; spider monkeys: Dubreuil et al., 2015). Moreover, many primates exhibit sex differences in vocal ontogeny of immatures (Green, 1981; Tomaszycki et al., 2001; Pistorio et al., 2006; Ey et al., 2007), which might reflect sex differences in adult vocalizations. For instance, adult indris (*Indri indri*) have a sex-specific song repertoire, and the female and male juveniles differ in their temporal song parameters (De Gregorio et al., 2022). However, little is known about the sex-specific development of primates' vocalization.

Gibbons (Hylobatidae) are excellent models to study the ontogeny and evolution of vocalizations because of their loud and elaborate songs and the phylogenetic position between great apes and monkeys (Raemaekers et al., 1984). Their songs are innate and species- and sex-specific, and these characteristics are inherited (Brockelman and Schilling, 1984). The most remarkable trait of Hylobatidae is duetting in pairs. These duets mainly consist of antiphonal or simultaneous emissions of stereotyped female great calls consisting of a series of notes with increasing tempo and pitch (Marshall and Marshall, 1976), and variable short-notes of males (Geissmann, 2002). However, *Hylobates moloch*, our study subject, is one of the only two non-duetting gibbon species together with Kloss's gibbons (*Hylobates klossii*), and females and males produce only solo songs (Tenaza, 1976; Kappeler, 1984). Lar gibbons (*Hylobates lar*) males and *H. moloch* females in captivity have been recorded to produce highly coordinated duets, the loss of duet in *H. moloch* is likely derived secondarily from duetting characteristics (Geissmann, 2002). Similarly, a loss of duet might be a derived trait from common ancestors (i.e., synapomorphy) shared by *H. moloch* and *H. klossii* (Chan et al., 2010; Roos, 2016; but see Gani et al., 2021). Moreover, *H. moloch* males rarely sing solos as well compared to females regularly sing solos (Kappeler, 1984; Geissmann and Nijman, 2006). Since the pattern of sexual dimorphism in the adult vocalization is different in duetting and non-duetting species, this may influence immature vocal development in non-duetting species differently from those in duetting species. Yet, little is known about the development of vocalization in gibbons.

Immature female gibbons develop their ability to perform great calls (i.e., vocal control and acoustic structures) by co-singing with their mothers, in several gibbon species (*H. lar*; Reichard, 2003, *H. klossii*; Tilson, 1981; Whitten, 1982, *Hylobates agilis;* Koda et al., 2013, and yellow-cheeked gibbons; *Nomascus gabriellae*; Merker and Cox, 1999). Co-singing can indicate the strength of bonding between mother and daughter (Koda et al., 2013; Koda, 2016). Furthermore, gibbon mothers may trigger and tutor their daughters by adjusting their vocal structures according to their daughters' responses (Koda et al., 2013). Immature male gibbons also sing female-specific parts when co-singing with their mothers. This behavior has been reported in duetting gibbons such as *H. lar* and agile gibbons (*H. agilis*; Koda et al., 2014), and *N. gabriellae,* northern white-cheeked gibbon (*Nomascus leucogenys*), and black crested gibbons (*Nomascus concolor*; Schilling, 1984; Hradec et al., 2016, 2017, 2021). With physical and sexual maturation, young males of *N. gabriellae* transitioned from female-like songs to a mix of both male and female-like song parts, to male calls only (Hradec et al., 2021). Similar to immature female gibbons, immature males may also co-sing with their mothers to practice vocalizing.

Adult male *H. moloch* sing solos extremely rare, for example only one single male song was heard during 130 full day survey in West Java (Kappeler, 1984) and male solos were 8.5% of all songs heard in Central Java (Geissmann and Nijman, 2006). However, we observed immature *H. moloch* males co-singing with their mothers' song. Male immatures' co-singing with mothers in non-duetting *H. moloch* may relate to territorial defense like the adult female songs or duets (Mitani, 1985; Geissmann and Orgeldinger, 2000; Fan et al., 2009; Ham et al., 2017). Primate vocalizations, especially duets, are often used to defend their territory by advertising the territorial border or their physical condition (i.e., fighting ability; Mitani, 1987; Barelli et al., 2013). For example, adult female *H. moloch* produced songs more often in the area of their home range that overlapped with neighboring groups than expected based on time spent in the overlapping area versus interior area (Ham et al., 2017; Yi et al., 2020). This suggested that songs of adult female *H. moloch* function for territorial defense by advertising territorial boundaries (Ham et al., 2017), like other primate species (southern brown howler monkeys: da Cunha and Jalles-Filho, 2007; titi monkeys: Robinson, 1979). Thus, *H. moloch* immature males may be triggered to co-sing with their mothers more frequently in similar contexts.

This study aims to understand the development of sexually dimorphic vocalization in Javan gibbons (*H. moloch*). To understand how the vocal ontogeny of *H. moloch* differs between sexes, we investigated vocal development in *H. moloch* offspring in Gunung Halimun-Salak National Park, Indonesia, between 2009 and 2021. First, we tested for sex differences in vocal development in offspring *H. moloch*. Since *H. moloch* adult females produce solo songs more frequently than adult males while there is no duetting, we hypothesized that the differences between adult female and adult male vocalizations should be reflected in co-singing with adult females during vocal ontogeny of offspring as well. We specifically predicted that (1) female offspring will start co-singing with their mothers at an earlier age than male offspring, (2) females will co-sing with their mothers more often than male offspring, and (3) the frequency of co-singing will increase as female offspring get older, in opposition with the co-singing activity of male offspring. Then, we investigated why male offspring *H. moloch* co-sing with their mothers despite the absence of duets and the rarity of adult male solo songs. We hypothesized that male offspring co-singing in *H. moloch* is used for territorial defense similar to that of mothers. Specifically, we predicted that male offspring would co-sing with their mothers more often (1) during intergroup encounters compared to the non-intergroup encounter context and (2) close to their home range border compared to the center of their home ranges. Finally, we reported opportunistically collected cases of adult male and offspring solos.

Methods and materials

Data collection

We have been habituating and following wild *H. moloch* in Citalahab Forest, Gunung Halimun-Salak National Park, West Java, Indonesia (S 6°44′19′′E 106°31′45′′), as part of a long-term project called the Javan Gibbon Research and Conservation Project (Kim et al., 2012; Ham et al., 2017; Oktaviani et al., 2018; Yi et al., 2020; Jang et al., 2021). We collected data from three habituated gibbon groups and each gibbon group consists of one adult female, one adult male and 2–3 offspring (**Table 1**). During the study period, we observed 13 offspring from the three gibbon groups (male; $N = 9$, female; $N = 4$). They were distributed in four age categories (infant: 0–2 year, juvenile: 2–5 year, adolescent: 5–8 year, subadult: >8 year; definitions following Brockelman et al., 1998). Whenever there were adult female or male vocal events, we recorded the presence of any offspring co-singing and the Global Positioning System (GPS) coordinates. We defined co-singing as two individuals (mother and her immature offspring in our study) vocalizing simultaneously (Koda et al., 2013; Hradec et al., 2021), while duetting is usually defined as coordinated vocalization between a female and male pair in gibbons (Marshall and Marshall, 1976). When an offspring co-sang with an adult, we also recorded the age of the offspring. We calculated the age of offspring from the date of birth. For the individuals without exact birth information ($N_{\text{offspring}} = 5$), we calculated the inter-birth interval of the study population between 2010 and 2021 ($N_{\text{offspring}} = 10$; $N\ \text{within} - \text{group inter} - \text{birth interval} = 7$; mean ± SD = 1,295 ± 242 days) and subtracted that from the date of birth of the next-born individual with exact birth information in the same group. We compared the body size and behaviors (e.g., spatial distance with other group members) of the individuals with putative age with those of the individuals with known exact age. Then we confirmed no obvious mismatch between the putative age and the body size/behaviors. We also recorded whether co-singing occurred during the intergroup encounters. We excluded the data for which the intergroup encounter context could not be determined for the analyses. We collected GPS coordinates of adult gibbons at 15-min intervals from 2014 to 2019 and 10-min from 2020 to 2021. We followed a gibbon group from a sleeping tree to the next between 0600 and 1700 h.

In addition, we opportunistically observed fourteen cases of offspring vocalization without an adult female vocalization prior. We also recorded 51 male solos from both the focal and the neighboring groups during the study period. Among these, we could confirm the identity of the singers for 41 cases (22 cases

TABLE 1 The life history of the study subjects and data collection period for each gibbon group in Gunung Halimun-Salak National Park, Indonesia between 2009 and 2021.

Group	Data collection period	Individual	Sex	Class	Date of birth	Record
A	Jul 2009–Mar 2010, Mar–Nov 2011, Feb–Jul 2012, Feb–Dec 2013, Dec 2014–Jul 2016, Nov 2016–Aug 2017, Dec 2018–Feb 2021	Aris	Male	Adult	Unknown	
		Ayu	Female	Adult	Unknown	
		Asri	Female	Offspring	Nov 2003*	Dispersed in Aug 2011
		Amran	Female	Offspring	Jun 2007*	Disappeared in Aug 2011
		Amore	Male	Offspring	Dec 2010	
		Awan	Male	Offspring	Dec 2013	
		Ajaib	?	Offspring	Oct 2018	
B	Jul 2009–Mar 2010, Mar–Nov 2011, Feb–Jul 2012, Feb–Dec 2013, Dec 2014–Jul 2016, Nov 2016–Aug 2017, Dec 2018–Feb 2021	Bang Kumis	Male	Adult	Unknown	
		Bu Keti	Female	Adult	Unknown	
		Bayi KumKum	Male	Offspring	Jul 2007	Disappeared in Sep 2007
		Bayi Kimkim	Male	Offspring	Apr 2011	Dispersed in Jan 2020
		Bayi Komeng	Male	Offspring	May 2014	
		Bayi Kendeng	Male	Offspring	Nov 2017	
S	Feb–Dec 2013, Dec 2014–Jul 2016, Nov 2016–Aug 2017, Dec 2018–Feb 2021	Sahri	Male	Adult	Unknown	
		Surti	Female	Adult	Unknown	
		Surono	Male	Offspring	Oct 2003*	Dispersed in Feb 2014
		Sendi	Female	Offspring	Apr 2007*	Dispersed in Apr 2016
		Salwa	Male	Offspring	Nov 2010*	
		Sanha	Female	Offspring	May 2014	
		Setia	Male	Offspring	Jan 2018	

*A star next to the date of birth indicates that it has been calculated from the inter-birth interval of the study population.

from the focal groups and 19 cases from non-focal groups with visual contact).

Data analyses

First, we checked the age at which offspring males and females start co-singing with their mothers. To do so, we investigated the first record of co-singing between offspring and adult females. Additionally, we calculated the mean age of the first *five* co-singing events. We chose five events to overcome the potential observation bias of describing only the first event as we may have missed some co-singing events, for instance, while not following the focal group. Following this protocol, we included sufficient data to calculate mean values ($N_{\text{offspring}}$ = 5), while restricting the range of age used to describe the starting age of co-singing. For those with assumed age calculated with inter-birth interval, we did not include them in the calculation in the first or first five co-singing events of co-singing.

To determine the co-singing rates of offspring with mothers, we created a list of female songs ($N_{\text{femalesong}}$ = 240) for each offspring individual per day (N_{total} = 556). When adult females stopped singing more than 5 mins before starting to sing again, we defined it as two different song bouts (hereafter "song" Geissmann, 2002). For example, if an offspring co-sang with at least one female song during the day, we marked the day as an offspring co-singing present. For example, if an adult female sang twice a day, and one of the songs was followed by a juvenile but not by an infant of the focal group, we had a line for each immature for that day, one marked with the juvenile co-singing present and the other marked with the infant co-singing absent. Adult females typically sang one song bouts a day and a maximum of up to three bouts. We analyzed our data on a daily base to compensate for the potential of missing data.

We ran a binomial Generalized Linear Mixed Model (GLMM) with the presence of offspring co-singing with the female songs as the response variable (N_{total} = 532), and the offspring age, sex, and the interaction between the offspring age and sex as test predictors, and interaction between the offspring sex and season (dry, semidry, and wet) to control for a potential seasonal effect. We included the gibbon group ID, offspring ID, and each adult female vocalization event ID as random effects. We added the offspring age and season within gibbon group and offspring ID as random slopes.

We estimated the annual home range of each focal group using 95% kernel density estimates (Worton, 1989) to investigate the effect of the location of female vocalizations and the intergroup encounter on the probability of co-singing in offspring males. We calculated the Euclidean distance between the location of female vocalization and the nearest home range border. If a female sang outside the estimated annual home range, we marked the distance to the closest border with a negative value to differentiate the coordinates inside and outside the home range. The distance was divided by the size of the annual home range (ha) to control the potential effect of the home range size difference between groups. We conducted the spatial analyses using ArcGIS Pro (version 2.8.5; Environmental Systems Research Institute, 2022). Then we ran the second binomial GLMM with the presence of offspring males co-singing to the female songs as the response variable. This has been calculated the same way as co-singing rates of offspring with mothers, we created a list of female songs of which the GPS location was identified ($N_{\text{femalesong}}$ = 110) for each male offspring individual per day (N_{total} = 242). We included the distance from the home range border divided by the home range size and the context (intergroup encounter/non-intergroup encounter) as test predictors. Then, we included the group ID and offspring ID as random effects, and the two predictors (i.e., the distance from home range border divided by the home range size, and the intergroup encounter context) within gibbon group and offspring ID as random slopes.

Before running the two binomial GLMMs, we visually checked the homogeneity of variance and *z*-transformed all covariates in the models. We also checked for multicollinearity using the R package *car* (Fox and Weisberg, 2019) and did not find a problem since the variance inflation factor (VIF) values ranged between 1.68 and 2.79. We presented the results only when the full-null model comparison was statistically significant, while the null model was only with the random effect and slopes included. All data were analyzed using R (version 4.1.0; R Core Team, 2021).

Results

We recorded 240 female songs, and among these, 95 cases of co-singing between offspring and mothers, and 51 male solos over 82 months of surveys between 2009 and 2021 (**Figure 1**). The age of the first recording of offspring co-singing with their mothers was 23.8 ± 18.8 months (mean ± SD; range: 7.0–60.0; $N_{\text{offspring}}$ = 8; **Figure 2**). The age of the first five recordings of offspring co-singing with their mothers was 42.4 ± 16.3 months (mean ± SD; range: 17.2–62.4; $N_{\text{offspring}}$ = 5; **Figure 2**). Since we had first recordings of only one female offspring, we could not compare the age of first recordings between sexes directly. Although not included in the study period, we observed a 2-month-old infant (born in November 2021) of unknown sex co-singing with the mother in group A, one of our focal gibbon groups, in February 2022 (**Supplementary Audio 1**). This is the earliest record of infant gibbon co-singing.

In general, the rate of each offspring co-singing with their mothers' songs (i.e., number of female songs that offspring co-sung with/total number of female songs produced) was 0.159 ± 0.153 (mean ± SD; $N_{\text{offspring}}$ = 13). Probability of co-singing with mothers changed across developmental stages in both female and male offspring (infancy: 0.091 ± 0.101, juvenility: 0.160 ± 0.141, adolescence: 0.234 ± 0.195, subadult period: 0; **Supplementary Figure 1**).

We found that the interaction between offspring age and sex and between offspring sex and season significantly affected the presence of offspring co-singing with their mother songs (full-null model comparison: χ^2 = 6.912, *df* = 2, *p* = 0.032; **Table 2** and **Supplementary Table 1**). The results indicate that as female offspring become older, the probability of female offspring co-singing with their mothers increased while that of male offspring did not (**Figure 3**). The full model testing the effect of distance from the home range border and the intergroup encounter on co-singing between male offspring and mothers did not fit significantly better than the null model (full-null model comparison: χ^2 = 2.059, *df* = 2, *p* = 0.357).

We found that a female offspring and two male offspring produced vocalizations without the presence of adult vocalization ($N_{\text{vocalization}}$ = 12). While female offspring produced vocalizations similar to female great calls even without adult female singing, male offspring produced adult male-like sounds which were not observed when they co-sung with their mothers. The average age of sex-specific vocalization of offspring without adults is 90 months ($N_{\text{offspring}}$ = 1) for a female offspring and 98 ± 2.8 (mean ± SD; $N_{\text{offspring}}$ = 2) months for male offspring. 75% of these vocalizations were observed during intergroup encounters ($N_{\text{vocalization}}$ = 9). All vocalizations of focal subadult males occurred during intergroup encounters ($N_{\text{vocalization}}$ = 7). Subadult males from two encountering groups simultaneously vocalized before the other group members arrived at the encounter location and after the other members of the groups left. Male offspring also vocalized when the adult male of the neighboring group produced vocalizations. However, we have not heard any offspring co-singing with their fathers, yet we are not able to confirm whether there was co-singing before we started daily follows at 0600, given that male gibbons produce solos at dawn before 0600 h (Raemaekers and Raemaekers, 1985; Geissmann and Nijman, 2006).

Discussion

To our knowledge, this study is the first to record the vocal developmental trajectories in female and male gibbons, covering the entire immature and mature period of offspring. We found considerable differences in vocal development between female and male offspring in non-duetting *H. moloch*. Both immature female and male *H. moloch* started co-singing with their mothers during infancy, which began around 7 months, and started co-singing more stably during early juvenility. However, surprisingly, we also observed a 2-month-old infant co-singing with its mother outside of our study period. At this point, we are not able to firmly conclude which sex starts co-singing with their mothers at an earlier age, since the sex ratio of our study subject is highly biased toward males for the first recordings (male; *N* = 7, female; *N* = 1). In another pair-living and duetting primate species, *I. indri*, immature females started participating in choruses earlier than immature males, which might be related to the earlier maturation of females than males in this species (De Gregorio et al., 2022). To clarify the sex difference in the emergence of co-singing in gibbons, an intensive investigation of their early infancy for both sexes would be needed.

As infant gibbons grew older, however, they exhibited a clear sex difference in the rate of co-singing with mothers, supporting our prediction. On average, *H. moloch* female

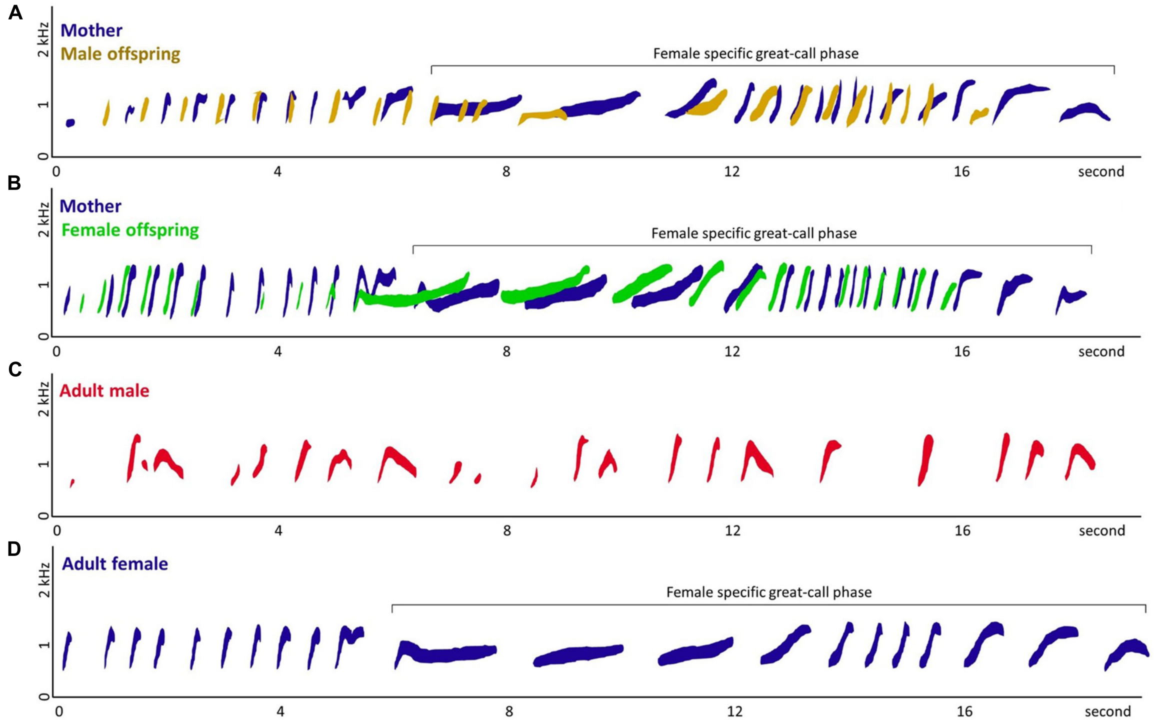


FIGURE 1
Spectrograms of **(A)** co-singing between mother and male offspring, **(B)** co-singing between mother and female offspring, **(C)** adult male solo, and **(D)** adult female solo of *Hylobates moloch*. Panel **(A)** was recorded from gibbon group S, Surti (mother), and Salwa (male offspring; subadult) in September 2019. Panel **(B)** was recorded from gibbon group A, Ayu (mother), and Asri (female offspring; adolescent) in September 2009. Panel **(C)** was recorded from the Javan Gibbon Center, Gunung Gede-Pangrango, Indonesia, in September 2009, due to the difficulty of recording rare adult male solos in the wild. Panel **(D)** was recorded from the gibbon group B, Bu Keti (adult female) in April 2011. Except **(C)** adult male solo, all were recorded from the focal gibbon groups in the study area, Gunung Halimun-Salak National Park, Indonesia. The spectrograms were extracted from the recordings using Raven Lite (Version 2.0.3; Yang, 2022) and we manually traced the spectrograms to specify the callers.

offspring co-sang with their mothers 2.4 times more often than male offspring. Moreover, female offspring co-sang with their mothers more often as they became older. These results support our predictions regarding the difference in vocal development between female and male offspring. This pattern is consistent with *N. gabriellae* females (Merker and Cox, 1999). In contrast, male offspring's engagement in co-singing stayed low throughout development. This sex difference in *H. moloch* seems to be expressed around 2 or 3 years of age, which is still early in their development, given that gibbons wean around 2 years of age (Treesucon, 1984). Co-singing probability with their mothers increased continuously until the female offspring *H. moloch* reached the subadult stage at 8 years of age. From this point on, they were no longer observed co-singing with their mothers, similarly to *H. lar* females (Reichard, 2003). The cessation of co-singing may indicate that female offspring have already acquired adult-level vocalization skills and social independence from mothers (Reichard, 2003; Koda et al., 2013).

In our study, *H. moloch* male offspring also completely stopped co-singing with their mothers at around 7 years of age despite a lower co-singing rate compared to female offspring. In addition, we observed that young males, after 8 years of age, vocalized alone without other group members. Remarkably, while they always produced only female-like great calls when they co-sang with their mothers before 7 years of age, young males produced male-like vocalizations of simple "wa" notes by themselves after 8 years of age. *Nomascus gabriellae* young males also stopped co-singing with their mothers around 7 years old, while gradually switching to adult male songs after producing both female and male type vocalizations between the age of 5–7 years (Hradec et al., 2021). We speculate that these sudden changes in *H. moloch* young male vocalizations might result from hormones related to age and sex, which also may influence the larynx growth and acoustic structures (Newman et al., 2000; Barelli et al., 2013). Further studies should investigate the ontogeny of vocalization with detailed acoustic analyses together with hormonal analyses.

We did not find any relationship between male offspring co-singing with their mothers and territorial defense. Unlike what we predicted, male offspring did not co-sing with their mothers more often during intergroup encounters compared

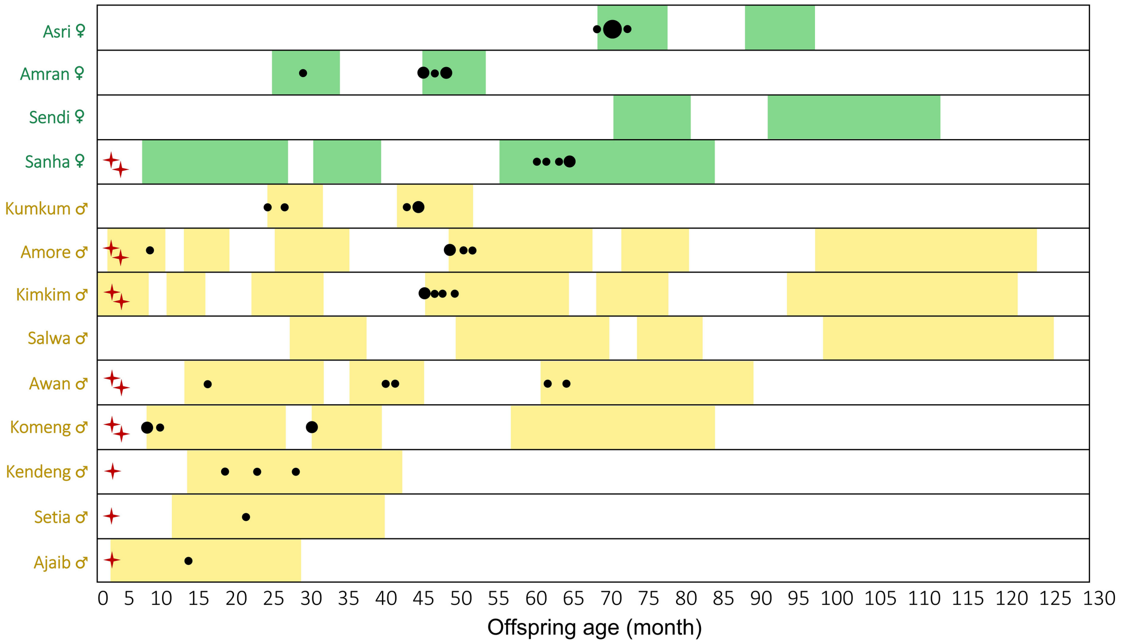


FIGURE 2
The age of the earliest (maximum) five recordings of each offspring *Hylobates moloch* co-singing with mothers in Gunung Halimun-Salak National Park, Indonesia, between 2009 and 2011. The colored bar indicates the data collection period covered in this study. The circle size represents the number of co-singing events (1–3 cases). Red crosses close to the individual name indicate that we included the individuals ($N_{\text{offspring}}$ = 8) in calculating the age of the first co-singing record, and two crosses indicate that we included the individuals ($N_{\text{offspring}}$ = 5) in calculating the age of the first five co-singing records. Individuals without crosses were not included in the first co-singing records because our study period does not cover their early infancy and likely missed the first records.

to the non-intergroup encounter context, or close to their home range border compared to the center. During the study period, we recorded only six cases of vocalizations produced in intergroup encounters from the focal adult males over the 82 months of the study period (e.g., 132 female songs produced in intergroup encounters in the same period). Considering *H. moloch* adult females defend territories by singing solos at the overlapping area with neighbors and advertising the

TABLE 2 The result of the binomial Generalized Linear Mixed Model (GLMM) investigating the effect of offspring age, sex, and season on the offspring *Hylobates moloch* co-singing with their mothers in Gunung Halimun-Salak National Park, Indonesia from 2009 to 2021.

	Estimate	Std. error	*z*-Value	*P*-Value
Offspring age[1]	1.484	0.560	2.648	0.008**
Offspring sex (male)[2]	1.785	1.412	1.265	0.206
Offspring age × sex (male)	−1.642	0.541	−3.035	0.002**
Season (semidry)[2]	4.290	1.560	2.750	0.006**
Season (dry)[2]	3.989	1.548	2.577	0.010**
Offspring sex (male) × Season (semidry)	−3.971	1.690	−2.350	0.018*
Offspring sex (male) × Season (dry)	−4.189	1.697	−2.469	0.014*

[1] *z*-transformed (original mean ± SD = 49.52 ± 31.02).
[2] Reference level for offspring sex: female, season: wet.
The significance levels are * < 0.05 and ** < 0.01.

FIGURE 3
The probability of offspring *Hylobates moloch* co-singing with their mothers throughout the ages in Gunung Halimun-Salak National Park, Indonesia, between 2009 and 2021.

borders (Ham et al., 2017), our finding implies that *H. moloch* adult males rarely produce solos for the same purpose of female's territorial defense. However, since *H. moloch* adult males participate in intergroup encounters mostly by chasing (Yi et al., 2020), we speculate that male offspring also do not relate vocalization or co-singing with adult females for territorial defense.

Then what could be the evolutionary function of co-singing in *H. moloch* male offspring? Male offspring may co-sing with their mothers to avoid competition with fathers by producing female-like great calls, and thus by relaying information about their immature status (Hradec et al., 2021). However, this is unlikely because in both *N. gabriellae* (Hradec et al., 2021) and *H. moloch* (present study), co-singing episodes between young males and mothers were no longer observed after 8 years of age. The competition with fathers should peak around the subadult stage (i.e., >8 year). As Geissmann (2002) suggested, the duetting behavior of *H. moloch* and *H. klossii* was likely lost secondarily. We speculate that in duetting ancestors, *H. moloch* male offspring co-singing with their mothers once had the evolutionarily benefit of facilitating the improvement of their vocalization. However, despite the lower benefits of co-singing compared to duetting ancestors, *H. moloch* male offspring might benefit from practicing to achieve varieties of adult male songs and strengthen their bonds with their mothers through co-singing.

Finally, we found that *H. moloch* males rarely but still considerably produced solo songs after 0630 h. We assume that *H. moloch* males vocalize more often than we report here from our opportunistic data, since male gibbons produce solos at dawn (Raemaekers and Raemaekers, 1985; Geissmann and Nijman, 2006). While previous studies mainly focused on the pre-dawn male chorus, there is no detailed description of post-dawn male songs. Despite their rarity, we found that male vocalizations occurred throughout a day between 0630 and 1700 h, and some male vocalizations lasted more than 40 min. Further studies on adult male vocalization will shed light on the evolution of young male vocalization in *H. moloch*.

Our study has limitations in several aspects. First, because of data collection methods (i.e., following one focal group a day), we could have missed co-singing between adults and offspring and offspring solos, even though we tried to overcome this limitation by counting the first five co-singing records. Also, we lack observations in the early dawn, in which adult males might produce solos more. Second, we do not know the exact ages of some offspring born before our research period. Even though we relied on the individuals with exact birth records in the first co-singing events and considered body size and behaviors as other proxies, there may be a slight difference from their actual age. Moreover, we have the biased sex ratio in the focal groups resulting in a small sample size for female offspring. Lastly, we could not compare vocal structures between female and male offspring to examine the similarity with their mothers since we did not have good quality recordings. Instead, we focus more on their behavioral ecology than detailed acoustic analyses. We recommend that future studies investigate the ontogeny of vocal structures in both female and male offspring.

Vocalizations of non-human primates have received great attention due to their phylogenetic closeness to humans. However, there is a dearth of data on the vocal development of non-human primates, probably because of prolonged immature periods, long lifespan, and different methodologies and parameters for various vocal patterns (Harvey and Clutton-Brock, 1985; van Schaik and Isler, 2012). Despite the limitations of our study, our results highlight that *H. moloch* develop their vocalization from early infancy throughout their development while interacting with their mothers. Given their elaborated song structures, co-singing with their mother would help young female *H. moloch* acquire adult-level communication skills. Even though young males co-sing with their mothers less, still, co-singing can help achieve a variety of adult male songs despite potentially lower benefits than their duetting ancestors. Our findings contribute to understanding the non-duetting gibbon's vocal development and emphasize further interdisciplinary and longitudinal studies encompassing social systems, life history, behaviors, and physiology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The animal study was reviewed and approved by Indonesian Ministry of Research and Technology (RISTEK), the Indonesian Ministry of Forestry's Department for the Protection and Conservation of Nature (PHKA), and the Gunung Halimun-Salak National Park.

Author contributions

YY and AC conceptualized the initial idea and drafted the manuscript and performed and interpreted the statistical analyses. YY, AC, SL, SH, RO, and HJ collected the behavioral data. YY, AC, SL, SH, RO, HJ, AM, and JC revised, contributed to the article, and confirmed the final version of the manuscript.

Funding

This research was partly supported by the National Research Foundation of Korea (NRF) funded by the Ministry of

Education (2021R1A6A3A03039709) to YY and Amore Pacific Academic and Cultural Foundation (AACF). YY was supported by the Foreign Youth Talent Program (QN2021014010L) from the Ministry of Science and Technology and Research Grants (2021K310C) from the Department of Human Resources and Social Security of Jiangsu Province of the People's Republic of China. This work was also supported by National Research Foundation of Korea, Posco International Corporation, and Ewha Womans University. This project was conducted in collaboration with the Department of Forest Resources Conservation and Ecotourism, Faculty of Forestry and Environment at the IPB University. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

We thank the Indonesian Ministry of Research and Technology (RISTEK), the Indonesian Ministry of Forestry and Environment, and the Gunung Halimun-Salak National Park (GHSNP) for the research permissions. We thank Rinekso Soekmadi, Mirza Kusrini, and GHSNP staff for their assistance and cooperation. We also thank our field assistants, Aris, Muhammad Nur, Ri Rudini, Isra Kurnia, Iyan Sopian, Nandar Pratama, Indra Lesmana, Muhammad A. Ajis, and Alan for their hard work in the field. We are grateful to Sangmi Lee for the gibbon illustration.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.910260/full#supplementary-material

References

Barelli, C., Mundry, R., Heistermann, M., and Hammerschmidt, K. (2013). Cues to androgens and quality in male gibbon songs. *PLoS One* 8:e82748. doi: 10.1371/journal.pone.0082748

Bene, J., and Zuberbueler, K. (2009). Sex differences in the use of vocalizations in wild olive *Colobus* monkeys. *Eur. J. Sci. Res.* 25, 266–279.

Bernstein, I. S., and Ehardt, C. L. (1985). Age-sex differences in the expression of agonistic behavior in rhesus monkey (*Macaca mulatta*) groups. *J. Comp. Psychol.* 99:115. doi: 10.1037/0735-7036.99.2.115

Brockelman, W. Y., and Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature* 312, 634–636. doi: 10.1038/312634a0

Brockelman, W. Y., Reichard, U., Treesucon, U., and Raemaekers, J. J. (1998). Dispersal, pair formation and social structure in gibbons (*Hylobates lar*). *Behav. Ecol. Sociobiol.* 42, 329–339. doi: 10.1007/s002650050445

Bruner, J. S. (1975). The ontogenesis of speech acts. *J. Child Lang.* 2, 1–19. doi: 10.1016/j.biosystems.2020.104264

Chan, Y.-C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C.-C., Pei, K. J.-C., et al. (2010). Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates* gibbons. *PLoS One* 5:e14419. doi: 10.1371/journal.pone.0014419

Chow, C. P., Mitchell, J. F., and Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B Biol. Sci.* 282:20150069. doi: 10.1098/rspb.2015.0069

Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour* 156, 1471–1494. doi: 10.1163/1568539X-00003575

da Cunha, R. G. T., and Jalles-Filho, E. (2007). The roaring of southern brown howler monkeys (*Alouatta guariba clamitans*) as a mechanism of active defence of borders. *Folia Primatol.* 78, 259–271. doi: 10.1159/000105545

De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., et al. (2022). Born to sing! Song development in a singing primate. *Curr. Zool.* 67, 585–596. doi: 10.1093/cz/zoab018

Dubreuil, C., Notman, H., and Pavelka, M. S. (2015). Sex differences in the use of whinny vocalizations in spider monkeys (*Ateles geoffroyi*). *Int. J. Primatol.* 36, 412–428. doi: 10.1007/s10764-015-9832-6

Environmental Systems Research Institute (2022). *ArcGIS Pro. (Version 2.8.5)*. Available online at: https://www.esri.com/en-us/home

Ey, E., Hammerschmidt, K., Seyfarth, R. M., and Fischer, J. (2007). Age-and sex-related variations in clear calls of *Papio ursinus*. *Int. J. Primatol.* 28, 947–960. doi: 10.1007/s10764-007-9139-3

Fan, P. F., Xiao, W., Huo, S., and Jiang, X. L. (2009). Singing behavior and singing functions of black-crested gibbons (*Nomascus concolor jingdongensis) at Mt. Wuliang, Central Yunnan, China. Am. J. Primatol.* 71, 539–547.

Fox, J., and Weisberg, S. (2019). *An {R} companion to applied regression*. Thousand Oaks, CA: Sage.

Gani, M., Rovie-Ryan, J. J., Sitam, F. T., Kulaimi, N. A. M., Zheng, C. C., Atiqah, A. N., et al. (2021). Taxonomic and genetic assessment of captive White-Handed Gibbons (*Hylobates lar*) in Peninsular Malaysia with implications towards conservation translocation and reintroduction programmes. *ZooKeys* 1076:25. doi: 10.3897/zookeys.1076.73262

Garcia de la Chica, A., Huck, M., Depeine, C., Rotundo, M., Adret, P., and Fernandez-Duque, E. (2020). Sexual dimorphism in the loud calls of Azara's owl monkeys (*Aotus azarae*): Evidence of sexual selection? *Primates* 61, 309–319. doi: 10.1007/s10329-019-00773-6

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/s1464793101005826

Geissmann, T., and Nijman, V. (2006). Calling in wild silvery gibbons (*Hylobates moloch*) in Java (Indonesia): Behavior, phylogeny, and conservation. *Am. J. Primatol.* 68, 1–19. doi: 10.1002/ajp.20203

Geissmann, T., and Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*. *Anim. Behav.* 60, 805–809. doi: 10.1006/anbe.2000.1540

Giacoma, C., Sorrentino, V., Rabarivola, C., and Gamba, M. (2010). Sex differences in the song of *Indri indri*. *Int. J. Primatol.* 31, 539–551. doi: 10.1007/s10764-010-9412-8

Green, S. M. (1981). Sex differences and age gradations in vocalizations of Japanese and lion-tailed monkeys (*Macaca fuscata* and *Macaca silenus*). *Am. Zool.* 21, 165–183. doi: 10.1093/icb/21.1.165

Gultekin, Y. B., and Hage, S. R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat. Commun.* 8:14046. doi: 10.1038/ncomms14046

Gultekin, Y. B., Hildebrand, D. G., Hammerschmidt, K., and Hage, S. R. (2021). High plasticity in marmoset monkey vocal development from infancy to adulthood. *Sci. Adv.* 7:eabf2938. doi: 10.1126/sciadv.abf2938

Ham, S., Lappan, S., Hedwig, D., and Choe, J. C. (2017). Female songs of the nonduetting Javan gibbons (*Hylobates moloch*) function for territorial defense. *Int. J. Primatol.* 38, 533–552. doi: 10.1007/s10764-017-9964-y

Harvey, P. H., and Clutton-Brock, T. H. (1985). Life history variation in primates. *Evolution* 39, 559–581. doi: 10.1111/j.1558-5646.1985.tb00395.x

Hradec, M., Bolechová, P., and Svobodová, I. (2016). Production of a female-specific great call in an immature male gibbon, the *Nomascus* genus. *Primates* 57, 445–448. doi: 10.1007/s10329-016-0569-4

Hradec, M., Illmann, G., Bartoš, L., and Bolechová, P. (2021). The transition from the female-like great calls to male calls during ontogeny in southern yellow-cheeked gibbon males (*Nomascus gabriellae*). *Sci. Rep.* 11, 1–9. doi: 10.1038/s41598-021-01648-x

Hradec, M., Linhart, P., Bartoš, L., and Bolechová, P. (2017). The traits of the great calls in the juvenile and adolescent gibbon males *Nomascus gabriellae*. *PLoS One* 12:e0173959. doi: 10.1371/journal.pone.0173959

Jang, H., Oktaviani, R., Kim, S., Mardiastuti, A., and Choe, J. C. (2021). Do Javan gibbons (*Hylobates moloch*) use fruiting synchrony as a foraging strategy? *Am. J. Primatol.* 83:e23319. doi: 10.1002/ajp.23319

Kappeler, M. (1984). "Vocal bouts and territorial maintenance in the moloch gibbon," in *The lesser apes : Evolutionary and behavioural biology*, eds H. Preuschoft, D. J. Chivers, W. Y. Brockelman, and N. Creel (Edinburgh: Edinburgh University Press), 376–389.

Kim, S., Lappan, S., and Choe, J. C. (2012). Responses of Javan Gibbon (*Hylobates moloch*) groups in submontane forest to monthly variation in food availability: Evidence for variation on a fine spatial scale. *Am. J. Primatol.* 74, 1154–1167. doi: 10.1002/ajp.22074

Koda, H. (2016). "Gibbon songs: Understanding the evolution and development of this unique form of vocal communication," in *Evolution of gibbons and Siamang: Phylogeny, morphology, and cognition*, eds U. H. Reichard, H. Hirai, and C. Barelli (New York, NY: Springer), 349–359. doi: 10.1007/978-1-4939-5614-2_15

Koda, H., Lemasson, A., Oyakawa, C., Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS One* 8:e71432. doi: 10.1371/journal.pone.0071432

Koda, H., Oyakawa, C., Kato, A., Shimizu, D., Koyama, Y., and Hasegawa, S. (2014). Immature male gibbons produce female-specific songs. *Primates* 55, 13–17. doi: 10.1007/s10329-013-0390-2

Lemasson, A., Ouattara, K., Petit, E. J., and Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evol. Biol.* 11:362. doi: 10.1186/1471-2148-11-362

Marshall, J. T. Jr., and Marshall, E. R. (1976). Gibbons and their territorial songs. *Science* 193, 235–237. doi: 10.1126/science.193.4249.235

Masataka, N. (1987). The perception of sex-specificity in long calls of the tamarin (*Saguinus labiatus labiatus*). *Ethology* 76, 56–64. doi: 10.1111/j.1439-0310.1987.tb00671.x

Merker, B., and Cox, C. (1999). Development of the female great call in *Hylobates gabriellae*: A case study. *Folia Primatol.* 70, 97–106. doi: 10.1159/000021680

Mitani, J. C. (1985). Gibbon song duets and intergroup spacing. *Behaviour* 92, 59–96.

Mitani, J. C. (1987). Territoriality and monogamy among agile gibbons (*Hylobates agilis*). *Behav. Ecol. Sociobiol.* 20, 265–269. doi: 10.1007/BF00292179

Newman, S.-R., Butler, J., Hammond, E. H., and Gray, S. D. (2000). Preliminary report on hormone receptors in the human vocal fold. *J. Voice* 14, 72–81. doi: 10.1016/s0892-1997(00)80096-x

Oktaviani, R., Kim, S., Cahyana, A., and Choe, J. (2018). "Nutrient composition of the diets of Javan gibbons (*Hylobates moloch*)," in *Proceedings of the IOP Conference Series: Earth and Environmental Science*, Vol. 197, (Bristol: IOP Publishing Ltd), 012048. doi: 10.1088/1755-1315/197/1/012048

Pistorio, A. L., Vintch, B., and Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* 120, 1655–1670. doi: 10.1121/1.2225899

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raemaekers, J. J., and Raemaekers, P. M. (1985). Field playback of loud calls to gibbons (*Hylobates lar*): Territorial, sex-specific and species-specific responses. *Anim. Behav.* 33, 481–493. doi: 10.1016/S0003-3472(85)80071-3

Raemaekers, J. J., Raemaekers, P. M., and Haimoff, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*): Repertoire, organisation and context. *Behaviour* 91, 146–189. doi: 10.1163/156853984X00263

Reichard, U. H. (2003). "Social monogamy in gibbons: The male perspective," in *Monogamy: Mating strategies and partnerships in birds, humans and other mammals*, eds U. H. Reichard and C. Boesch (Cambridge: Cambridge University Press), 190–213. doi: 10.1017/CBO9781139087247.013

Robinson, J. G. (1979). Vocal regulation of use of space by groups of titi monkeys *Callicebus moloch*. *Behav. Ecol. Sociobiol.* 5, 1–15. doi: 10.1007/BF00302691

Roos, C. (2016). "Phylogeny and classification of gibbons (Hylobatidae)," in *Evolution of gibbons and siamang: Phylogeny, morphology, and cognition*, eds U. H. Reichard, H. Hirai, and C. Barelli (New York, NY: Springer), 151–165. doi: 10.1007/978-1-4939-5614-2_7

Schilling, D. (1984). "Song bouts and duetting in the concolor gibbon," in *The lesser apes : Evolutionary and behavioural biology*, eds H. Preuschoft, D. J. Chivers, W. Y. Brockelman, and N. Creel (Edinburgh: Edinburgh University Press), 390–403.

Takahashi, D. Y., Liao, D. A., and Ghazanfar, A. A. (2017). Vocal learning via social reinforcement by infant marmoset monkeys. *Curr. Biol.* 27, 1844–1852.e6. doi: 10.1016/j.cub.2017.05.004

Tenaza, R. R. (1976). Songs, choruses and countersinging of Kloss' gibbons (Hylobates klossii) in Siberut Island, Indonesia. *Z. Tierpsychol.* 40, 37–52. doi: 10.1111/j.1439-0310.1976.tb00924.x

Tilson, R. L. (1981). Family formation strategies of Kloss's gibbons. *Folia Primatol.* 35, 259–287. doi: 10.1159/000155979

Tomasello, M. (2010). *Origins of human communication*. Cambridge, MA: MIT press.

Tomaszycki, M. L., Davis, J. E., Gouzoules, H., and Wallen, K. (2001). Sex differences in infant rhesus macaque separation–rejection vocalizations and effects of prenatal androgens. *Horm. Behav.* 39, 267–276. doi: 10.1006/hbeh.2001.1659

Treesucon, U. (1984). *Social development of young gibbons (Hylobates lar) in Khao Yai National Park, Thailand*. MSc. thesis. Bangkok: Mahidol University.

van Schaik, C. P., and Isler, K. (2012). "Life-history evolution in primates," in *The evolution of primate societies*, eds J. C. Mitani, J. Call, P. M. Kappeler, R. A. Palombit, and J. B. Silk (Chicago, IL: The University of Chicago Press), 220–244.

Whitten, A. J. (1982). The ecology of singing in Kloss gibbons (*Hylobates klossii*) on Siberut Island, Indonesia. *Int. J. Primatol.* 3, 33–51. doi: 10.1007/BF02693489

Winter, P., Schott, D., Ploog, D., and Handley, P. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47, 230–239. doi: 10.1163/156853973x00085

Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70, 164–168. doi: 10.2307/1938423

Yang, K. L. (2022). *Raven lite: Interactive sound analysis software (Version 2.0.3)* [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology.

Yi, Y., Fichtel, C., Ham, S., Jang, H., and Choe, J. C. (2020). Fighting for what it's worth: Participation and outcome of inter-group encounters in a pair-living primate, the Javan gibbon (*Hylobates moloch*). *Behav. Ecol. Sociobiol.* 74, 1–15. doi: 10.1007/s00265-020-02879-0

Zhang, Y. S., and Ghazanfar, A. A. (2018). Vocal development through morphological computation. *PLoS Biol.* 16:e2003933. doi: 10.1371/journal.pbio.2003933

Frontiers in Ecology and Evolution

TYPE Systematic Review
PUBLISHED 01 December 2022
DOI 10.3389/fevo.2022.987253

OPEN ACCESS

EDITED BY
Patrice Adret,
Universidad Autónoma Gabriel René
Moreno, Bolivia

REVIEWED BY
Koen de Reus,
Vrije University Brussel, Belgium
Asif A. Ghazanfar,
Princeton University, United States

*CORRESPONDENCE
Filipa Abreu
filipa.abreu@uni-osnabrueck.de
Simone Pika
simone.pika@uni-osnabrueck.de

SPECIALTY SECTION
This article was submitted to
Behavioral and Evolutionary Ecology,
a section of the journal
Frontiers in Ecology and Evolution

RECEIVED 06 July 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Abreu F and Pika S (2022) Turn-taking skills in mammals: A systematic review into development and acquisition.
Front. Ecol. Evol. 10:987253.
doi: 10.3389/fevo.2022.987253

Turn-taking skills in mammals: A systematic review into development and acquisition

Filipa Abreu* and Simone Pika*

Comparative BioCognition, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

How human language evolved remains one of the most intriguing questions in science, and different approaches have been used to tackle this question. A recent hypothesis, the Interaction Engine Hypothesis, postulates that language was made possible through the special capacity for social interaction involving different social cognitive skills (e.g., joint attention, common ground) and specific characteristics such as face-to-face interaction, mutual gaze and turn-taking, the exchange of rapid communicative turns. Recently, it has been argued that this turn-taking infrastructure may be a foundational and ancient mechanism of the layered system of language because communicative turn-taking has been found in human infants and across several non-human primate species. Moreover, there is some evidence for turn-taking in different mammalian taxa, especially those capable of vocal learning. Surprisingly, however, the existing studies have mainly focused on turn-taking production of adult individuals, while little is known about its emergence and development in young individuals. Hence, the aim of the current paper was 2-fold: First, we carried out a systematic review of turn-taking development and acquisition in mammals to evaluate possible research bias and existing gaps. Second, we highlight research avenues to spur more research into this domain and investigate if distinct turn-taking elements can be found in other non-human animal species. Since mammals exhibit an extended development period, including learning and strong parental care, they represent an excellent model group in which to investigate the acquisition and development of turn-taking abilities. We performed a systematic review including a wide range of terms and found 21 studies presenting findings on turn-taking abilities in infants and juveniles. Most of these studies were from the last decade, showing an increased interest in this field over the years. Overall, we found a considerable variation in the terminologies and methodological approaches used. In addition, studies investigating turn-taking abilities across different development periods and in relation to different social partners were very rare, thereby hampering direct, systematic comparisons within and across species. Nonetheless, the results of some studies suggested that specific turn-taking elements are innate, while others are acquired during development (e.g., flexibility). Finally, we pinpoint fruitful research avenues and hypotheses to move the field of turn-taking development forward and improve our understanding of the impact of turn-taking on language evolution.

KEYWORDS

language evolution, social interaction, development, learning, turn-taking, mammals, primates, infants

Introduction

Language has been proposed to be uniquely human (Christiansen and Kirby, 2003a; Corballis, 2009; McNeill, 2010) because it involves specific characteristics such as high variation, complexity, open-endedness, and the use of linguistic and socially learned symbols to direct the attentional and mental states of others (e.g., Christiansen and Kirby, 2003a; Pika et al., 2005). However, although the evolution of language has intrigued scientific scholars for centuries (Darwin, 1859) and across scientific disciplines (e.g., Christiansen and Kirby, 2003b; Fitch, 2010; Corballis, 2011; Hauser et al., 2014; Killin, 2017), it still remains a mystery (Knight et al., 2000). Attempts to shed light on language evolution have used different approaches and methods (e.g., comparative approach and purpose of language), focused on different research disciplines (e.g., biology, linguistics, and neuroscience), and used different model systems (e.g., songbirds, great apes; for an overview see Fitch, 2010). In addition, several hypotheses have been postulated ranging from different communicative modalities as starting points (e.g., Hewes et al., 1973; Armstrong and Sherman, 2007; Cheney and Seyfarth, 2010; McNeill, 2010), proto-languages (e.g., Wray, 1998) to the purpose of language (e.g., Shannon, 1948; Hauser et al., 2010; Seyfarth et al., 2010a).

One important approach to investigating the evolution of language is the comparative approach, which investigates similarities and differences between human and non-human animal species, especially non-human primates (hereafter primates) to then draw informed inferences about the abilities of our extinct ancestors (Pika, 2015). Due to the analogy to speech, first comparative studies investigated the vocal abilities of primates with a special focus on Hocketts' design features of language (e.g., interchangeability, semanticity, displacement, flexibility, learnability; Kellogg and Kellogg, 1933; Hayes and Hayes, 1951; Hockett, 1960, 1963). The first studies investigating vocalizations of different monkey species showed that distinct call types are characterized by "semanticity" and "arbitrariness," while there is no evidence yet for the features "displacement" and "traditional transmission" (e.g., Hockett and Hockett, 1960; Seyfarth et al., 2010b; Slocombe et al., 2011; Pika and Fröhlich, 2019; Janik and Knörnschild, 2021). Furthermore, the majority of studies on primate vocalizations provided evidence that call morphology and usage seem to have limited flexibility, with learning playing a relatively small role only (e.g., Tomasello and Zuberbühler, 2002; Hammerschmidt and Fischer, 2008; Seyfarth and Cheney, 2010; but see Crockford et al., 2004; Schel et al., 2013).

In parallel, researchers also examined gestural abilities of primates in interactions with humans (e.g., Ladygina-Kohts, 1935), by for instance teaching great ape individuals American Sign language (Gardner and Gardner, 1969; Patterson, 1978) and observing natural communicative interactions between conspecifics (e.g., Call and Tomasello, 2007; Liebal et al., 2012; Schel et al., 2022). The studies showed that gestural signaling of primates involves distinct design features such as interchangeability, semanticity, and arbitrariness (e.g., Pika et al., 2003; Cartmill and Byrne, 2007; Hobaiter and Byrne, 2014; Fröhlich et al., 2016a,b). In contrast to vocalizations, however, some studies also provided evidence for the features productivity, traditional transmission, and flexibility (e.g., Leavens and Hopkins, 1998; Call and Tomasello, 2007; Hobaiter et al., 2017; Pika and Deschner, 2019; Prieur et al., 2020).

However, currently no consensus has been reached concerning the evolutionary trajectory of language (Arbib et al., 2008; Cheney and Seyfarth, 2010; Slocombe et al., 2011; Fischer, 2017; Fröhlich et al., 2019a). In addition, Levinson (2006, 2016) recently proposed the "*Interaction Engine*" hypothesis which suggests that it is not language that makes human communication possible but a special capacity for social interaction. This capacity rests on a layered assemblage of different social cognitive skills, including joint attention (see Box 1 for definitions), common ground, collaboration, and reasoning about communicative intent (Clark, 1996). It also deploys the specific characteristics of face-to-face interaction, frequent employment of mutual gaze, and the exchange of rapid communicative turns—conversational turn-taking (Sacks et al., 1974; Levinson, 2006).

The first systematic framework of conversational turn-taking has been provided by Sacks et al. (1974) in the last century. It consists of an exchange of communicative turns with at least two interlocutors and is governed by specific rules (e.g., avoidance of overlaps, specific temporal relationships, adjacency pairs, communicative repair; see Box 1 for definitions). A recent study by Stivers et al. (2009) investigated the temporal relationships of turns across 10 languages of varied types, geographical locations, and cultural settings and showed that they were all characterized by a similar distribution of response offsets (unimodal peak of response within 200 ms of the end of a given question). The study, therefore, suggested a strong universal basis for turn-taking behavior and emphasizes the antiquity of the turn-taking system. Furthermore, studies on human infants revealed that turn-taking interactions first start around the age of 3 months with infants coordinating actions and signals with caretakers (such as smiles, gaze looking, and facial expressions; Bates et al., 1975; Bateson, 1975; Gratier et al., 2015).

In addition, Levinson and Holler (2014) and Levinson (2016) proposed that the turn-taking infrastructure for conversations may be one of the most ancient layer of the language system with evolutionary precursors present in all the major primate branches. Subsequently, Pika et al. (2018) provided a comprehensive overview of the existing research on turn-taking and related phenomena in the animal kingdom. They showed that although the study of turn-taking abilities in animal species has been growing in the last decades (e.g.,

BOX 1 Definitions of terms used.

Adjacency pair-like sequences	The term adjacency pair sequence was first used by conversational analysts (Schegloff, 2007; Levinson, 2013) and is composed of two utterances by two speakers, one after the other such as question-answer or greeting-greeting response. It was adapted by researchers of animal communication and described as a sequence of at least two turns between different animal individuals involving matched signal-response sequences (such as interactions with the same call types; Luef and Pika, 2017; Pika et al., 2018).
Antiphonal communication	This term refers to non-human animals. Exchange of signals between two or more individuals characterized by a call-and-response interaction (Yoshida and Okanoya, 2005; Pika et al., 2018).
Communicative repair	This term refers to human and non-human animals. The ability to "fix" a signal by the signaler (self-repair) or the recipient (other-repair) after the initial communication attempt failed using strategies such as modifying, repeating or revising the signal or using a different signal (e.g., Sacks et al., 1974; Dingemanse et al., 2015).
Duetting	This term refers to non-human animals. Coordinated exchange of stereotyped signals that follow a temporal precision between the first signal and its reply, generally in the vocal modality (e.g., mating system in some insect species; Yoshida and Okanoya, 2005; de Reus et al., 2021).
Flexibility of turn-taking organization	This term refers to non-human animals. The element mirrors the ability to voluntarily change, adjust and combine signals/actions and thus the degree of underlying cognitive flexibility. It can be operationalized by quantifying the number, frequency and degree of repetition of signals and actions produced in turn-taking events, their combination (e.g., A-B-A; A-B-C), distribution of roles between participants (e.g., role reversal), and intentionality involved (e.g., persistence; Sacks et al., 1974; Pika et al., 2018).
Intentionality	This term refers to non-human animals and can be divided into different levels (Dennett, 1983). In the first-order intentionality, the signaler produce a signal with the intention of producing a response in the recipient, thus recognizing that the recipient is a distinct agent; however, it does not require that recipient will understand (Dennett, 1983; Townsend et al., 2017; Hobaiter et al., 2022). Second-order intentionality is more complex and requires the signaler recognize that the recipient is a distinct individual with his own "potentially distinct, understanding and knowledge of the world" (Dennett, 1983; Hobaiter et al., 2022, p. 5).
Joint attention	This term refers to human and non-human animals. The capacity to engage and coordinate the attention of a recipient toward a third entity (e.g., individuals or objects; Bakeman and Adamson, 1984; Mundy and Newell, 2007; Mocha et al., 2019).
Bimodal- and multimodal communication	In comparative gesture research (but see for behavioral ecologists: Partan and Marler, 1999; Hebets and Papaj, 2005), these terms refer to the simultaneous or sequential integration of signals from at least two 'modalities' (Liebal et al., 2013; Luef and Pika, 2017) and represent a form of signal flexibility (e.g., Davila-Ross et al., 2015). Modalities denote signal categories such as facial expressions, gestures and vocalizations.
Participation-framework	Establishment of the interaction partners and the choice of the individual that will be part of the communicative interaction (e.g., Goodwin and Heritage, 1990; Gibson, 2003; Levinson, 2013). In non-human animals, parameters to measure participation frameworks could involve (i) body orientation toward recipient(s), (ii) gaze direction of signaller, (iii) response waiting, and (iv) whether recipient(s) can perceive the signal (e.g., being in the visual or auditory field; Pika et al., 2018).
Singing behavior	This term refers to non-human animals. Notes with different frequency variation and hierarchical structures that can be uttered solo (one individual only), in a duet (dyadic interaction), or in a chorus (multiple individuals; Pika et al., 2018; De Gregorio et al., 2022).
Temporal relationships	First described by Sacks et al. (1974) as a system of allocation of single turns between two speakers with few gaps or overlaps during transitions. For the animal communication field, it was described as the time elapsed between a turn and its response (Stivers et al., 2009; Pika et al., 2018). If this time is positive and with little gap between turns, it is referred to as a standard conversation gap. If this time is positive and with a long gap, it is referred to as a delayed response. If the time is negative, it is considered an overlap between signal and response (adapted from Fröhlich et al., 2016c). These timings depend on the study species.
Turn-taking production	This term refers to human and non-human animals. Turn-taking production: An individual's capacity to reply to a first turn or initiate a turn-taking interaction by using communicative signals or actions (adapted from Seyfarth and Cheney, 2010). For the present study, turn-taking production was defined as producing communicative signals (e.g., gestures, vocalizations, oro-facial sounds, and bi-modal signals) or behaviors resulting in a turn-taking interaction.
Turn-taking comprehension	This term refers to human and non-human animals. Turn-taking comprehension: An individual's capacity to understand that other individuals engage in a turn-taking interaction (adapted from Seyfarth and Cheney, 2010). Here, we considered only those studies that involved individuals observing and reacting to turn-taking interactions of other interactants.

Miller et al., 2004; Méndez-Cárdenas and Zimmermann, 2009; Morisaka et al., 2013; Takahashi et al., 2015; Terleph et al., 2018; see Figure 1), the field is strongly biased toward investigations involving primates (e.g., Rossano, 2013; Takahashi et al., 2013; Fröhlich et al., 2016c; Snowdon, 2017; Pougnault et al., 2020).

Furthermore, Vanderhoff and Bernal Hoverud (2022) recently provided an overview of communicative exchanges in non-primate mammals with a special focus on antiphonal calling, duetting, and counter-singing (for definitions of terms, see Box 1) and showed that some singing species also possess turn-taking abilities. For instance, Alston's signing mice (*Scotinomys teguina*) show the ability to combine singing with turn-taking behaviors. These results are aligned with studies on several ape and monkey species, which exhibit some aspects of conversational turn-taking involving signal-signal and signal-action exchanges (*Callithrix jacchus*: Chow et al., 2015;

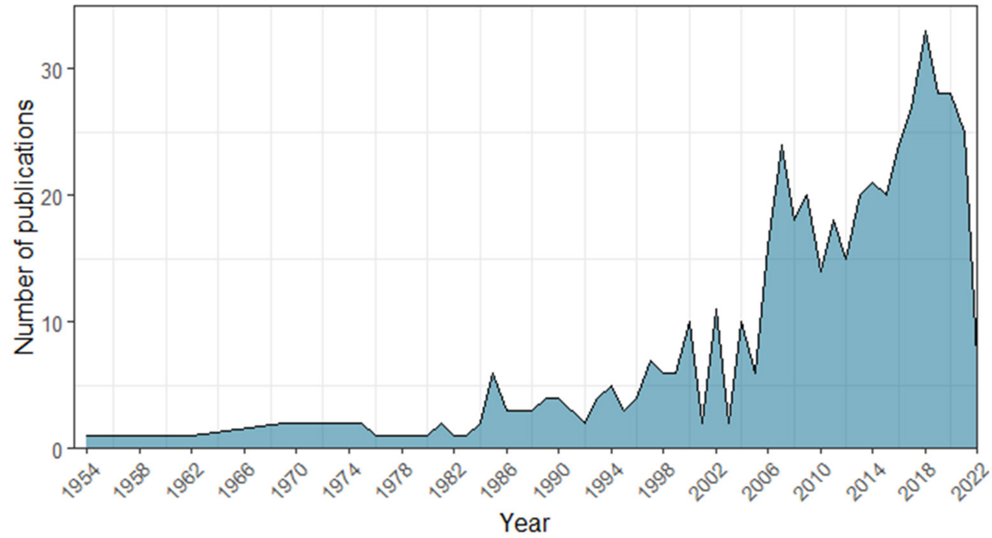


FIGURE 1
Number of studies indexed in the Scopus and Web of Science databases that included the terms "duet" or "turn-taking" or "antiphonal calling" combined with the words "animal" or "non-human" or "Animalia" or "fauna" across years (1954–2022).

Takahashi et al., 2015; *Gorilla g. gorilla*: Luef et al., 2016; *Pan paniscus* and *Pan troglodytes*: Fröhlich et al., 2016c; *Cercocebus torquatus*: Aychet et al., 2021; *Alouatta pigra*: Briseño-Jaramillo et al., 2021; *Pan troglodytes*: Pougnault et al., 2021a), distinct temporal relationships (avoiding overlap and presence of gaps between signal-response; e.g., *Callithrix jacchus*: Yamaguchi et al., 2009; *Macaca fuscata*: Katsu et al., 2018; *Hylobates lar*: Terleph et al., 2018; *Indri indri*: De Gregorio et al., 2019), and adjacency pair-like sequences (*Hylobates agilis:* Koda et al., 2013; *Macaca fuscata*: Bouchet et al., 2017; *Ateles geoffroyi:* Briseño-Jaramillo et al., 2018). Other mammal species and groups such as bats, cetaceans, meerkats, and Alston's singing mice, also show some elements of conversational turn-taking. For instance, they interact and exchange signals, mainly vocal ones, and these turn-exchanges adhere to specific temporal relationships (e.g., *Loxodonta africana*: Leighty et al., 2008; O'Connell-Rodwell et al., 2012; *Physeter macrocephalus*: Schulz et al., 2008; *Diaemus youngi*: Carter et al., 2009; *Delphinapterus leucas*: Morisaka et al., 2013; *Suricata suricatta*: Demartsev et al., 2018; *Lagenorhynchus obliquidens*: Mishima et al., 2018; *Scotinomys teguina*: Okobi et al., 2019; for a review see Vernes, 2017).

Overall, specific temporal relationships in turn-taking interactions have been found in a variety of non-mammal taxa including amphibians, birds, and insects (see for a recent review Pika et al., 2018; Pougnault et al., 2020; de Reus et al., 2021). Pika et al. (2018), however, concluded that considerable methodological confounds and the employment of different terminologies in the existing studies (e.g., antiphonal calling and duetting) have significantly hampered insightful comparisons across species and an in-depth understanding of turn-taking complexity. To counteract these problems, they proposed a new comparative framework focusing particularly on four key elements of human social action during conversations: involved flexibility, adjacency-pair like sequences, temporal relationships, and participation-framework (Pika et al., 2018; see Box 1 for definitions).

Furthermore, relatively little is also known about the development of turn-taking skills and the acquisition of involved elements (Levinson, 2016; Pika et al., 2018). This is surprising since especially long-living mammal species exhibit extended developmental periods, including social learning, strong parental care, and cooperative behaviors (e.g., Kappeler and Van Schaik, 2005; Yamamoto, 2005; Hudson and Trillmich, 2008; Kerth, 2008; Clutton-Brock, 2009; Kölliker et al., 2012; Rosenbaum and Gettler, 2018; Janik and Knörnschild, 2021). These characteristics make them an excellent group to investigate and understand the acquisition and development of turn-taking abilities and draw inferences about its phylogenetic trajectory.

To date, most studies on turn-taking development have concerned two cooperative breeding species, humans (*Homo sapiens*; e.g., Henrich et al., 2010; Hilbrink et al., 2015; Nomikou et al., 2017) and common marmosets (*Callithrix jacchus*), a New World monkey (Chow et al., 2015; Takahashi et al., 2016). The studies on human children were strongly biased toward individuals living in western, educated, industrialized, rich and democratic (WEIRD) societies (Henrich et al., 2010). Furthermore, investigations in both species focused on production rather than comprehension of turn-taking (e.g., Chow et al., 2015; Hilbrink et al., 2015; Takahashi et al., 2015) and specifically examined the onset of turn-taking, temporal relationships, and the role of learning (Snow, 1977; Jaffe et al., 2001; Casillas et al., 2016; Takahashi et al., 2016). Overall, turn-taking production in these two highly social primate species seemed to start relatively early during development (e.g., 0–2 months: Takahashi et al., 2016; 4–6 months: Nomikou et al., 2017), with some studies suggesting that distinct elements (such as temporal relationships and adjacency pair-like sequences) are learned and rely on input and active shaping by caretakers (Chow et al., 2015).

Concerning other species and taxa, however, an in-depth understanding of the acquisition and development of turn-taking and involved elements is currently missing (but see Briseño-Jaramillo et al., 2018; Fröhlich et al., 2019b; Araya-Salas et al., 2020; Ames et al., 2021; Dafreville et al., 2021). Moreover, the few studies available have used different terms, research approaches, and focused on non-comparable age classes, thereby hampering cross-species comparisons and a general understanding of the learning processes involved. Hence, the goal of the current review was 2-fold: First, we aimed to provide a comprehensive overview of the current knowledge of turn-taking acquisition and development in non-human mammals with a special focus on methodologies employed, distribution of studies across species and modalities and components of communication investigated to identify current gaps and research biases. Second, we pinpoint fruitful research avenues to spur more research into this domain and to gain a better understanding of the role of learning, shaping and social tradition for turn-taking development and involved elements.

Methods

Search protocol

We applied the PRISMA search protocol (O'Dea et al., 2021) and used the online search engines *Scopus* and *Web of Science*. The following terms were utilized to search titles, abstracts, and keywords of publications: "turn-taking," "taking turns," "conversation," "duet*," "antiphon*," "chorus*," "communicative interaction," "communicative interactions," "interactive communication," "interactive communications," "dialog*," "vocal exchang*," "vocal cooperation," "vocal production," "vocal sequence," "vocal interact*," "vocal timing," "vocal overlap," "verbal exchang*," "verbal cooperation," "verbal production," "verbal sequence," "verbal interact*," "verbal timing," "verbal overlap," "call exchang*," "call cooperation," "call production," "call sequence," "call interact*," "call timing," "call overlap," "signal exchang*," "signal cooperation," "signal production," "signal sequence," "signal interact*," "signal timing," "signal overlap," "gesture exchang*," "gesture cooperation," "gesture production," "gesture sequence," "gesture interact*," "gesture timing" and "gesture overlap" combined with the words "develop*," "learn*," "ontogen*," "age," "offspring," "cub," "infant," "calf," "group," "descendant," "young*," "litter," "progeny," "bab*," "pup*," "calves," "piglet*," "juvenile" and "immature." We also combined all these terms with "mammal*," "primate," "monkey," "chiroptera," "bat," "rodent*," "rat," "soricomorpha," "carnivora," "fox," "wolf," "bear," "racoon," "dog," "cat," "mongoose," "hyena," "bear," "weasel," "pinniped," "seal," "ungulate," "cetacean," "whales," "dolphin," "porpoise," "beluga," "pig," "hippopotamus," "antelope," "deer," "giraffe," "camel," "llama," "alpaca," "sheep," "goat," "cattle," "marsupial," "kangaroo," "koala," "wallaby," "wombat," "possum," "lagomorph," "pika," "rabbit," "marsupial," "opossum," "mole," "hedgehogs," "armadillos," "shrew," "horse," "zebra," "rhinoceroses," "tapir," "elephant," "sloth," "echidna." The search (December 2021) returned a total of 2098 manuscripts (without duplicates).

Since the use of the term "turn-taking" has only increased in the last decade (see Figure 2), with former studies applying different terms to refer to turn-taking abilities (such as antiphonal conversation or duetting; Pika et al., 2018; Ravignani et al., 2019), we designed our search protocol as broad as possible. Similarly, we also used both the taxonomic order and common names to search for the groups of organisms of our interest (e.g., bats and chiropteran).

Eligibility criteria and data extraction

To evaluate the manuscripts, we first read the manuscript titles and abstracts and selected only those studies that presented empirical findings on the acquisition and development of coordinated exchanges in non-human mammal species. At this stage, we included in our review all those publications that investigated communicative interactions, defined as an exchange of signals or actions between at least two individuals (where one of them needed to be an infant/juvenile). We also included studies in this review that did not (1) use the term "turn-taking"; (2) focus specifically on turn-taking interactions, and (3) measure specific elements involved in conversational turn-taking (see for definitions Pika et al., 2018). Since the field of comparative turn-taking is a new one, we aimed to provide a relatively broad overview to inform this useful research avenue and inspire future research. We excluded event publications, theses, reviews, methodological articles, and publications containing only human findings.

This method resulted in a total of 74 manuscripts, which were then screened to assess whether they matched the criteria outlined above. We excluded 45 manuscripts and maintained a total of 29 studies matching the selection criteria. To expand our search, we applied the PRISMA protocol (O'Dea et al., 2021) and verified the bibliographic references of the articles chosen. This method resulted in a total of five additional

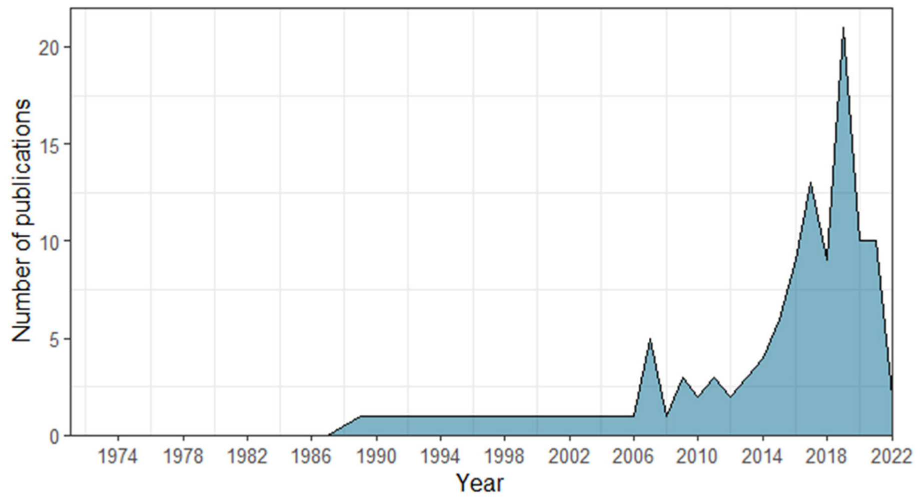


FIGURE 2
Number of non-human animal studies indexed in the Scopus and Web of Science databases, including the term "turn-taking" across years (1971–2022).

articles. These contained relevant information about turn-taking development and acquisition but did not appear in our search protocol due to the lack of key search terms in the title, abstract, and keywords (for more details on the extraction process, see Supplementary Figure 1).

The articles selected in the previous stage were then systematically screened and read to enable data extraction. Here, we followed a specific "protocol" spreadsheet (e.g., O'Dea et al., 2021; Rodrigues et al., 2021; Ferreira et al., 2022; Supplementary material) to extract the following data from text, figures, tables, or Supplementary material: (i) article information; (ii) subjects studied; (iii) methodology used; (iv) components of the turn-taking investigated; (v) modalities studied and (vi) elements of turn-taking explored following the framework proposed by Pika et al. (2018; Table 1). The information from *i* to *vi* was always available in the main manuscript text or Supplementary material. Therefore, we did not contact any authors requesting additional data.

Data analysis

Since we included all publications that presented findings on communicative exchanges, we decided to provide as a first step descriptive statistics to better assess the existing knowledge on the acquisition and development of turn-taking abilities in mammals. Therefore, we clustered the publications into two main categories: Studies that (1) failed or (2) succeeded to present findings that could increase our knowledge of turn-taking abilities in infant/juvenile individuals. Following Pika et al. (2018; Box 1) we defined turn-taking as "purely communicative signals or behaviors between individuals characterized by principles for the coordination of turn transfer, which result in observable temporal regularities. The communicative signals delivered by turns can vary, as can the size and the order of turns, and techniques used to allocate turns to specific individuals." Second, the studies included in the second category also required to include assessments of at least one element characterizing human social action during conversations (see Pika et al., 2018). For all publications incorporated into this category, we used descriptive statistics (absolute number, frequency, etc.) to compare trends and biases according to (i) terminology used; (ii) taxa used in studies of the development of turn-taking abilities; (iii) research design; (iv) components and modalities of communication; (v) elements of the comparative turn-taking framework; (vi) social factors investigated, and (vii) number of studies that investigated the development of turn-taking abilities over time. We also included and descriptively reported the results of studies that showed a link between turn-taking elements and development.

Results

Overall, we found 34 studies that reported findings concerning communicative interactions (signal-signal or signal-action) between at least one infant or juvenile individual and another conspecific. Thirteen of these studies (38%) were categorized as studies that failed to provide information about turn-taking abilities. They were biased toward species of the primate order (62%), followed by species of the order artiodactyla (14%), chiroptera (8%), rodentia (8%), and proboscidea (8%). Five of these studies focused on the

TABLE 1 Information extracted from the articles.

Categories	Information extracted
Characteristics of subjects	Taxon, species, number of individuals, age of individuals.
Methodology	Terminology used, design (experimental, observational or both), research setting (captivity or natural environment), ontogeny (across development or in one specific age-class), developmental milestones reported[a] (yes/no, which), and social partners (interaction partner).
Components of turn-taking	Production or comprehension, interactors (signaler or recipient).
Modalities	Vocal, gestural, action, facial expression, and multimodal.[b]
Comparative turn-taking framework	Descriptors used in the different elements (flexibility, temporal relationship, adjacency pair-like sequence, and participation-framework).

Table shows all information extracted from the articles including specific characteristics of the subjects used, the methodology employed, components and modalities of communication addressed, and elements examined based on the comparative turn-taking framework proposed by Pika et al. (2018).

[a] If the study reported developmental milestones for the individuals tested (e.g., start of locomotory or feeding independence, ceasing of breastfeeding).

[b] Use of different signal types combined (e.g., gestures and vocalizations; for a detail definition see Box 1).

development of communicative interactions (38%), while eight did not address this aspect (62%). None of these studies investigated any developmental markers important to perform comparisons across species such as weaning, locomotion, or feeding independence (see Supplementary Table 1). Some of these studies addressed distinct elements characterizing human social action during conversations (Pika et al., 2018) such as flexibility of turns and temporal relationships. For instance, one study focused specifically on assessing intentionality, thereby also enabling inferences about the flexibility involved in communicative interactions (chimpanzee *Pan troglodytes*: Bard et al., 2014). Another one examined the timing between the onset of a vocalization and the onset of an action (baboons *Papio cynocephalus ursinus*: Fischer et al., 2000). The remaining studies focused on vocal recognition and interactions between mother-infant dyads (38%, e.g., Cow *Bos taurus*: Marchant-Forde et al., 2002; Sheep *Ovis aries*: Sèbe et al., 2010) or investigated which elements characterized coordinated communicative exchanges [46%, e.g., Gibbons, duets: *Nomascus gabriellae* (formerly genus *Hylobates*): Merker and Cox, 1999; *Hylobates agilis* and *Hylobates lar*: Koda et al., 2014; *Nomascus leucogenys*: Hradec et al., 2016; antiphonal calling: Bulldog bat *Noctilio albiventris*, Brown et al., 1983, see Box 1 for definitions]. For instance, Elowson et al. (1998) observed in a group of pygmy marmosets (*Cebuella pygmea*) in captivity that until the age of 20 months, crying infants were more likely to change the behavior of a given adult individual (by being carried, being groomed, or getting the opportunity to climb on the back more often) than non-crying infants. Moreover, an experimental study performed on mice (*Mus musculus*) in captivity reported that mothers responded to the calls of their newborn offspring by increasing the frequency of their maternal behavior (e.g., licking and changing their body position: Ehret and Bernecker, 1986).

The remaining 21 studies involved findings to aid in increasing our knowledge of turn-taking skills in infant or juvenile individuals (62%; see Supplementary Table 1 for more details). In the following paragraphs, we will introduce and discuss these studies in more detail by paying specific attention to (i) terminology used, (ii) distribution across taxa, (iii) research design utilized, (iv) social factors, (v) development, (vi) components and modalities of communication investigated, and (vii) involved turn-taking elements.

Terminology

The first scientific article addressing some elements characterizing turn-taking was published in the 1970's (Matsumura, 1979). It did not explicitly use the term "turn-taking," but described behaviors exchanged between mother-infant dyads of horseshoe bats (*Rhinolophus ferrumequinum nippon*). The first paper using the term "turn-taking" was published in the twenty-first century by Lemasson et al. (2011). It focused on the production and comprehension of communicative turn-taking in one group of Japanese macaques *(Macaca fuscata)* in captivity involving five juvenile individuals. From the 21 citations extracted, only six studies used the term "turn-taking" (29%). Moreover, only one study (5%) defined the terminology in the methods section (Chow et al., 2015), and only one paid attention to specific elements of turn-taking (Fröhlich et al., 2016c).

In sum, the majority of articles found and extracted did not use the term turn-taking but referred indirectly to turn-taking interactions by utilizing terms such as "exchanges" and/or "interactions." The few studies that specifically used the term "turn-taking" were conducted on primate species and were published in the last decade. This is probably due to the coining of the term and predominant usage in the field of conversational analysis (Sacks et al., 1974), with comparative researchers only recently grasping its importance and implications for language

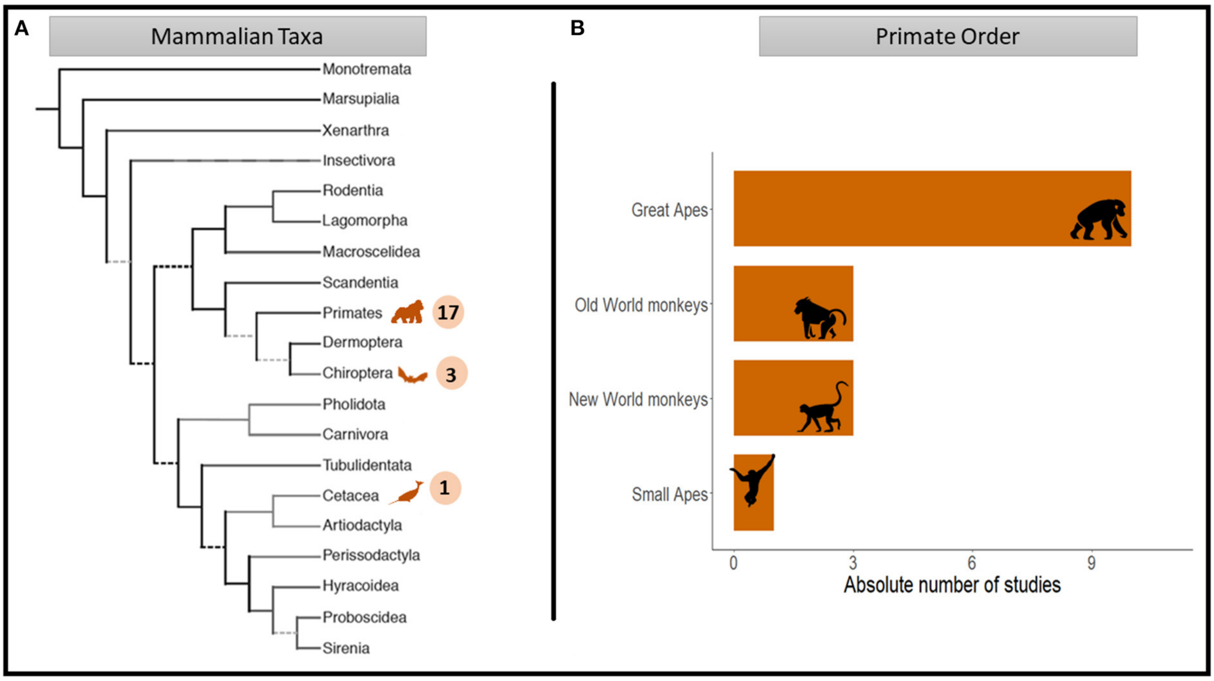


FIGURE 3
Mammal orders represented in the studies reviewed as a function of the total numbers of turn-taking studies distributed across **(A)** mammal orders ($n = 21$) and **(B)** the primate order ($n = 17$; adapted from Figure 1A in Springer et al., 2004).

evolution (Pika et al., 2018; Rossano, 2018; Ravignani et al., 2019).

Distribution across taxa

Across all 21 studies, a total of 12 different species were investigated. Most of the studies focused on primates (81%), followed by chiropterans (14%) and cetaceans (5%). Within the primate studies, the majority focused on great ape species (59%), with a strong bias toward chimpanzees (*Pan troglodytes*; 60%). Turn-taking interactions of Old and New World monkeys were reported in a comparable number of studies (17%), followed by small apes (6%; see Figure 3). Only one study investigated turn-taking abilities in more than one species (chimpanzees and bonobos *Pan paniscus*: Fröhlich et al., 2016c).

Research designs used

Of the 21 studies considered here, 14 (67%) used observational methods, and seven (33%) applied experimental set-ups. More than half of the observational studies were performed with individuals and species living in their natural environments (64%), while the others were carried out in captive settings (36%). Of those studies using experimental designs, only one was conducted with a species living in its natural environment (14%), while all other studies were performed with species living in captive settings (86%; see Figure 4). The number of individuals observed and tested showed a relatively high variation: Nine studies included <10 individuals (43%), nine between 10 and 13 individuals (43%), and three studies included more than 13 individuals (14%). For example, studies on narwhals (*Monodon monoceros*: cetacean) and bonobos (primates) used only one and two individuals, respectively, while the study on horseshoe bats (chiropteran) included a total of 26 individuals. The ages of the individuals investigated in the studies were also relatively broad. For instance, nine studies included infant individuals (ranging from 4 days to 2 years depending on the species, 43%), five investigated turn-taking abilities in juvenile individuals (ranging from 4 months to 4 years, 24%), and seven included both infant and juvenile individuals in their studies (33%; see Figure 5). From those, five compared infant and juvenile individuals or younger and older infant age classes (71%). Only four studies investigated the use and onset of turn-taking abilities of individuals directly after birth (but see for common marmosets, Takahashi et al., 2016; Spix's disc-winged bats Thyroptera tricolor, Araya-Salas et al., 2020).

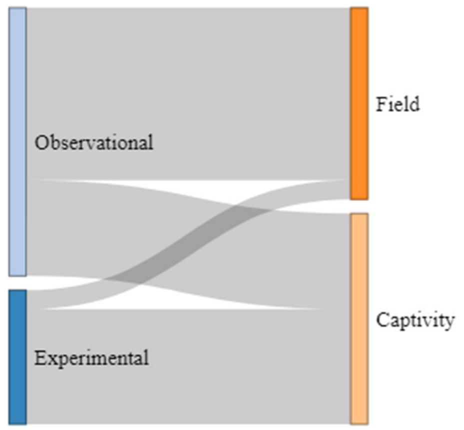


FIGURE 4
Distribution of research designs employed in the reviewed studies.

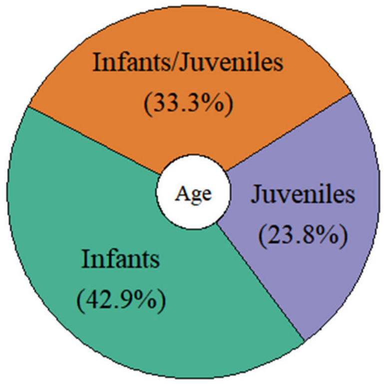


FIGURE 5
Different age classes investigated in the reviewed studies.

In sum, we found a wide variety concerning the research designs employed, age groups and number of individuals tested, and very little research concerning turn-taking skills of very young and newborn individuals.

Social factors

The majority of studies that investigated turn-taking acquisition and development focused solely on mother-infant dyads (48%), followed by interactions with conspecifics (individuals from the whole social group; 43%), and parents and siblings (9%; see Figure 6). All species were observed/tested interacting only with their close family members (mothers, fathers, and siblings). However, the studies examining interactions between infants/juveniles and their group members mainly concerned primate species (but see the study on Spix's disc-winged bats, Araya-Salas et al., 2020). For instance, Fröhlich et al. (2016a), who studied the communicative behavior of mother-infant dyads of two different chimpanzee communities (Kanyawara, Kibale National Park, Uganda; Taï South, Taï National Park, Côte D'Ivoire) belonging to two different subspecies (*Pan troglodytes schweinfurthii*; *Pan troglodytes verus*) observed that the majority of play initiations by infants were produced toward mothers than toward other individuals. Similarly, in a subsequent paper, Fröhlich et al. (2019b) showed that the likelihood of receiving an inappropriate response across the contexts of joint travel, social play, and food sharing was higher when chimpanzee infants interacted with non-maternal conspecifics than with their mothers. Moreover, Chow et al. (2015) studied communicative interactions of two groups of common marmosets in captivity. They showed that juveniles aged 10–12 months started to interact with their mothers much earlier than their fathers. In addition, the juveniles were more likely to interrupt their fathers but not their mothers' vocalization during Phee call exchanges (used for group coordination; Bezerra and Souto, 2008). Also, the frequency of vocal exchanges decreased in the 1st year of life when exchanging signals with parents but remained constant in sibling-sibling interactions (Chow et al., 2015).

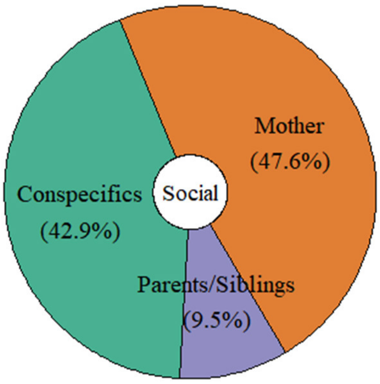


FIGURE 6
Distribution of the distinct social partners interacting with infants/juveniles in the reviewed studies.

In sum, these studies suggest that parents in some pair-bonded primate species (e.g., owl monkeys and common marmosets) and mothers in polygamous species (e.g., gorillas and chimpanzees) seem to play a crucial role in the acquisition and shaping of distinct turn-taking elements. They act as role models, and shape their infants' responses by providing them with appropriate responses, possibly qualifying as teaching (Musgrave et al., 2016). Teaching—high-fidelity social learning—occurs in the presence of a naïve learner, involves some cost or at least no benefit to the teacher, and facilitates learning in another individual (Caro and Hauser, 1992).

Development of turn-taking

Across all studies considered here, eight studies did not address the development of turn-taking abilities and elements (38%). In contrast, they reported turn-taking skills at a specific time point (e.g., juveniles) or a specific age (e.g., 2 months). For instance, Matsumura (1979) found that captive 1-week-old horseshoe bats, when separated from their mothers, emitted "attractive" calls that only ceased after the mothers approached and took their infants under the wings. Furthermore, the only two studies that addressed turn-taking comprehension showed that juvenile individuals did not adhere to the respective "turn-taking rules" more (e.g., call matching) in comparison to adult individuals (Campbell's monkeys: Lemasson et al., 2011; Japanese macaques: Bouchet et al., 2017). However, the authors did not evaluate the development of turn-taking skills across different age classes. Similarly, Ames et al. (2021) and Knörnschild and von Helversen (2008) observed the behavior of wild narwhals at 5 months and wild greater sac-winged bats at the age of 6 weeks (*Saccopteryx bilineata*) respectively and showed that at this young age infants already replied vocally to the vocalizations of their mothers.

Additionally, 13 of the 21 studies addressed the development of turn-taking abilities across different age classes (62%). Seven of these studies investigated the development of turn-taking skills continuously (54%), followed by distinctions between age classes (e.g., "infant-juvenile-adolescent" or "baby-younger-older;" 31%) and two specific developmental time points (e.g., "volants and non-volants" which refers to the ability to fly; 15%). The studies addressing turn-taking interactions continuously across a specific ontogenetic time period showed that younger individuals seemed to improve certain elements (e.g., temporal relationships and adjacency pair-like sequences) across ages by stopping to overlap parents' vocalizations and increase call matching. They, therefore, argued that these elements might be shaped and adjusted during ontogeny (e.g., common marmosets: Chow et al., 2015; Takahashi et al., 2016). Furthermore, one of the studies that compared turn-taking skills across different age classes reported that young and old immature spider monkeys living in their natural environments replied less frequently and answered less with the same call type than adults (Briseño-Jaramillo et al., 2018).

Only a minor proportion of these articles investigated developmental milestones with regard to turn-taking development (23%). The examined developmental milestones were locomotion independence ("volant and non-volant:" if an individual can fly; Araya-Salas et al., 2020) and independence from the mother (breastfeeding and locomotion; Genty, 2019 and Dafreville et al., 2021). For instance, Araya-Salas et al. (2020) showed that very young bats (around 5 days of age) living in their natural environments that could not yet fly produced response calls. They also uttered first inquiry calls at the age of 40 days (vocalizations produced when already volant and only during flight) when held on the experimenter's hand, mimicking flight conditions. These findings may suggest that the motivation to respond and engage in turn-taking interactions may be present in some species from very early on, and may not be learned from conspecifics, and only produced in different stages of the development or in the presence of a specific stimulus.

Furthermore, two studies on the development of two great apes' species (chimpanzees and bonobos) provided insights that may be useful to draw inferences to the development of turn-taking skills. For instance, Dafreville et al. (2021), re-using a data set collected on a chimpanzee community in Uganda, showed that it is only when chimpanzees gain full independence from their mothers (around 103–180 months) that they are capable of adjusting the type of gestural signals to the mother's visual attention (considered by the authors when the mother had a full view of the infant) during a signal-action turn-taking interaction. Similarly, Genty (2019), who studied the behavior of seven bonobo infants living in "Lola ya Bonobo" sanctuary, DRC, reported that as infants become more independent from their mothers, their gesture specificity during signal-action turn-taking interaction also increases, showing a developing of adjacency pair-like sequences element across ages. These studies suggest that distinct communication modalities can have different developing times and a need for different cognitive capacities, possibly with vocal responses preceding gestural ones (Fröhlich et al., 2016b). Therefore, the first appearance of turn-taking skills may change according to the modality observed (see Box 1 for definitions of bimodal and multimodality).

In sum, although almost none of the previous studies specifically addressed the development of turn-taking abilities, they suggested that individuals at relatively early ages are capable of engaging in some form of turn-taking with closely related individuals, mainly mothers. However, full-blown adult-like turn-taking abilities may only be present with increasing age.

Components and modalities of turn-taking

Almost all studies investigated the production (90%) but not the comprehension of turn-taking (5%). One study examined both the production and the comprehension (5%). Of the 19 studies that addressed the production of turn-taking, 10 focused solely on the behavior of the initiator of the interaction (53%), and one addressed the recipient's behavior (5%). The remaining studies investigated both the signalers' and the recipients' behavior (42%).

The majority of studies investigated one modality of turn-taking interactions (e.g., gestural or vocal; 67%) only. Of the 10 studies that addressed the signalers' behavior, three investigated vocalizations only (30%), three gestures only (30%),

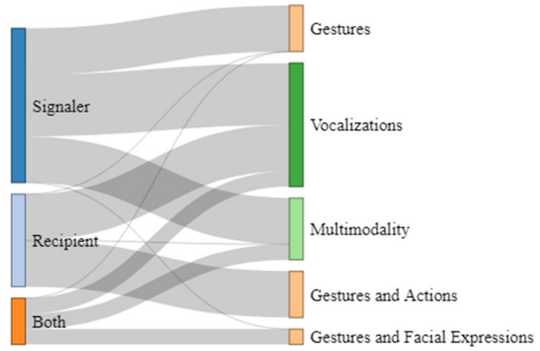


FIGURE 7
Components and modalities of turn-taking interactions addressed in the reviewed studies.

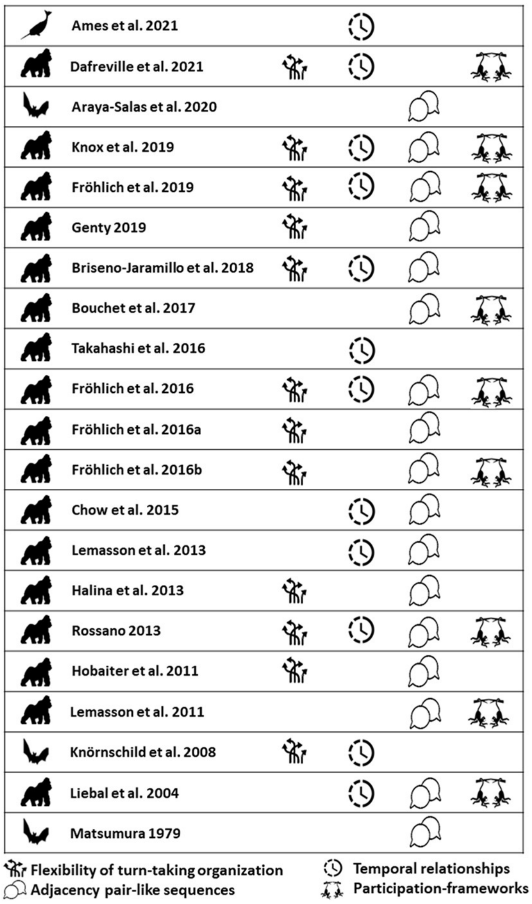
and one focused on gestures and facial expressions (10%). The other three studies examined multimodal communication (30%) in different ape species (chimpanzees: Fröhlich et al., 2019b; bonobos: Genty, 2019; siamang: Liebal et al., 2004). The study investigating the behavior of the recipient considered vocalizations only (100%). Concerning the studies that did not discriminate between signalers and recipients, the modalities investigated were mainly vocal (50%), followed by gestural signals and actions (38%) and multimodal signaling (12%; see Figure 7).

The comparative turn-taking framework

Adopting a previously proposed framework by Pika et al. (2018), we analyzed which papers examined four main key elements of human conversational turn-taking with regards to development: Flexibility of turn-taking organization, temporal relationships, adjacency pair-like sequences, and participation framework (presented below in more detail). In sum, from the 21 studies reviewed here, six measured a single element (29%), seven two elements (33%), and four investigated three elements (19%). All four elements were only investigated in four of the 21 studies (19%). The results also showed that all studies that paid attention to all four elements were carried out in the current century (2010 forward, see Figure 8).

Flexibility of turn-taking organization

Twelve of the 21 articles investigated abilities crucial to voluntarily influence and adjust signals/actions (57%). Moreover, only four studies investigated interactions using a multimodal approach (14%). Although not explicitly focusing on turn-taking interactions, they may help inform future studies on turn-taking development. We therefore provide a brief overview here. The 12 studies focused on four specific



🐦	Ames et al. 2021		✓		
🦍	Dafreville et al. 2021	✓	✓		✓
🦇	Araya-Salas et al. 2020			✓	
🦍	Knox et al. 2019	✓	✓	✓	✓
🦍	Fröhlich et al. 2019	✓	✓	✓	✓
🦍	Genty 2019	✓		✓	
🦍	Briseno-Jaramillo et al. 2018	✓	✓	✓	
🦍	Bouchet et al. 2017			✓	✓
🦍	Takahashi et al. 2016		✓		
🦍	Fröhlich et al. 2016	✓	✓	✓	✓
🦍	Fröhlich et al. 2016a	✓		✓	
🦍	Fröhlich et al. 2016b	✓		✓	✓
🦍	Chow et al. 2015		✓	✓	
🦍	Lemasson et al. 2013		✓	✓	
🦍	Halina et al. 2013	✓		✓	
🦍	Rossano 2013	✓	✓	✓	✓
🦍	Hobaiter et al. 2011	✓		✓	
🦍	Lemasson et al. 2011			✓	✓
🦇	Knörnschild et al. 2008	✓	✓		
🦍	Liebal et al. 2004		✓	✓	✓
🦇	Matsumura 1979			✓	

Flexibility of turn-taking organization
Temporal relationships
Adjacency pair-like sequences
Participation-frameworks

FIGURE 8
Overview of the four key elements of the comparative framework proposed by Pika et al. (2018) to enable systematic comparisons across species.

parameters (57%): audience checking (e.g., directing eyes, head, or body at the recipient before signaling), response waiting (e.g., visual attention on the recipient after signaling), goal persistence (e.g., using same signal when the first one did not result in the desired interaction), and elaboration of signaling (e.g., using different signals when the first one did not result in the desired interaction), representing key criteria of flexibility. For instance, Fröhlich et al. (2019b), studying the communicative development of wild chimpanzees at the Taï and Kanyawara communities, found that audience checking and goal persistence but not sensitivity to the recipient's attentional state increased with age. The recipient's attentional state already occurred at 9 months, while goal persistence and audience checking were only fully developed at the age of 70 months. The authors argued that even at a relatively young age, chimpanzees need to be attentive to the visual orientation of their mothers because it is crucial for mother-infant coordination. However, around 15 months of age, when chimpanzee infants start to interact with other group members, audience checking and goal persistence also begin to play a crucial role (Bründl et al., 2021). In contrast, Dafreville et al. (2021), also working with wild chimpanzees, showed that the adjustment in the use of gestures in relation to the mother's attentional state was age-dependent, with only adolescent chimpanzees adjusting their communication appropriately. The authors explain the difference between their results and Fröhlich et al. (2019b) findings because, in their study, maternal visual attention was considered only when the mother had a full view of the infant.

Moreover, two studies on mother-infant interactions in seven pairs of orangutans living in the Sabangau peat-swamp forest, Borneo, Indonesia (*Pongo pygmaeus wurmbii*, Knox et al., 2019) and two pairs of bonobos living in the Leipzig Zoo in Germany (Rossano, 2013) showed that individuals aged three and 2 years, respectively, changed and adjusted their signals during communicative exchanges. Rossano (2013) also showed that the use of response waiting started around the age of 2 years in captive bonobos.

A small number of studies examined the flexibility involved in communicative exchanges with a special focus on the variability of signals used during turn-taking interactions (23%). For instance, Genty (2019) showed that the number and specificity of signals by infant bonobos living in captivity to request different actions from their mothers increased with age. Similarly, Briseño-Jaramillo et al. (2018) found that age positively influenced call rates during turn-taking interactions in one group of wild spider monkeys. In sum, the reported findings, provided by studies that did not directly investigate turn-taking interactions, can be helpful to gain knowledge about the development of distinct cognitive skills thereby helping to formulate hypotheses to be tested in future studies.

Overall, the reviewed studies suggested that key criteria of interactional flexibility such as response waiting, goal persistence, elaboration of signaling, and adjustment to audience effects may be acquired in non-human primates during interactions with mothers and other group members. These may act as models, nurturing and even actively influencing the learning process.

Temporal relationships

Twelve studies investigated the temporal relationships present in communicative interactions (57%) by measuring the time span between the onset of the first signal and the response of the recipient. However, the majority of studies did not measure the time between signal and response. Instead, they delimit the time between the first signal and the answer as a turn-taking event. For example, Lemasson et al. (2013) and Briseño-Jaramillo et al. (2018) considered that a call was emitted as a response if it occurred within 2 and 3 s, respectively. Lemasson et al. (2013), who studied a social group of captive Japanese macaques (*Macaca fuscata*), found that response rates (coo calls given as a response to others and uttered <2 s after the initial call) were less frequently produced by juvenile females aged 8–10 months than by adult females. Moreover, they also found that juvenile females did not adhere to the commonly used temporal relationships in the group (2 s) and produced several consecutive calls repeatedly disrespecting turn-taking principal. Similarly, Briseño-Jaramillo et al. (2018) found that the response rates of wild spider monkeys to calls of conspecifics (calls made within 3 s in response to another call) increased with age.

Furthermore, a study by Takahashi et al. (2016) on three family groups of common marmosets living in captivity showed that individuals avoided overlapping the vocalizations of their parents with increasing age. Moreover, during the 1st month, the infants already engaged in vocal turn-taking interactions with their parents. Similarly, Fröhlich et al. (2016c), who studied bonobos and chimpanzees living in four different communities (LuiKotale at the fringe of Salonga National Park, DRC; Wamba in the Luo Scientific Reserve, DRC, and Kanyawara in Kibale National Park, Uganda; Taï South in Taï National Park, Côte d'Ivoire) in the wild, investigated temporal relationship in interactions by assessing the timing between signals and the respective responses. They divided them into three categories: Immediate responses (<2 s), delayed responses (>2 s), and overlapping responses (<0 s or <1 s). The authors found that overlapping responses were more frequent in interactions between younger infants than between older infants of both species and that chimpanzees, but not bonobos, produced more delayed responses across ages.

Ames et al. (2021) also reported that a narwhal calf of 5 months living in the Scoresby Sound fjord was able to produce a call that either overlapped or occurred within 1 s after the mother's call (78%). The mother only replied to the calf's vocalizations in 16.7% of the time. Since the authors considered a response as "a signal that occurred overlapping or within 1 s of an initiating call," the mother could be replying after this predetermined time. Thus, it was not considered by the authors as a response. However, to verify whether narwhales learn to engage in turn-taking interactions and the species-typical response times, a better sample size and investigations across development will be needed. In addition, Rossano (2013) showed that the temporal relationships of young bonobos (between the age of 1- and 2-years) in captivity when replying to gestures are very similar to those of their mothers. However, this study did not address the development of the temporal relationship element and just compared this element between infants and adults.

Adjacency pair-like sequences

This element was investigated in the majority of studies (81%) by focusing on the presence of "signal-response" pairs (82%). For instance, Briseño-Jaramillo et al. (2018) found that call matching increased across ages in wild spider monkeys. The authors suggested that this turn-taking element may be learned during development. In a similar vein, two studies on two groups of common marmosets in captivity also showed that younger individuals (age 4–6 months) replied to the Phee calls of their parents by using a non-matching call type, a Twitter call. However, at the age of 8 months they started to use the matching call type to engage in turn-taking interactions with their parents (Chow et al., 2015, but see Takahashi et al., 2016 for another explanation of turn-taking development in common marmosets). Moreover, the authors suggested that some turn-taking elements—adjacency pair-like sequences and temporal relationships—in common marmosets seem to be learned during ontogeny and are actively shaped by the parents. A study by Fröhlich et al. (2016c) investigating bonobos and chimpanzees in natural environments found that age influenced specific turn-taking elements, including adjacency pair-like sequences (considered as the number of gesture-response pairs and response waiting). They found that the number of gesture-response pairs (defined by the authors as the number of gestures produced by the signaler and replied by the recipient with a minimum interval of 1 s) decreased with age in bonobos, whereas it increased with age in chimpanzees. The infants of both species were also more likely to wait for a response than their mothers. Moreover, a study performed by Araya-Salas et al. (2020) showed that wild individuals of Spix's disc-winged bats living in a natural environment already produced matched response calls at the age of 4–6 days. This species is known to engage in antiphonal calling by producing an inquiry call that is usually replied with a "response" call during the flight in a roost (Chaverri and Gillam, 2010; Chaverri et al., 2013).

Additionally, the two comprehension studies (Lemasson et al., 2011; Bouchet et al., 2017) that investigated the behavior of Campbell's monkeys (around 2–3 years) in captivity (Rennes University, France), and captive females of Japanese macaques (12–16 months) in captivity (Primate Research Institute, Japan), found that individuals did not pay attention to species-specific turn-taking rules during playback experiments. For instance, juvenile Campbell's monkeys, exposed to two different stimuli (appropriate vocal exchange: A1BA2 and inappropriate vocal exchange: BA1A2), did not show differences in their looking behavior toward the loudspeaker. Similarly, juvenile Japanese macaques were exposed to matching calls (e.g., A_bB_b) and non-matching calls (e.g., A_bB_w) *via* loudspeakers. Their response showed a random distribution regardless of the type of stimulus. The authors attributed these results to the lack of experience of the young individuals in turn-taking interactions, thus, suggesting a possible role of social learning for the production and comprehension of different turn-taking elements. Furthermore, they also argued that participating and being exposed to turn-taking interactions may be a necessary step for fully understanding the turn-taking rules.

Other studies addressed if the signaler was "satisfied" with the response from the recipient (gestures or actions) by investigating intentional gesturing and ontogenetic ritualization or included this parameter as a requirement in the study methods. For instance, Hobaiter and Byrne (2011) and Halina et al. (2013), who studied one community of chimpanzees at the Budongo Forest Reserve (Uganda) and 10 mother-infant bonobo dyads from six zoos, respectively, focused only on signals produced by infants and juveniles that presented a satisfactory outcome to the signaler. However, they did not investigate the changes across development. Similarly, some studies examined the use of intentional signals by chimpanzee infants and juveniles to start interactions—for example, playing and traveling—but did not specifically address the questions of distinct gesture-response pairs (Fröhlich et al., 2016a,b).

Although some of the presented studies did not provide detailed information concerning the development of adjacency pair-like sequences in mammals, they seem to suggest that this ability may be learned in some orders (e.g., primates) and present in others from birth (e.g., bats; Montero and Gillam, 2015; Araya-Salas et al., 2020). However, further research is needed to rule out the possibility of fast learning (Knörnschild, 2014). Moreover, this element may be crucial in species where communicative responses increase survival and reproductive success (e.g., Montero and Gillam, 2015; Araya-Salas et al., 2020).

Participation-frameworks

Nine of the 21 articles addressed the element of participation frameworks in communicative interactions (43%). All these studies used gaze and body orientation as testing parameter (100%). One also measured the distance between the signaler and the receiver. For instance, Rossano (2013) found that 2-year-old captive bonobos established participation frameworks by looking toward their mothers before signaling. Dafreville et al. (2021) and Fröhlich et al. (2019b) included in their studies "eye gaze toward a recipient" as a parameter to be able to consider an exchange of signal-action as an interaction. However, the authors did not measure this element across ages nor did they present analyses of this specific element in their results. Moreover, the studies suggest the presence of this capability at younger ages in distinct chimpanzee communities. Finally, a study conducted by Fröhlich et al. (2016c) with wild bonobos and chimpanzees highlighted that body orientation and initiation distance increased with infant age in both species. These results suggested that similar to the participation-framework element, other turn-taking elements also improve during ontogeny in these two primate species.

In sum, the presented studies suggest that the elements flexibility, and participation-frameworks seem to have a strong learning component, in which full-blown adult-like behavior only appears with increasing age, especially in different primate species. In contrast, other studies suggested that some turn-taking elements, such as temporal relationship and adjacency pair-like sequences, seem to show distinct developmental trajectories according to the non-human species, with some mammal species presenting developed turn-taking elements early in life (e.g., Matsumura, 1979; Rossano, 2013; Araya-Salas et al., 2020). For instance, bats of the species *Thyroptera tricolor* seem to be able to use adjacency pair-like sequences directly after birth. It may be possible that this feature is present since birth, but due to the lack of systematic investigations, the possibility of fast learning cannot be ruled out (Knörnschild, 2014). In contrast, some primate species (such as common marmosets or spider monkeys) appear to learn how to match calls across their development. However, the majority of studies that enabled insights into species' capacities to engage in turn-taking interactions at early ages did not investigate the development of turn-taking longitudinally or during the 1st days of life.

Discussion

The present review aimed to summarize the current knowledge of turn-taking acquisition and development in non-human mammals by carrying out a systematic review of the existing body of research. This approach resulted in a total of 21 studies using experimental and observational methods to investigate the development and acquisition of turn-taking abilities in infant and juvenile individuals of a total of 12 mammal species, mostly primates. Overall, the studies showed considerable variation in methodological approaches and terminologies and were biased toward specific model species (e.g., chimpanzees, common marmosets, and bonobos), and social factors (e.g., mother-infant interactions). As a result, systematic comparisons across species and a detailed understanding of the acquisition and development of turn-taking abilities across mammals is currently not yet possible. In the following paragraphs, we will highlight and discuss the existing gaps and biases in more detail with a special focus on species, developmental milestones, social factors, and turn-taking elements. We will also pinpoint fruitful research avenues to spur more research into this intriguing and new research domain.

Terminologies used

Similar to a recent cross-species review on turn-taking skills by Pika et al. (2018), we found a high degree of heterogeneity concerning the terminologies used. Quite naturally, the term has mainly been used by linguists since the first systematic framework originated in this field (e.g., Sacks et al., 1974; de Ruiter et al., 2006; Stivers et al., 2009). In the twenty-first century, Levinson (2006) stirred considerable interest in turn-taking and involved cognitive processes, particularly in the fields of cognitive science and animal communication (e.g., Logue and Stivers, 2012; Wilkinson et al., 2012; Levinson and Torreira, 2015). Hence, the field of comparative turn-taking is just emerging, and the term may be embraced more in future studies and research. Furthermore, many studies used the term when referring to and investigating temporal relationships only or utilized traditional ethological terms such as "antiphonal calling" and "duetting" (e.g., for recent overviews Pika et al., 2018; Ravignani et al., 2019).

Species and methodological bias

We found a strong research bias toward non-human primates, specifically great ape species (e.g., chimpanzees and bonobos). Furthermore, we found some evidence for turn-taking in infants of other mammalian taxa, especially those capable of vocal learning (e.g., bats and cetaceans). Due to their close phylogenetic proximity to humans (Langergraber et al., 2012; Prüfer et al., 2012), great apes and particularly chimpanzees (Beck, 1982; Gruber and Clay, 2016; Bezanson and McNamara, 2019) have been the focus of a lot of research studies (e.g., Lemasson et al., 2018; Dezecache et al., 2019; Miglietta et al., 2021). For several decades, great apes have been investigated regarding their gestural, vocal, and bimodal communication (e.g., Call and Tomasello, 2007; Genty et al., 2009; Slocombe et al., 2011). Consequently, this attention has also resulted in several studies addressing turn-taking skills in great apes, with a considerable research bias on chimpanzees and bonobos (e.g., Rossano, 2013; Fröhlich et al., 2016c; Genty, 2019) as well as adult individuals (e.g., Luef and Pika, 2017; Levréro et al., 2019; Pougnault et al., 2021b; Rodrigues et al., 2021; Cornec et al., 2022). While studying turn-taking in our closest living relatives is crucial and may aid in developing, in comparison with data from modern humans, more accurate estimates of our extinct ancestors (e.g., Wrangham, 1987; Gruber and Clay, 2016; Muller, 2018), these studies offer only limited insight into abilities derived by convergent evolution (e.g., Emery and Clayton, 2004). A better understanding of the role of turn-taking for sophisticated communication systems and the selective pressures involved can therefore only be gained by studying and comparing turn-taking skills also in and across more distantly related species that live in comparable social settings or show some comparable social aspects (e.g., corvids, cetaceans, New World primates, and Strepsirrhines).

The assessment of research designs employed (e.g., age, the number of individuals, and study design) showed that research interest in this new field increased considerably. However, the studies were unevenly distributed between and within mammal species. In addition, the majority of studies conducted in captivity used experimental designs, while studies carried out in natural settings applied both observational and experimental designs. Moreover, we found a wide variety on the sample size and the age of investigated individuals. Given that infants of different mammal species develop at different rates and show distinct time dependencies and attachments to their mothers (e.g., gorillas: Hoff et al., 1983; common marmosets: Schiel and Huber, 2006; Wang et al., 2014; dolphins: von Streit et al., 2013; bats: Mehdizadeh et al., 2018; chimpanzees: Bründl et al., 2021), future longitudinal studies (infancy to adulthood) could be useful to better understand the linkage between turn-taking skills and developmental milestones thereby avoiding that age becomes a confounding factor. For instance, in marmosets, weaning and locomotory independence starts at the age of ∼1 and 3 months, respectively (Tardif et al., 2003; Schultz-Darken et al., 2016). In contrast, chimpanzees start to walk independently only after the age of 6 months (Goodall, 1986; Bründl et al., 2021) and stop breastfeeding after ∼4 years of age (Samuni et al., 2020). Moreover, offspring of species with prolonged periods to gain independence and extended attachment periods with their mothers are exposed to more learning opportunities and interaction partners, teaching and scaffolding to learn, develop and fine-tune their turn-taking skills. In support of this hypothesis, mother-infant bonding has been shown to be correlated with social communication (e.g., interacting with others in adult life, affective communication) and relationship preferences in different long-living mammal species (e.g., Boccia et al., 1991; Suomi, 2005; Maestripieri, 2018; Verderane et al., 2020). In addition, mothers in these species are quite naturally the first role models for social learning (Whiten and van de Waal, 2018).

Moreover, social and ecological factors have been shown to affect and shape communicative repertoires, usage and, consequently, the exchange of signals (e.g., Fröhlich et al., 2019b, 2021; Pika and Fröhlich, 2019; Roberts and Roberts, 2020). For instance, Fröhlich et al. (2019b) showed that gesture frequency and repertoire size in wild chimpanzees increased with higher interaction rates with non-maternal conspecifics. Thus, future developmental studies could pay special attention to species' biology and control for the influence of social and ecological factors (see also Bräuer et al., 2020).

The influence of social factors

Overall, half of the studies that investigated the role of social factors focused on interactions between mothers and their infants, whereas a smaller proportion also investigated interactions with non-related group members (e.g., Hobaiter and Byrne, 2011; Briseño-Jaramillo et al., 2018). For instance, Hobaiter and Byrne (2011) and Liebal et al. (2004) examined the understanding of intentionally produced signals of infants/juveniles with members of their social group in chimpanzees living in a natural environment and siamangs living in captivity, respectively. Studies focusing on adult individuals have, however, already shown the influence of distinct social factors on turn-taking skills (e.g., Leong et al., 2003; Digweed et al., 2007; Lemasson et al., 2010; Arlet et al., 2015; Levréro et al., 2019; Jenikejew et al., 2020; Pougnault et al., 2021b). For instance, Levréro et al. (2019) showed that social affinity (measured by spatial proximity) influenced the response rate of vocal calls (mainly Peep yelps and Peeps) in captive bonobos. Lemasson et al. (2010) found that captive Campbell's monkeys replied vocally more frequently to older individuals. Moreover, the strength of social bonds seems to be the best predictor of vocal and gestural exchanges in adult individuals of different mammal species (e.g., Fedurek et al., 2013; Roberts and Roberts, 2016; Fröhlich et al., 2017; Toarmino et al., 2017; Kavanagh et al., 2021; Chereskin et al., 2022).

Therefore, it may be possible that some of these factors also shape the communicative development of infants/juveniles. For instance, parents (mother or father) seem to have different influences on the learning processes involved to acquire distinct turn-taking skills. For instance, Chow et al. (2015) showed that in common marmosets living in captivity, parents play essential roles in the development of turn-taking, with juveniles replying differently to the vocalizations of their mothers and fathers compared to their siblings. Moreover, it may be possible that distinct turn-taking elements develop at different developmental rates in cooperative, solitary, or pair-bonding living species due to the number of individuals available to interact with infants/juveniles. Social learning opportunities are provided mainly through mothers in many mammal species (e.g., Bender et al., 2009; van Schaik et al., 2017; Whiten and van de Waal, 2018). However, other group members can also act as role models (e.g., Thornton and Clutton-Brock, 2011; Allen, 2019). Van Boekholt et al. (2021) suggested that a higher number of individuals in the group positively influences learning opportunities in a wide range of behaviors. Similarly, in humans, variability of the interactions (e.g., heterogeneity and numerosity) positively affect learning in different domains, including language (Raviv et al., 2022). Moreover, Fröhlich et al. (2017) demonstrated that in chimpanzees in the wild, interaction rates with other group members crucially influenced communicative exchanges of infants and resulted in a higher number of gestures used in their interactions and hence their gestural repertoires. In the present review, we also found that success and frequency of turn-taking interactions were more common with mothers (Fröhlich et al., 2016c, 2019b). Overall, it seems that different role models (mother, father, or non-related group members) provide crucial but also different learning opportunities to infants. Thus, observing the development of turn-taking with regards to the whole complexity of the respective social group may significantly strengthen our understanding of how and which social factors influence the acquisition and development of turn-taking abilities.

Linkage to developmental milestones

Only relatively little research focused on the acquisition and development of turn-taking and a possible linkage to developmental milestones (Genty, 2019; Araya-Salas et al., 2020; Dafreville et al., 2021). For instance, Genty (2019) and Dafreville et al. (2021) investigated turn-taking abilities of bonobos living in captive settings and chimpanzees living in natural environment and correlated the ages to two developmental milestones, breastfeeding and locomotion. Araya-Salas et al. (2020) investigated the developmental milestone "volant and non-volant" in Spix's disc-winged bats in the wild to assess whether they are able to engage in call-response exchanges. Since the life cycles of mammals can be very different (Western, 1979), the linkage between developmental milestones and turn-taking abilities is essential to enable systematic comparisons across mammal species. Moreover, it is also crucial to better understand how turn-taking abilities and involved elements correlate with the social development of a given species. For instance, some studies showed that several social behaviors in chimpanzees only start later in life and are shaped and scaffolded during ontogeny (e.g., mutual grooming, nut-cracking; Boesch and Boesch, 1983; Goodall, 1986; Matsuzawa, 1994). However, some skills crucial to engage in turn-taking may develop earlier (e.g., goal persistence and audience checking in chimpanzees: Plooij, 1978; Fröhlich et al., 2019b).

In sum, some turn-taking skills and underlying cognitive prerequisites are acquired before mammals engage in frequent social interactions with mothers, possibly shaping and scaffolding these learning processes (e.g., Luef and Pika, 2013; Chow et al., 2015; Musgrave et al., 2016; Whiten and van de Waal, 2018). Thus, linking developmental milestones to turn-taking skills may offer crucial insights into similarities and differences of turn-taking skills and involved elements between mammal species and beyond.

Components and modalities of turn-taking

The investigated components and modalities of communication (production vs. comprehension, signaler vs. recipient, type of signals used) also diverged across studies. Most studies investigated the development of turn-taking production but not comprehension. Although this bias is probably due to studying comprehension being more complicated than production, experimental field studies and current advancements in technology (e.g., observer gaze paradigm and cognitive field experiments using tablets; Hayashi et al., 2020; Lewis and Krupenye, 2022) may enable a methodological balance and a systematic understanding of the cognitive processes needed to understand turn-taking in others.

We also found that not all studies analyzed both the signalers' and recipients' perspectives (e.g., Briseño-Jaramillo et al., 2018; Knox et al., 2019). Although all studies focused on interactions between signalers and recipients, the analyses were biased toward signalers. However, investigating both the behavior of the signaler and the recipient is crucial since some aspects of human conversational turn-taking, such as temporal relationships and communicative repair (Sacks et al., 1974), can only be explored when focusing on both interlocutors (Heesen et al., 2022; Kolff and Pika, 2022).

Furthermore, studies investigating multimodal turn-taking exchanges were relatively limited (e.g., Fröhlich, 2017; Fröhlich et al., 2019b), mirroring a general bias in animal communication research (e.g., Slocombe et al., 2011; Liebal et al., 2012; Prieur et al., 2020; but see Genty et al., 2014). All studies investigated either the vocal or the gestural modality only and the modality changed according to specific model systems studied. For example, gestural interactions were mainly investigated in great apes (e.g., Bard et al., 2014; Knox et al., 2019; Dafreville et al., 2021) whereas vocal exchanges were studied in other primate species and non-primate mammals (e.g., Chow et al., 2015; Araya-Salas et al., 2020). However, as already reported in human children, the use of distinct modalities and multimodal combinations may have different developmental trajectories (e.g., Bates et al., 1975; Holler et al., 2015; Fröhlich et al., 2016c), thereby affecting the first onset and appearance of turn-taking skills. Thus, unimodal and multimodal turn-taking, with the later probably reflecting a higher degree of cognitive flexibility, may also be characterized by different acquisition and developmental times in other mammal species. Hence, using a more holistic approach onto communicative signaling may be important to gain a better understanding of the acquisition and development of turn-taking skills and the importance of turn-taking for language to evolve (Levinson and Holler, 2014; Fröhlich et al., 2019a; Holler and Levinson, 2019). It may also enable better comparisons between mammal species, including humans.

The comparative framework

Recently, Pika et al. (2018) developed a systematic framework to enable systematic comparisons of turn-taking abilities across different species (Pika et al., 2018). Although, they pointed out already that scholars used a wide variety of different terms to describe similar phenomena, we still found considerable variation of terminologies used across recent mammal studies focusing on communicative exchanges. Moreover, the studies were biased toward specific turn-taking elements such as temporal relationships and adjacency pair-like sequences (e.g., Matsumura, 1979; Lemasson et al., 2013; Takahashi et al., 2016; Ames et al., 2021). One explanation for this finding is that temporal relationships can be reliably and consistently measured in both captive and wild individuals and settings and across communicative modalities (e.g., Wong et al., 2004; Kondo et al., 2010; Ames et al., 2021). In addition, measuring the temporal aspects of signals has a long tradition in ethology (Catchpole and Slater, 2008; Pika et al., 2018; Ravignani et al., 2019; de Reus et al., 2021).

The few studies that examined all four elements of the comparative framework involved only older individuals (around 1 year old or more; Rossano, 2013; Fröhlich et al., 2016a,b, 2019b; Knox et al., 2019) and did not always measure each element using the same parameters. For instance, parameters used to assess temporal relationship and flexibility of turns were very broad and were investigated from different approaches. While some studies provided a limited interval time between the offset of the first signal and the onset of the response to consider the exchange as a turn-taking interaction (e.g., Briseño-Jaramillo et al., 2018), others measured the time between the two signals providing the mean of all interval times (e.g., Rossano, 2013). One explanation may be that studies collecting data and measuring and analyzing all involved elements are very time-consuming. In addition, reliable assessments of specific parameters underlying some elements, such as intentionality (Dennett, 1983), may be difficult (Rodrigues and Fröhlich, 2021). However, the quickly developing field of machine learning may offer new solutions to overcome these challenges in the future.

In sum, although the form of turn-taking exchanges of young individuals differed from full-blown turn-taking interactions in adults—similarly to human children—they were characterized by the elements of flexibility, participation frameworks, and temporal relationships. For instance, the element flexibility seems to be positively correlated with age, indicating a possible learning process involved. In addition, studies examining participation frameworks and adjacency pair-like sequences showed that these elements are learned during development in primates (Fröhlich et al., 2016c; Takahashi et al., 2016; Briseño-Jaramillo et al., 2018). However, it is important to note that adjacency pair-like sequences in some non-primate species may already be present at birth (Montero and Gillam, 2015; Araya-Salas et al., 2020). For instance, Araya-Salas et al. (2020) showed that newborn non-volant bats already produced inquiry calls (only produced during flight) when mimicking flying conditions. Further studies could focus on these elements and investigate them across different ages and mammal species. Moreover, the use of signals in appropriate contexts and circumstances may change with regards to the involved costs, benefits, and survival risk (Krebs and Dawkins, 1985; Zeifman, 2001; Laidre and Johnstone, 2013). For example, the survival of young individuals of species where the mothers leave their offspring for considerable time periods to collect food (e.g., seals, bats) relies heavily on correctly replying to their mother's signal. Therefore, this capacity and the intrinsic motivation to reply but also to recognize the mother's call needs to be present early in life in these species.

On the other hand, although the element of temporal relationships is one of the most frequent elements investigated in the existing literature, the results are also the most contradictory in both human and non-human species. For instance, some studies showed that younger animals use response times similar to those of adult individuals (e.g., bonobos: Rossano, 2013; belugas *Delphinapterus leucas*: Vergara et al., 2010), while other studies revealed that young individuals changed their response times with increasing age by decreasing the time of overlap between signals and converging to adult response times (e.g., common marmosets: Takahashi et al., 2016). Similarly, studies investigating temporal relationships in turn-taking interactions of human children also produced mixed results (e.g., Hilbrink et al., 2015; Dominguez et al., 2016; but see Nguyen et al., 2022 for a systematic review in the development of timing in adult–child turn-taking interactions). However, the differences may be due to comparisons between subjects and study groups of different ages (e.g., 1- to 2-year-old children) with older individuals possibly having learned the temporal relationships *via* active and passive shaping in interactions with their caretakers. Thus, further studies may investigate individuals at younger ages to gain a better understanding of the evolution and development of this element and the linkage to the ecology of a given species (see also Bräuer et al., 2020).

Limitations, future directions, and concluding remarks

One of the major limitations of the current review was the lack of available studies investigating interactions of individuals at early ages in different mammal species (e.g., newborns or individuals aged 1–6 months of life), longitudinal studies as well as studies linking turn-taking abilities to developmental milestones (Dafreville et al., 2021). Even when the focus was on newborns, the observational periods were restricted to 4 weeks and 2 months, respectively (e.g., Matsumura, 1979; Takahashi et al., 2016). In contrast, studies that addressed longer developmental time spans did not observe individuals of younger ages (e.g., 10–24 months: Halina et al., 2013; 9–36 months: Fröhlich et al., 2016a) or included a limited number of individuals [e.g., 11 individuals distributed in three age classes (infants-juveniles-adolescents): Dafreville et al., 2021].

However, this research field is still very new, with few published data but with a high potential to help us gain a better understanding of turn-taking, the impact of prosociality on turn-taking evolution and cooperation and the role for language evolution (Yoshida and Okanoya, 2005; Pika et al., 2018). We thus hope to have stirred interest in this new research field to increase future research efforts, and longitudinal studies. Moreover, the review also revealed the challenges of collecting behavioral data, especially when filming interactions with infants in natural settings due to poor visibility, restricted access, and difficulties in following animals through longer time-periods. Thus, one possible solution is to join forces and work collaboratively with other researchers to create large datasets, as has already been done in projects such as ManyPrimates, ManyBirds, ManyBabies, and 1000PAN (Primates et al., 2019; Lambert et al., 2021; Comparative BioCognition, 2022).

Although we found a considerable number of cross-sectional studies, these were biased toward great ape species (e.g., chimpanzees and bonobos). While studying turn-taking abilities in our closest living relatives has key importance, these findings offer only limited insights into the selective pressures favoring the onset and development of cooperative communication (e.g., Vygotsky, 1978; Tomasello, 2008; Pika and Bugnyar, 2011). Carrying out systematic investigations of turn-taking abilities and their development across selected mammal species differing in distinct ecological and social factors will contribute to a more profound knowledge of involved elements and the evolutionary precursors and trajectory of skills constituting "the interaction engine" (Levinson, 2010, 2016). Moreover, we also found a high variation in the elements investigated and a lack of essential measurable variables in each turn-taking element in most of the considered studies, preventing us from drawing reliable conclusions regarding a possible evolutionary trajectory of turn-taking elements.

Nonetheless, the results of the present review suggest that turn-taking abilities and involved elements may have different evolutionary and ontogenetic trajectories depending on the species, social and ecological factors. These findings enable the formation of predictions and hypotheses that can be addressed and tested in future studies to move the field of comparative turn-taking forward (Pika et al., 2018).

For instance, we hypothesize that the onset of cooperative communication is tightly linked to the ecology of a given species and arose to increase reproductive fitness. If this hypothesis is true, mammal species with feeding ecologies that require the mothers to leave their offspring for extended periods and then locate them again respond to contact calls earlier than species where the offspring grows up clinging to the mother's body (e.g., primates) or stays in nests or close proximity (e.g., mice).

We also predict that the onset of the turn-taking elements flexibility of turn-taking organization and adjacency pair-like sequence are correlated with developmental milestones (e.g., timing of weaning, feeding and spatial independence). These elements may be present earlier in species that are characterized by shorter rates of independence, weaning, or that possess shorter periods in close body contact and proximity with their mothers/caretakers.

Moreover, the speed of development of turn-taking elements might be linked to demographic and social factors (e.g., sex, mating system, and parental care strategies). For instance, infants of mammal species that live in large groups or groups that possess cooperative parental care and provide their infants with a higher number of interactions and learning possibilities may show faster learning processes than solitary or pair-bonded species. The possible influence from social and parental care systems on turn-taking abilities has also been supported by Ravignani et al. (2022). Furthermore, since the singing behavior of mammals may also be influenced by the social system and the degree of territoriality (e.g., De Gregorio et al., 2022), future studies into turn-taking skills and acquisition patterns of singing mammals may be crucial to test whether evolutionary new inferential processes ensue when communication becomes governed by more cooperative motives (Vygotsky, 1978; Pika and Bugnyar, 2011).

In conclusion, we highlight five "take-home messages" to nurture the design, implementation, and comparisons of future studies when investigating the acquisition and development of turn-taking abilities: (1) The "turn-taking" terminology should be included in the abstract, title, or keywords of the manuscript; (2) data are needed in systematically selected model systems of mammals differing with regards to social system, parental care and ecology; (3) different social factors, and more extensive developmental periods should be investigated; (4) a more holistic approach to communicative interactions is needed involving different communicative modalities and multi-modal interactions; (5) given the variation across the elements used, the inclusion and usage of specific, measurable variables for non-human animals can be of extreme relevance. These five bullet points will hopefully open up future opportunities for this research field, allowing a better assessment and comparison of the acquisition and development of turn-taking abilities in mammal species. Moreover, based on recent experimental and conceptual studies that investigated the neural circuit mechanisms of vocal turn-taking in different mammal and bird species (e.g., Banerjee and Vallentin, 2022; Ravignani et al., 2022), non-invasive neuroethological approaches may also be very fruitful to move the field forward.

We thus hope that the present review served to highlight the gaps and trends in the study of the acquisition and development of turn-taking in mammal species and pinpointed the challenges and difficulties of this research field.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

SP had the idea for the present paper and designed it with FA. FA collected and analyzed the data and wrote the first draft. FA and SP edited and finalized the manuscript. Both authors contributed to the article and approved the submitted version.

Funding

This paper is part of a project that was funded by an EU-Consolidator grant (772000, TurnTaking) to SP of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.

Acknowledgments

We are grateful to S. Cosper for his help editing the manuscript and constructive discussions. We also thank the entire Comparative BioCognition (CBC) team for all support and helpful discussions and the reviewers for very constructive comments to increase the quality of the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.987253/full#supplementary-material

References

Allen, J. A. (2019). Community through culture: from insects to whales. *BioEssays* 41, 1900060. doi: 10.1002/bies.201900060

Ames, A. E., Blackwell, S. B., Tervo, O. M., and Heide-Jørgensen, M. P. (2021). Evidence of stereotyped contact call use in narwhal (*Monodon monoceros*) mother-calf communication. *PLoS ONE* 16, 1–19. doi: 10.1371/journal.pone.0254393

Araya-Salas, M., Hernández-Pinsón, H. A., Rojas, N., and Chaverri, G. (2020). Ontogeny of an interactive call-and-response system in Spix's disc-winged bats. *Anim. Behav.* 166, 233–245. doi: 10.1016/j.anbehav.2020.05.018

Arbib, M. A., Liebal, K., and Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* 49, 1053–1076. doi: 10.1086/593015

Arlet, M., Jubin, R., Masataka, N., and Lemasson, A. (2015). Grooming-at-a-distance by exchanging calls in non-human primates. *Biol. Lett.* 11, 711. doi: 10.1098/rsbl.2015.0711

Armstrong, D. F., and Sherman, E. W. (2007). *The Gestural Origin of Language*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195163483.001.0001

Aychet, J., Blois-heulin, C., and Lemasson, A. (2021). Sequential and network analyses to describe multiple signal use in captive mangabeys. *Anim. Behav.* 9, 5. doi: 10.1016/j.anbehav.2021.09.005

Bakeman, R., and Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Dev.* 55, 1278. doi: 10.2307/1129997

Banerjee, A., and Vallentin, D. (2022). Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Curr. Opin. Neurobiol.* 73, 102529. doi: 10.1016/j.conb.2022.102529

Bard, K. A., Dunbar, S., Maguire-Herring, V., Veira, Y., Hayes, K. G., and Mcdonald, K. (2014). Gestures and social-emotional communicative development in chimpanzee infants. *Am. J. Primatol.* 76, 14–29. doi: 10.1002/ajp.22189

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill. Palmer. Q.* 21, 205–226.

Bateson, M. C. (1975). Mother-infant exchanges: the epigenesis of conversational interaction. *Ann. N. Y. Acad. Sci.* 263, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575.x

Beck, B. B. (1982). Chimpocentrism: bias in cognitive ethology. *J. Hum. Evol.* 11, 3–17. doi: 10.1016/S0047-2484(82)80027-4

Bender, C. E., Herzing, D. L., and Bjorklund, D. F. (2009). Evidence of teaching in atlantic spotted dolphins (*Stenella frontalis*) by mother dolphins foraging in the presence of their calves. *Anim. Cogn.* 12, 43–53. doi: 10.1007/s10071-008-0169-9

Bezanson, M., and McNamara, A. (2019). The what and where of primate field research may be failing primate conservation. *Evol. Anthropol.* 28, 166–178. doi: 10.1002/evan.21790

Bezerra, B. M., and Souto, A. (2008). Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int. J. Primatol.* 29, 671–701. doi: 10.1007/s10764-008-9250-0

Boccia, M. L., Reite, M., and Laudenslager, M. L. (1991). Early social environment may alter the development of attachment and social support: two case reports. *Infant Behav. Dev.* 14, 253–260. doi: 10.1016/0163-6383(91)90009-H

Boesch, C., and Boesch, H. (1983). Optimisation of nut-cracking with natural hammers by wild chimpanzees. *Behaviour* 83, 265–286. doi: 10.1163/156853983X00192

Bouchet, H., Koda, H., and Lemasson, A. (2017). Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Anim. Behav.* 129, 81–92. doi: 10.1016/j.anbehav.2017.05.012

Bräuer, J., Hanus, D., Pika, S., Gray, R., and Uomini, N. (2020). Old and new approaches to animal cognition: there is not "one cognition." *J. Intell.* 8, 1–25. doi: 10.3390/jintelligence8030028

Briseño-Jaramillo, M., Berthet, M., Estrada, A., Biquand, V., and Lemasson, A. (2021). Socially mediated overlap in vocal interactions between free-ranging black howler monkeys. *Am. J. Primatol.* 83, 1–14. doi: 10.1002/ajp.23297

Briseño-Jaramillo, M., Ramos-Fernández, G., Palacios-Romo, T. M., Sosa-López, J. R., and Lemasson, A. (2018). Age and social affinity effects on contact call interactions in free-ranging spider monkeys. *Behav. Ecol. Sociobiol.* 72, 2. doi: 10.1007/s00265-018-2615-2

Brown, P. E., Brown, T. W., and Grinnell, A. D. (1983). Echolocation, development, and vocal communication in the lesser bulldog bat, *Noctilio albiventris*. *Behav. Ecol. Sociobiol.* 13, 287–298. doi: 10.1007/BF00299676

Bründl, A. C., Tkaczynski, P. J., Nohon Kohou, G., Boesch, C., Wittig, R. M., and Crockford, C. (2021). Systematic mapping of developmental milestones in wild chimpanzees. *Dev. Sci.* 24, 1–13. doi: 10.1111/desc.12988

Call, J., and Tomasello, M. (2007). *The Gestural Communication of Apes and Monkeys*. Oxfordshire: Taylor & Francis Group/Lawrence Erlbaum Associates.

Caro, T., and Hauser, M. (1992). Is there teaching in nonhuman animals? *Q. Rev. Biol.* 67, 151–174. doi: 10.1086/417553

Carter, G. G., Fenton, M. B., and Faure, P. A. (2009). White-winged vampire bats (*Diaemus youngi*) exchange contact calls. *Can. J. Zool.* 87, 604–608. doi: 10.1139/Z09-051

Cartmill, E. A., and Byrne, R. W. (2007). Orangutans modify their gestural signaling according to their audience's comprehension. *Curr. Biol.* 17, 1345–1348. doi: 10.1016/j.cub.2007.06.069

Casillas, M., Bobb, S. C., and Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *J. Child Lang.* 43, 1310–1337. doi: 10.1017/S0305000915000689

Catchpole, C. K., and Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations, 2nd Edn.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511754791

Chaverri, G., and Gillam, E. H. (2010). Cooperative signaling behavior of roost location in a leaf-roosting bat. *Commun. Integr. Biol.* 3, 599–601. doi: 10.4161/cib.3.6.13277

Chaverri, G., Gillam, E. H., and Kunz, T. H. (2013). A call-and-response system facilitates group cohesion among disc-winged bats. *Behav. Ecol.* 24, 481–487. doi: 10.1093/beheco/ars188

Cheney, D. L., and Seyfarth, R. M. (2010). "Primate communication and human language: continuities and discontinuities," in *Mind the Gap*, eds P. Kappeler, J. Silk (Berlin; Heidelberg, Springer), 13. doi: 10.1007/978-3-642-02725-3_13

Chereskin, E., Connor, R. C., Friedman, W. R., Sørensen, P. M., Kru, M., and King, S. L. (2022). Allied male dolphins use vocal exchanges to "bond at a distance". *Curr. Biol.* 1657–1663. doi: 10.1016/j.cub.2022.02.019

Chow, C. P., Mitchell, J. F., and Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B.* 282, 69. doi: 10.1098/rspb.2015.0069

Christiansen, M. H., and Kirby, S. (2003a). Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 300–307. doi: 10.1016/S1364-6613(03)00136-0

Christiansen, M. H., and Kirby, S. (2003b). *Language Evolution*. New York, NY: Oxford University Press.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539

Clutton-Brock, T. (2009). Structure and function in mammalian societies. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3229–3242. doi: 10.1098/rstb.2009.0120

Comparative BioCognition (2022). *100pan*. Available online at: https://www.comparative-biocognition.de/1000pan/about-1000pan (accessed June 18, 2022).

Corballis, M. (2011). The evolution of language. *Soc. Commun.* 96, 409–434.

Corballis, M. C. (2009). The evolution of language. *Ann. N. Y. Acad. Sci.* 1156, 19–43. doi: 10.1111/j.1749-6632.2009.04423.x

Cornec, C., Ngofuna, M., Lemasson, A., Monghiemo, C., Narat, V., and Levréro, F. (2022). A pilot study of calling patterns and vocal turn-taking in wild bonobos *Pan paniscus*. *Ethol. Ecol. Evol.* 2022, 1–18. doi: 10.1080/03949370.2022.2044387

Crockford, C., Herbinger, I., Vigilant, L., and Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* 110, 221–243. doi: 10.1111/j.1439-0310.2004.00968.x

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., and Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state: a developmental study on mother–infant dyads in wild chimpanzees (*Pan troglodytes schweinfurthii*). *Am. J. Primatol.* 2021, ajp.23339. doi: 10.1002/ajp.23339

Darwin, C. (1859). On the origin of species. John Murray, London.

Davila-Ross, M., Jesus, G., Osborne, J., and Bard, K. A. (2015). Chimpanzees (*Pan troglodytes*) produce the same types of "laugh faces" when they emit laughter and when they are silent. *PLoS ONE* 10, 1–11. doi: 10.1371/journal.pone.0127337

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., et al. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Curr. Zool.* 65, 89–97. doi: 10.1093/cz/zoy058

de Reus, K., Soma, M., Anichini, M., Gamba, M., De Heer Kloots, M., Lense, M., et al. (2021). Rhythm in dyadic interactions. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 337. doi: 10.1098/rstb.2020.0337

de Ruiter, J. P., Mitterer, H., and Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82, 515–535. doi: 10.1353/lan.2006.0130

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., and Manser, M. (2018). Vocal turn-taking in meerkat group calling sessions. *Curr. Biol.* 28, 3661–3666.e3. doi: 10.1016/j.cub.2018.09.065

Dennett, D. C. (1983). Intentional systems in cognitive ethology: the "Panglossian paradigm" defended. *Behav. Brain Sci.* 6, 343–355. doi: 10.1017/S0140525X00016393

Dezecache, G., Crockford, C., and Zuberbühler, K. (2019). The development of communication in alarm contexts in wild chimpanzees. *Behav. Ecol. Sociobiol.* 73, 6. doi: 10.1007/s00265-019-2716-6

Digweed, S. M., Fedigan, L. M., and Rendall, D. (2007). Who cares who calls? Selective responses to the lost calls of socially dominant group members in the white-faced capuchin (*Cebus capucinus*). *Am. J. Primatol.* 69, 829–835. doi: 10.1002/ajp.20398

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal principles in the repair of communication problems. *PLoS ONE* 10, 1–15. doi: 10.1371/journal.pone.0136100

Dominguez, S., Devouche, E., Apter, G., and Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant Child Dev.* 25, 240–255. doi: 10.1002/icd.1976

Ehret, G., and Bernecker, C. (1986). Low-frequency sound communication by mouse pups (*Mus musculus*): wriggling calls release maternal behaviour. *Anim. Behav.* 34, 821–830. doi: 10.1016/S0003-3472(86)80067-7

Elowson, A. M., Snowdon, C. T., and Lazaro-Perea, C. (1998). Infant "babbling" in a nonhuman primate: complex vocal sequences with repeated call types. *Behaviour* 135, 643–664. doi: 10.1163/156853998792897905

Emery, N. J., and Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306, 1903–1907. doi: 10.1126/science.1098410

Fedurek, P., Machanda, Z. P., Schel, A. M., and Slocombe, K. E. (2013). Pant hoot chorusing and social bonds in male chimpanzees. *Anim. Behav.* 86, 189–196. doi: 10.1016/j.anbehav.2013.05.010

Ferreira, M. J., Levis, C., Chaves, L., Clement, C. R., and Soldati, G. T. (2022). Indigenous and traditional management creates and maintains the diversity of ecosystems of south american tropical savannas. *Front. Environ. Sci.* 10, 1–18. doi: 10.3389/fenvs.2022.809404

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychon. Bull. Rev.* 24, 72–78. doi: 10.3758/s13423-016-1076-8

Fischer, J., Cheney, D. L., and Seyfarth, R. M. (2000). Development of infant baboons' responses to graded bark variants. *Proc. R. Soc. B Biol. Sci.* 267, 2317–2321. doi: 10.1098/rspb.2000.1285

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779

Fröhlich, M. (2017). Taking turns across channels: conversation-analytic tools in animal communication. *Neurosci. Biobehav. Rev.* 80, 201–209. doi: 10.1016/j.neubiorev.2017.05.005

Fröhlich, M., Bartolotta, N., Fryns, C., Wagner, C., Momon, L., Jaffrezic, M., et al. (2021). Orangutans have larger gestural repertoires in captivity than in the wild - a case of weak innovation? *iScience* 24, 103304. doi: 10.1016/j.isci.2021.103304

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., et al. (2016c). Unpeeling the layers of language: bonobos and chimpanzees engage in cooperative turn-taking sequences. *Sci. Rep.* 6, srep25887. doi: 10.1038/srep25887

Fröhlich, M., Müller, G., Zeiträg, C., Wittig, R. M., and Pika, S. (2017). Gestural development of chimpanzees in the wild: the impact of interactional experience. *Anim. Behav.* 134, 271–282. doi: 10.1016/j.anbehav.2016.12.018

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., van Schaik, C. P., Fröhlich, M., et al. (2019a). Multimodal communication and language origins: integrating gestures and vocalizations. *Biol. Rev.* 94, 1809–1829. doi: 10.1111/brv.12535

Fröhlich, M., Wittig, R. M., and Pika, S. (2016a). Play-solicitation gestures in chimpanzees in the wild: flexible adjustment to social circumstances and individual matrices. *R. Soc. Open Sci.* 3, 160278. doi: 10.1098/rsos.160278

Fröhlich, M., Wittig, R. M., and Pika, S. (2016b). Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. *Anim. Cogn.* 19, 483–500. doi: 10.1007/s10071-015-0948-z

Fröhlich, M., Wittig, R. M., and Pika, S. (2019b). The ontogeny of intentional communication in chimpanzees in the wild. *Dev. Sci.* 22, 1–14. doi: 10.1111/desc.12716

Gardner, R. A., and Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science* 165, 664–672. doi: 10.1126/science.165.3894.664

Genty, E. (2019). Vocal–gestural combinations in infant bonobos: new insights into signal functional specificity. *Anim. Cogn.* 22, 505–518. doi: 10.1007/s10071-019-01267-0

Genty, E., Breuer, T., Hobaiter, C., and Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. *Anim. Cogn.* 12, 527–546. doi: 10.1007/s10071-009-0213-4

Genty, E., Clay, Z., Hobaiter, C., and Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PLoS ONE* 9, e84738. doi: 10.1371/journal.pone.0084738

Gibson, D. R. (2003). Participation shifts: order and differentiation in group conversation. *Soc. Forces* 81, 1335–1381. doi: 10.1353/sof.2003.0055

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge: Cambridge University Press.

Goodwin, C., and Heritage, J. (1990). Conversation analysis. *Annu. Rev. Anthropol.* 19, 283–307. doi: 10.1146/annurev.an.19.100190.001435

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., and Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol.* 6, 1–10. doi: 10.3389/fpsyg.2015.01167

Gruber, T., and Clay, Z. (2016). A comparison between bonobos and chimpanzees: a review and update. *Evol. Anthropol.* 25, 239–252. doi: 10.1002/evan.21501

Halina, M., Rossano, F., and Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. *Anim. Cogn.* 16, 653–666. doi: 10.1007/s10071-013-0601-7

Hammerschmidt, K., and Fischer, J. (2008). Constraints in primate vocal production. *Evol. Commun. Flex.* 5, 92–119. doi: 10.7551/mitpress/9780262151214.003.0005

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2010). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., et al. (2014). The mystery of language evolution. *Front. Psychol.* 5, 1–12. doi: 10.3389/fpsyg.2014.00401

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., et al. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions *via* medial prefrontal cortex. *Cell Rep.* 30, 4433–4444.e5. doi: 10.1016/j.celrep.2020.03.013

Hayes, K. J., and Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proc. Am. Philos. Soc.* 95, 105–109. doi: 10.1037/e533882004-001

Hebets, E. A., and Papaj, D. R. (2005). Complex signal function: developing a framework of testable hypotheses. *Behav. Ecol. Sociobiol.* 57, 197–214. doi: 10.1007/s00265-004-0865-7

Heesen, R., Fröhlich, M., Sievers, C., Woensdregt, M., and Dingemanse, M. (2022). Coordinating social action: a primer for the cross- species investigation of communicative repair. *Philos. Trans. R. Soc B.* 2021, 110. doi: 10.1098/rstb.2021.0110

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hewes, G. W., Andrew, R. J., Carini, L., Choe, H., Gardner, R. A., Kortlandt, A., et al. (1973). Primate communication and the gestural origin of language [and Comments and Reply]. *Curr. Anthropol.* 14, 5–24. doi: 10.1086/201401

Hilbrink, E. E., Gattis, M., and Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* 6, 1–12. doi: 10.3389/fpsyg.2015.01492

Hobaiter, C., and Byrne, R. W. (2011). Serial gesturing by wild chimpanzees: its nature and function for communication. *Anim. Cogn.* 14, 827–838. doi: 10.1007/s10071-011-0416-3

Hobaiter, C., and Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Curr. Biol.* 24, 1596–1600. doi: 10.1016/j.cub.2014.05.066

Hobaiter, C., Byrne, R. W., and Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behav. Ecol. Sociobiol.* 71, 1. doi: 10.1007/s00265-017-2325-1

Hobaiter, C., Graham, K. E., and Byrne, R. W. (2022). Are ape gestures like words? Outstanding issues in detecting similarities and differences between human language and ape gesture. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 301. doi: 10.1098/rstb.2021.0301

Hockett, C. D., and Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 88–96. doi: 10.1038/scientificamerican0960-88

Hockett, C. F. (1960). "Logical considerations in the study of animal communication," in *Animal Sounds and Communication*, eds W. E. Lanyon and W. N. Tavolga (Washington, DC, American Institute of Biological Sciences), 392–430.

Hockett, C. F. (1963). "The problem of universals in language," in *Universals of Language*, ed J. H. Green-berg (Cambrigde, MIT Press), 1–22.

Hoff, M. P., Nadler, R. D., and Maple, T. L. (1983). Maternal transport and infant motor development in a captive group of lowland gorillas. *Primates* 24, 77–85. doi: 10.1007/BF02381455

Holler, J., Kendrick, K. H., Casillas, M., and Levinson, S. C. (2015). Editorial: turn-taking in human communicative interaction. *Front. Psychol.* 6, 1–4. doi: 10.3389/fpsyg.2015.01919

Holler, J., and Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends Cogn. Sci.* 23, 639–652. doi: 10.1016/j.tics.2019.05.006

Hradec, M., Bolechová, P., and Svobodová, I. (2016). Production of a female-specific great call in an immature male gibbon, the *Nomascus* genus. *Primates* 57, 445–448. doi: 10.1007/s10329-016-0569-4

Hudson, R., and Trillmich, F. (2008). Sibling competition and cooperation in mammals: challenges, developments and prospects. *Behav. Ecol. Sociobiol.* 62, 299–307. doi: 10.1007/s00265-007-0417-z

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., and Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev.* 66, 1–132. doi: 10.1111/1540-5834.00137

Janik, V. M., and Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. B Biol. Sci.* 376, 244. doi: 10.1098/rstb.2020.0244

Jenikejew, J., Chaignon, B., Linn, S., and Scheumann, M. (2020). Proximity-based vocal networks reveal social relationships in the Southern white rhinoceros. *Sci. Rep.* 10, 72052. doi: 10.1038/s41598-020-72052-0

Kappeler, P. M., and Van Schaik, C. P. (2005). *Cooperation in Primates and Humans: Mechanisms and Evolution*. Berlin; Heidelberg: Springer. doi: 10.1007/3-540-28277-7

Katsu, N., Yamada, K., Okanoya, K., and Nakamichi, M. (2018). Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Curr. Zool.* 65, 99–105. doi: 10.1093/cz/zoy077

Kavanagh, E., Street, S. E., Angwela, F. O., Bergman, T. J., Blaszczyk, M. B., Bolt, L. M., et al. (2021). Dominance style is a key predictor of vocal use and evolution across nonhuman primates. *R. Soc. Open Sci.* 8, 210873. doi: 10.1098/rsos.210873

Kellogg, W. N., and Kellogg, L. A. (1933). *The Ape and the Child*. New York, NY: McGraw Hill. doi: 10.1097/00006324-193311000-00009

Kerth, G. (2008). Causes and consequences of sociality in bats. *Bioscience* 58, 737–746. doi: 10.1641/B580810

Killin, A. (2017). Where did language come from? Connecting sign, song, and speech in hominin evolution. *Biol. Philos.* 32, 759–778. doi: 10.1007/s10539-017-9607-x

Knight, C., Studdert-Kennedy, M., and Hurford, J. (2000). *The Evolutionary Emergence of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511606441

Knörnschild, M. (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* 28, 80–85. doi: 10.1016/j.conb.2014.06.014

Knörnschild, M., and von Helversen, O. (2008). Nonmutual vocal mother-pup recognition in the greater sac-winged bat. *Anim. Behav.* 76, 1001–1009. doi: 10.1016/j.anbehav.2008.05.018

Knox, A., Markx, J., How, E., Azis, A., Hobaiter, C., van Veen, F. J. F., et al. (2019). Gesture use in communication between mothers and offspring in wild orangutans (*Pongo pygmaeus wurmbii*) from the Sabangau peat-swamp forest, Borneo. *Int. J. Primatol.* 40, 393–416. doi: 10.1007/s10764-019-00095-w

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., and Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS ONE* 8, 71432. doi: 10.1371/journal.pone.0071432

Koda, H., Oyakawa, C., Kato, A., Shimizu, D., Rizaldi, Koyama, Y., et al. (2014). Immature male gibbons produce female-specific songs. *Primates* 55, 13–17. doi: 10.1007/s10329-013-0390-2

Kolff, K., and Pika, S. (2022). "The evolutionary roots of human communicative repair: the case of chimpanzee grooming," *The Evolution of Language: Proceedings of the Joint Conference on Language Evolution (JCoLE), Japan*, eds A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, et al. (Kanazawa).

Kölliker, M., Royle, N., and Smiseth, P. T. (2012). *The Evolution of Parental Care*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199692576.001.0001

Kondo, N., Watanabe, S., and Izawa, E. I. (2010). A temporal rule in vocal exchange among Large-billed Crows *Corvus macrorhynchos* in Japan. *Ornithol. Sci.* 9, 83–91. doi: 10.2326/osj.9.83

Krebs, J. R., and Dawkins, R. (1985). "Animal signals: mind-reading and manipulation," in *Behavioral Ecology: An Evolutionary Approach*, eds J. R. Krebs and N. B. Davies (Oxford; London; Blackwell Scientific Publications), 381–403.

Ladygina-Kohts, N. N. (1935). *Infant Chimpanzee and Human Child. A Classic Comparative Study of Ape Emotions and Intelligence*. New York, NY: Oxford University Press.

Laidre, M. E., and Johnstone, R. A. (2013). Animal signals. *Curr. Biol.* 23, 829–833. doi: 10.1016/j.cub.2013.07.070

Lambert, M. L., Farrar, B. G., Garcia-Pelegrin, E., Reber, S., and Miller, R. (2021). ManyBirds: a multi-site collaborative open science approach to avian cognition and behaviour research. *Anim. Behav. Cogn.* 9, 133–152. doi: 10.26451/abc.09.01.11.2022

Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., et al. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. U. S. A.* 109, 15716–15721. doi: 10.1073/pnas.1211740109

Leavens, D. A., and Hopkins, W. D. (1998). Intentional communication by chimpanzees: a cross-sectional study of the use of referential gestures. *Dev. Psychol.* 34, 813–822. doi: 10.1037/0012-1649.34.5.813

Leighty, K. A., Soltis, J., Leong, K., and Savage, A. (2008). Antiphonal exchanges in African elephants (*Loxodonta africana*): collective response to a shared stimulus, social facilitation, or true communicative event? *Behaviour* 145, 297–312. doi: 10.1163/156853908783402885

Lemasson, A., Gandon, E., and Hausberger, M. (2010). Attention to elders' voice in non-human primates. *Biol. Lett.* 6, 325–328. doi: 10.1098/rsbl.2009.0875

Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., et al. (2011). Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Sci. Rep.* 1, srep00022. doi: 10.1038/srep00022

Lemasson, A., Guilloux, M., Rizaldi, Barbu, S., Lacroix, A., and Koda, H. (2013). Age- and sex-dependent contact call usage in Japanese macaques. *Primates* 54, 283–291. doi: 10.1007/s10329-013-0347-5

Lemasson, A., Pereira, H., and Levréro, F. (2018). Social basis of vocal interactions in western lowland gorillas (*Gorilla g. gorilla*). *J. Comp. Psychol.* 132, 141–151. doi: 10.1037/com0000105

Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D., Savage, A., Leong, K. M., et al. (2003). Quantifying acoustic and temporal characteristics of vocalizations for a group of captive african elephants *Loxodonta africana*. *Bioacoustics* 13, 213–231. doi: 10.1080/09524622.2003.9753499

Levinson, S. C. (2006). On the human "interaction engine". *Roots Hum. Soc. Cult. Cogn. Interact.* 3, 39–69. doi: 10.4324/9781003135517-3

Levinson, S. C. (2010). Interactional foundations of language: the interaction engine hypothesis. *Hum. Lang. From Genes Brain Behav.* 18, 189–200. doi: 10.7551/mitpress/10841.003.0018

Levinson, S. C. (2013). Recursion in pragmatics. *Language* 89, 149–162. doi: 10.1353/lan.2013.0005

Levinson, S. C. (2016). Turn-taking in human communication - origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Levinson, S. C., and Holler, J. (2014). The origin of human multi-modal communication. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130302. doi: 10.1098/rstb.2013.0302

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6, 731. doi: 10.3389/fpsyg.2015.00731

Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J.-P., and Lemasson, A. (2019). Social bonding drives vocal exchanges in bonobos. *Sci. Rep.* 9, 9. doi: 10.1038/s41598-018-36024-9

Lewis, L. S., and Krupenye, C. (2022). Eye-tracking as a window into primate social cognition. *Am. J. Primatol.* 2022, 1–18. doi: 10.1002/ajp.23393

Liebal, K., Pika, S., and Tomasello, M. (2004). Social communication in siamangs (*Symphalangus syndactylus*): use of gestures and facial expressions. *Primates* 45, 41–57. doi: 10.1007/s10329-003-0063-7

Liebal, K., Waller, B. M., Burrows, A. M., and Slocombe, K. E. (2012). *Primate Communication: A Multimodal Approach.* Cambridge: Cambridge University Press.

Liebal, K., Waller, B. M., Burrows, A. M., and Slocombe, K. E. (2013). *Primate Communication: A Multimodal Approach.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139018111

Logue, D. M., and Stivers, T. (2012). Squawk in interaction: a primer of conversation analysis for students of animal communication. *Behaviour* 149, 1283–1298. doi: 10.1163/1568539X-00003031

Luef, E. M., Breuer, T., and Pika, S. (2016). Food-associated calling in gorillas (Gorilla g. gorilla) in the wild. *PLoS One.* 11. doi: 10.1371/journal.pone.0144197

Luef, E. M., and Pika, S. (2013). Gorilla mothers also matter! New insights on social transmission in gorillas (*Gorilla gorilla gorilla*) in captivity. *PLoS ONE* 8, 79600. doi: 10.1371/journal.pone.0079600

Luef, E. M., and Pika, S. (2017). Reciprocal greeting in chimpanzees (*Pan troglodytes*) at the Ngogo community. *J. Neurolinguistics* 43, 263–273. doi: 10.1016/j.jneuroling.2016.11.002

Maestripieri, D. (2018). Maternal influences on primate social development. *Behav. Ecol. Sociobiol.* 72, 2547. doi: 10.1007/s00265-018-2547-x

Marchant-Forde, J. N., Marchant-Forde, R. M., and Weary, D. M. (2002). Responses of dairy cows and calves to each other's vocalisations after early separation. *Appl. Anim. Behav. Sci.* 78, 19–28. doi: 10.1016/S0168-1591(02)00082-5

Matsumura, S. (1979). Mother-infant communication in a horseshoe bat (*Rhinolophus ferrumequinum nippon*): development of vocalization. *J. Mammal.* 60, 76–84. doi: 10.2307/1379760

Matsuzawa, T. (1994). "Field experiments on use of stone tools by chimpanzees in the wild," in *Chimpanzee Cultures*, eds R. W. Wrangham, W. C. McGrew, F. de Wall, and P. G. Heltne (Cambridge, MA: Harvard University Press), 351–370.

McNeill, D. (2010). *How Language Began: Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.

Mehdizadeh, R., Eghbali, H., and Sharifi, M. (2018). Postnatal growth and vocalization development in the long-fingered bat, *Myotis capaccinii* (Chiroptera, vespertilionidae). *Zool. Stud.* 57, 37. doi: 10.6620/ZS.2018.57-37

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting - a mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi*)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Merker, B., and Cox, C. (1999). Development of the female great call in *Hylobates gabriellae*: a case study. *Folia Primatol.* 70, 97–106. doi: 10.1159/000021680

Miglietta, S., Bardino, G., Sotto-Mayor, A., Galli, A. S., Meulman, E., Breuer, T., et al. (2021). Absence of specific individuals and high food abundance elicit food calls in wild western gorillas. *Behav. Ecol. Sociobiol.* 75, 3027. doi: 10.1007/s00265-021-03027-y

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca. Anim. Behav.* 67, 1099–1107. doi: 10.1016/j.anbehav.2003.06.017

Mishima, Y., Morisaka, T., Mishima, Y., Sunada, T., and Miyamoto, Y. (2018). Redefinition and sexual difference of contact calls in belugas (*Delphinapterus leucas*). *Aquat. Mamm.* 44, 538–554. doi: 10.1578/AM.44.5.2018.538

Mocha, Y., Ben Mundry, R., and Pika, S. (2019). Joint attention skills in wild Arabian babblers (*Turdoides squamiceps*): a consequence of cooperative breeding? *Proc. R. Soc. B Biol. Sci.* 286, 147. doi: 10.1098/rspb.2019.0147

Montero, K., and Gillam, E. H. (2015). Geographic variation in contact calls emitted by a leaf-roosting bat suggest distinct modes of vocal transmission. *J. Acoust. Soc. Am.* 138, 1932–1932. doi: 10.1121/1.4934089

Morisaka, T., Yoshida, Y., Akune, Y., Mishima, H., and Nishimoto, S. (2013). Exchange of "signature" calls in captive belugas (*Delphinapterus leucas*). *J. Ethol.* 31, 141–149. doi: 10.1007/s10164-013-0358-0

Muller, M. N. (2018). "Introduction: chimpanzees and human evolution," in *Chimpanzees and Human Evolution*, eds N. M. Muller, R. W. Wrangham and D. R. Pilbeam (Cambridge, MA; London: Harvard University Press), 3–21. doi: 10.4159/9780674982642

Mundy, P., and Newell, L. (2007). Attention, joint attention & social cognition. *Soc. Cogn.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Musgrave, S., Morgan, D., Lonsdorf, E., Mundry, R., and Sanz, C. (2016). Tool transfers are a form of teaching among chimpanzees. *Sci. Rep.* 6, 1–7. doi: 10.1038/srep34783

Nguyen, V., Versyp, O., Cox, C., and Fusaroli, R. (2022). A systematic review and Bayesian meta- - analysis of the development of turn taking in adult - child vocal interactions. *Child Dev.* 2022, 1–20. doi: 10.1111/cdev.13754

Nomikou, I., Leonardi, G., Radkowska, A., Raczaszek-Leonardi, J., and Rohlfing, K. J. (2017). Taking up an active role: emerging participation in early mother–infant interaction during peekaboo routines. *Front. Psychol.* 8, 1–19. doi: 10.3389/fpsyg.2017.01656

O'Connell-Rodwell, C. E., Wood, J. D., Wyman, M., Redfield, S., Puria, S., and Hart, L. A. (2012). Antiphonal vocal bouts associated with departures in free-ranging African elephant family groups (*Loxodonta africana*). *Bioacoustics* 21, 215–224. doi: 10.1080/09524622.2012.686166

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., et al. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev*. 96, 1695–1722. doi: 10.1111/brv.12721

Okobi, D. E., Banerjee, A., Matheson, A. M. M., Phelps, S. M., and Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988. doi: 10.1126/science.aau9480

Partan, S., and Marler, P. (1999). Communication goes multimodal. *Science* 1272–1273. doi: 10.1126/science.283.5406.1272

Patterson, F. (1978). Conversations with a gorilla. *National Geographic* 134, 438–465.

Pika, S. (2015). Gestural communication in nonhuman species. *Emerg. Trends Soc. Behav. Sci.* 2015, 1–11. doi: 10.1002/9781118900772.etrds0147

Pika, S., and Bugnyar, T. (2011). The use of referential gestures in ravens (*Corvus corax*) in the wild. *Nat. Commun.* 2, 1567. doi: 10.1038/ncomms1567

Pika, S., and Deschner, T. (2019). A new window onto animal culture: the case of chimpanzee gesturing. *Gesture* 18, 239–260. doi: 10.1075/gest.19012.pik

Pika, S., and Fröhlich, M. (2019). Gestural acquisition in great apes: the Social Negotiation Hypothesis. *Anim. Cogn.* 22, 551–565. doi: 10.1007/s10071-017-1159-6

Pika, S., Liebal, K., Call, J., and Tomasello, M. (2005). The gestural communication of apes and human infants. *Gesture* 5, 41–56. doi: 10.1075/gest.5.1.05pik

Pika, S., Liebal, K., and Tomasello, M. (2003). Gestural communication in young gorillas (*Gorilla gorilla*): gestural repertoire, learning, and use. *Am. J. Primatol*. 60, 95–111. doi: 10.1002/ajp.10097

Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B Biol. Sci*. 285, 598. doi: 10.1098/rspb.2018.0598

Plooij, F. (1978). "Some basic traits of language in wild chimpanzees?" in *Action, Gesture and Symbol: The Emergence of Language*, ed A. Lock (London: Academic Press), 111–131.

Pougnault, L., Lemasson, A., Mulot, B., and Levréro, F. (2021a). Temporal calling patterns of a captive group of chimpanzees (*Pan troglodytes*). *Int. J. Primatol.* doi: 10.1007/s10764-021-00262-y

Pougnault, L., Levréro, F., Leroux, M., Paulet, J., Bombani, P., Dentressangle, F., et al. (2021b). Social pressure drives "conversational rules" in great apes. *Biol. Rev*. doi: 10.1111/brv.12821

Pougnault, L., Levréro, F., Mulot, B., and Lemasson, A. (2020). Breaking conversational rules matters to captive gorillas: a playback experiment. *Sci. Rep*. 10, 7. doi: 10.1038/s41598-020-63923-7

Prieur, J., Barbu, S., Blois-Heulin, C., and Lemasson, A. (2020). The origins of gestures and language: history, current advances and proposed theories. *Biol. Rev*. 95, 531–554. doi: 10.1111/brv.12576

Primates, M., Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., et al. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLoS ONE* 14, e0223675. doi: 10.1371/journal.pone.0223675

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., et al. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486, 527–531. doi: 10.1038/nature11128

Ravignani, A., Lumaca, M., and Kotz, S. A. (2022). Interhemispheric brain communication and the evolution of turn-taking in mammals. *Front. Ecol. Evol.* 10, 1–7. doi: 10.3389/fevo.2022.916956

Ravignani, A., Verga, L., and Greenfield, M. D. (2019). Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Ann. N. Y. Acad. Sci*. 1453, 12–21. doi: 10.1111/nyas.14230

Raviv, L., Lupyan, G., and Green, S. C. (2022). How variability shapes learning and generalization. *Trends Cogn. Sci*. 26, 462–483. doi: 10.1016/j.tics.2022.03.007

Roberts, A. I., and Roberts, S. G. B. (2016). Wild chimpanzees modify modality of gestures according to the strength of social bonds and personal network size. *Sci. Rep*. 6, 1–13. doi: 10.1038/srep33864

Roberts, S. G. B., and Roberts, A. I. (2020). Social and ecological complexity is associated with gestural repertoire size of wild chimpanzees. *Integr. Zool.* 15, 276–292. doi: 10.1111/1749-4877.12423

Rodrigues, E. D., and Fröhlich, M. (2021). Operationalizing intentionality in primate communication: social and ecological considerations. *Int. J. Primatol*. 21, 248. doi: 10.1007/s10764-021-00248-w

Rodrigues, E. D., Santos, A. J., Hayashi, M., Matsuzawa, T., and Hobaiter, C. (2021). Exploring greetings and leave-takings: communication during arrivals and departures by chimpanzees of the Bossou community, Guinea. *Primates* 21, 957. doi: 10.1007/s10329-021-00957-z

Rosenbaum, S., and Gettler, L. T. (2018). With a little help from her friends (and family) part II: non-maternal caregiving behavior and physiology in mammals. *Physiol. Behav*. 193, 12–24. doi: 10.1016/j.physbeh.2017.12.027

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interact. Stud*. 14, 160–189. doi: 10.1075/is.14.2.02ros

Rossano, F. (2018). Social manipulation, turn-taking and cooperation in apes: implications for the evolution of language-based interaction in humans. *Interact. Stud*. 19, 151–166. doi: 10.1075/is.17043.ros

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn- taking in conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Samuni, L., Tkaczynski, P., Deschner, T., Löhrrich, T., Wittig, R. M., and Crockford, C. (2020). Maternal effects on offspring growth indicate post-weaning juvenile dependence in chimpanzees (*Pan troglodytes verus*). *Front. Zool.* 17, 1–12. doi: 10.1186/s12983-019-0343-8

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis I*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208

Schel, A. M., Bono, A., Aychet, J., Pika, S., and Lemasson, A. (2022). Intentional gestural communication amongst red-capped mangabeys (*Cercocebus torquatus*). *Anim. Cogn*. 2022, 7. doi: 10.1007/s10071-022-01615-7

Schel, A. M., Machanda, Z., Townsend, S. W., Zuberbühler, K., and Slocombe, K. E. (2013). Chimpanzee food calls are directed at specific individuals. *Anim. Behav*. 86, 955–965. doi: 10.1016/j.anbehav.2013.08.013

Schiel, N., and Huber, L. (2006). Social influences on the development of foraging behavior in free-living common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 68, 1150–1160. doi: 10.1002/ajp.20284

Schultz-Darken, N., Braun, K. M., and Emborg, M. E. (2016). Neurobehavioral development of common marmoset monkeys. *Dev. Psychobiol*. 58, 141–158. doi: 10.1002/dev.21360

Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Anim. Behav*. 76, 1977–1988. doi: 10.1016/j.anbehav.2008.07.032

Sèbe, F., Duboscq, J., Aubin, T., Ligout, S., and Poindron, P. (2010). Early vocal recognition of mother by lambs: contribution of low- and high-frequency vocalizations. *Anim. Behav*. 79, 1055–1066. doi: 10.1016/j.anbehav.2010.01.021

Seyfarth, R. M., and Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain Lang*. 115, 92–100. doi: 10.1016/j.bandl.2009.10.003

Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., and Hammerschmidt, K. (2010a). The central importance of information in studies of animal communication. *Anim. Behav*. 80, 3–8. doi: 10.1016/j.anbehav.2010.04.012

Seyfarth, R. M., Cheney, D. L., and Marler, P. (2010b). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav*. 28, 1070–1094. doi: 10.1016/S0003-3472(80)80097-2

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J*. 27, 623–656. doi: 10.1002/j.1538-7305.1948.tb00917.x

Slocombe, K. E., Waller, B. M., and Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Anim. Behav*. 81, 919–924. doi: 10.1016/j.anbehav.2011.02.002

Snow, C. E. (1977). "Mothers' speech research: from input to interaction," in *Talking to Children*, eds C. E. Snow and C. A. Ferguson (Cambridge: Cambridge University Press), 31–49.

Snowdon, C. T. (2017). Learning from monkey "talk." *Science* 355, 1120–1122. doi: 10.1126/science.aam7443

Springer, M. S., Stanhope, M. J., Madsen, O., and Jong, W. W. (2004). Molecules consolidate the placental mammal tree. *Trends Ecol. Evol*. 19, 430–438. doi: 10.1016/j.tree.2004.05.006

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Suomi, S. J. (2005). Mother-infant attachment, peer relationships, and the development of social networks in rhesus monkeys. *Hum. Dev.* 48, 67–79. doi: 10.1159/000083216

Takahashi, D. Y., Fenley, A. R., and Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Trans. R. Soc. B Biol. Sci.* 2015, 371. doi: 10.1098/rstb.2015.0370

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., et al. (2015). The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734–738. doi: 10.1126/science.aab1058

Takahashi, D. Y., Narayanan, D. Z., and Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168. doi: 10.1016/j.cub.2013.09.005

Tardif, S. D., Smucny, D. A., Abbott, D. H., Mansfield, K., Schultz-Darken, N., and Yamamoto, M. E. (2003). Reproduction in captive common marmosets (*Callithrix jacchus*). *Comp. Med.* 53, 364–368.

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018). An analysis of white-handed gibbon male song reveals speech-like phrases. *Am. J. Phys. Anthropol.* 166, 649–660. doi: 10.1002/ajpa.23451

Thornton, A., and Clutton-Brock, T. (2011). Social learning and the development of individual and group behaviour in mammal societies. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 978–987. doi: 10.1098/rstb.2010.0312

Toarmino, C. R., Wong, L., and Miller, C. T. (2017). Audience affects decision-making in a marmoset communication network. *Biol. Lett.* 13, 20160934. doi: 10.1098/rsbl.2016.0934

Tomasello, M. (2008). *Origins of Human Communications.* Cambridge, Massachusetts: MIT Press.

Tomasello, M., and Zuberbühler, K. (2002). "Primate vocal and gestural communication," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, eds M. Bekoff, C. S. Allen and G. Burghardt (Cambridge: MIT Press), 293–299.

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., et al. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biol. Rev.* 92, 1427–1433. doi: 10.1111/brv.12289

Van Boekholt, B., van de Waal, E., and Sterck, E. H. M. (2021). Organized to learn: the influence of social structure on social learning opportunities in a group. *iScience* 24, 102117. doi: 10.1016/j.isci.2021.102117

van Schaik, C., Graber, S., Schuppli, C., and Burkart, J. (2017). The ecology of social learning in animals and its link with intelligence. *Span. J. Psychol.* 19, 1–12. doi: 10.1017/sjp.2016.100

Vanderhoff, E. N., and Bernal Hoverud, N. (2022). Perspectives on antiphonal calling, duetting and counter-singing in non-primate mammals: an overview with notes on the coordinated vocalizations of bamboo rats (*Dactylomys spp.,* Rodentia: Echimyidae). *Front. Ecol. Evol.* 10, 1–10. doi: 10.3389/fevo.2022.906546

Verderane, M. P., Aguiar, R. M., and Izar, P. (2020). Face–to–face interactions between mothers and female infants in wild bearded capuchin monkeys (*Sapajus libidinosus*). *Dev. Psychobiol.* 62, 941–949. doi: 10.1002/dev.21948

Vergara, V., Michaud, R., and Barrett-Lennard, L. G. (2010). What can captive whales tell us about their wild counterparts? Identification, usage, and ontogeny of contact calls in belugas (*Delphinapterus leucas*). *Int. J. Comp. Psychol.* 23, 278–309. doi: 10.46867/IJCP.2010.23.03.08

Vernes, S. C. (2017). What bats have to say about speech and language. *Psychon. Bull. Rev.* 24, 111–117. doi: 10.3758/s13423-016-1060-3

von Streit, C., Ganslosser, U., and von Fersen, L. (2013). Behavioral development of two captive mother-calf dyads of bottlenose dolphins (*Tursiops truncatus*) in the calves' first year. *Int. J. Comp. Psychol.* 26, 5. doi: 10.46867/ijcp.2013.26.03.05

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wang, Y., Fang, Q., and Gong, N. (2014). Motor assessment of developing common marmosets. *Neurosci. Bull.* 30, 387–393. doi: 10.1007/s12264-013-1395-y

Western, D. (1979). Size, life history and ecology in mammals. *Afr. J. Ecol.* 17, 185–204. doi: 10.1111/j.1365-2028.1979.tb00256.x

Whiten, A., and van de Waal, E. (2018). The pervasive role of social learning in primate lifetime development. *Behav. Ecol. Sociobiol.* 72, 80. doi: 10.1007/s00265-018-2489-3

Wilkinson, R., Leudar, I., and Pika, S. (2012). "Requesting behaviours within episodes of active sharing. A new look on chimpanzee signalling," in *Developments in Primate Gesture Research*, eds S. Pika and K. Liebal (Amsterdam: John Benjamins Publishing Company), 199–221. doi: 10.1075/gs.6.12wil

Wong, B. B. M., Cowling, A. N. N., Cunningham, R. B., Donnelly, C. F., and Cooper, P. D. (2004). Do temperature and social environment interact to affect call rate in frogs (*Crinia signifera*)? *Aust. Ecol.* 29, 209–214. doi: 10.1111/j.1442-9993.2004.01338.x

Wrangham, R. W. (1987). "The significance of African apes for reconstructing human social evolution," in *Primate Models of Hominid Evolution*, ed W. G. Kinzey (Albany: SUNY Press), 51–71.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Lang. Commun.* 1, 47–67. doi: 10.1016/S0271-5309(97)00033-5

Yamaguchi, C., Izumi, A., and Nakamura, K. (2009). Temporal rules in vocal exchanges of phees and trills in common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 71, 617–622. doi: 10.1002/ajp.20697

Yamamoto, M. E. (2005). Infant care in callitrichids: cooperation and competition. *Annu. Rev. Biomed. Sci.* 7, 149–160. doi: 10.5016/48

Yoshida, S., and Okanoya, K. (2005). Evolution of turn-taking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165. doi: 10.11225/jcss.12.153

Zeifman, D. M. (2001). An ethological analysis of human infant crying: answering Tinbergen's four questions. *Dev. Psychobiol.* 39, 265–285. doi: 10.1002/dev.1005

Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fevo.2023.1071640

A workflow for the automated detection and classification of female gibbon calls from long-term acoustic recordings

Dena J. Clink[1]*, Isabel Kier[1], Abdul Hamid Ahmad[2] and Holger Klinck[1]

[1]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States, [2]Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

OPEN ACCESS

EDITED BY
Marco Gamba,
University of Turin,
Italy

REVIEWED BY
Lydia Light,
University of North Carolina at Charlotte,
United States
Tim Sainburg,
University of California,
San Diego,
United States

*CORRESPONDENCE
Dena J. Clink
✉ dena.clink@cornell.edu

SPECIALTY SECTION
This article was submitted to Behavioral and Evolutionary Ecology, a section of the journal Frontiers in Ecology and Evolution

RECEIVED 16 October 2022
ACCEPTED 09 January 2023
PUBLISHED 09 February 2023

CITATION
Clink DJ, Kier I, Ahmad AH and Klinck H (2023) A workflow for the automated detection and classification of female gibbon calls from long-term acoustic recordings.
Front. Ecol. Evol. 11:1071640.
doi: 10.3389/fevo.2023.1071640

Passive acoustic monitoring (PAM) allows for the study of vocal animals on temporal and spatial scales difficult to achieve using only human observers. Recent improvements in recording technology, data storage, and battery capacity have led to increased use of PAM. One of the main obstacles in implementing wide-scale PAM programs is the lack of open-source programs that efficiently process terabytes of sound recordings and do not require large amounts of training data. Here we describe a workflow for detecting, classifying, and visualizing female Northern grey gibbon calls in Sabah, Malaysia. Our approach detects sound events using band-limited energy summation and does binary classification of these events (gibbon female or not) using machine learning algorithms (support vector machine and random forest). We then applied an unsupervised approach (affinity propagation clustering) to see if we could further differentiate between true and false positives or the number of gibbon females in our dataset. We used this workflow to address three questions: (1) does this automated approach provide reliable estimates of temporal patterns of gibbon calling activity; (2) can unsupervised approaches be applied as a post-processing step to improve the performance of the system; and (3) can unsupervised approaches be used to estimate how many female individuals (or clusters) there are in our study area? We found that performance plateaued with >160 clips of training data for each of our two classes. Using optimized settings, our automated approach achieved a satisfactory performance (F1 score~80%). The unsupervised approach did not effectively differentiate between true and false positives or return clusters that appear to correspond to the number of females in our study area. Our results indicate that more work needs to be done before unsupervised approaches can be reliably used to estimate the number of individual animals occupying an area from PAM data. Future work applying these methods across sites and different gibbon species and comparisons to deep learning approaches will be crucial for future gibbon conservation initiatives across Southeast Asia.

KEYWORDS

machine learning, *Hylobates*, R programing language, signal processing, bioacoustics, Southeast Asia

Introduction

Passive acoustic monitoring

Researchers worldwide are increasingly interested in passive acoustic monitoring (PAM), which relies on autonomous recording units to monitor vocal animals and their habitats. Increased availability of low-cost recording units (Hill et al., 2018; Sethi et al., 2018; Sugai et al., 2019), along with advances in data storage capabilities, makes the use of PAM an attractive option for monitoring vocal species in inaccessible areas where the animals are difficult to monitor visually (such as dense rainforests) or when the animals exhibit cryptic behavior (Deichmann et al., 2018). Even in cases where other methods such as visual surveys are feasible, PAM may be superior as it may be able to detect animals continuously for extended periods of time, at a greater range than visual methods, can operate under any light conditions, and is more amenable to automated data collection than visual or trapping techniques (Marques et al., 2013). In addition, PAM provides an objective, non-invasive method that limits observer bias in detection of target signals.

One of the most widely recognized benefits of using acoustic monitoring, apart from the potential to reduce the amount of time needed for human observers, is that there is a permanent record of the monitored soundscape (Zwart et al., 2014; Sugai and Llusia, 2019). In addition, the use of archived acoustic data allows for multiple analysts at different times to review and validate detections/classifications, as opposed to point-counts where one or multiple observers, often with varying degrees of experience, collect the data *in-situ*. It is, therefore, not surprising that, in many cases, analysis of recordings taken by autonomous recorders can be more effective than using trained human observers in the field. For example, a comparison of PAM and human observers to detect European nightjars (*Caprimulgus europaeus*) showed that PAM detected nightjars during 19 of 22 survey periods, while surveyors detected nightjars on only six of these occasions (Zwart et al., 2014). An analysis of 21 bird studies that compared detections by human observers and detections from acoustic data collected using autonomous recorders found that for 15 of the studies, manual analysis of PAM acoustic data led to results that were equal to or better than results from point counts done using human observers (Shonfield and Bayne, 2017). Despite the rapidly expanding advances in PAM technology, the use of PAM is limited by a lack of widely applicable analytical methods and the limited availability of open-source audio processing tools, particularly for the tropics, where soundscapes are very complex (Gibb et al., 2018).

Interest in the use of PAM to monitor nonhuman primates has increased in recent years, with one of the foundational papers using PAM to estimate occupancy of three signal types: chimpanzee buttress drumming (*Pan troglodytes*) and the loud calls of the Diana monkey (*Cercopithecus diana*) and king colobus monkey (*Colobus polykomos*) in Taï National Park, Côte d'Ivoire (Kalan et al., 2015). The authors found that occurrence data from PAM combined with automated processing methods was comparable to that collected by human observers. Since then, PAM has been used to investigate chimpanzee group ranging and territory use (Kalan et al., 2016), vocal calling patterns of gibbons (*Hylobates funereus;* Clink et al., 2020b) and howler monkeys (*Alouatta caraya;* Pérez-Granados and Schuchmann, 2021), occupancy modeling of gibbons (*Nomascus gabriellae;* Vu and Tran, 2019) and density estimation of pale fork-marked lemurs (*Phaner pallescens*) based on calling bout rates (Markolf et al., 2022).

Acoustic analysis of long-term datasets

Traditional approaches for finding signals of interest include hand-browsing spectrograms to identify signals of interest using programs such as Raven Pro (K. Lisa Yang Center for Conservation Bioacoustics, Ithaca, NY, USA). This approach can reduce processing time relative to listening to the recordings but requires trained analysts and substantial human investment. Another approach is hand-browsing of long-term spectral averages (LTSAs), which still requires a significant time investment, but allows analysts to process data at a faster rate than hand-browsing of regular spectrograms, as LTSAs provide a visual representation of the soundscape over a larger time period [days to weeks to years (Wiggins, 2003; Clink et al., 2020b)]. However, particularly with the advances in data storage capabilities and deployment of arrays of recorders collecting data continuously, the amount of time necessary for hand-browsing or listening to recordings for signals of interest is prohibitive and is not consistent with conservation goals that require rapid assessment. This necessitates reliable, automated approaches to efficiently process large amounts of acoustic data.

Automated detection and classification

Machine listening, a fast-growing field in computer science, is a form of artificial intelligence that "learns" from training data to perform particular tasks, such as detecting and classifying acoustic signals (Wäldchen and Mäder, 2018). Artificial neural networks (Mielke and Zuberbühler, 2013), Gaussian mixture models (Heinicke et al., 2015), and Support Vector Machines (Heinicke et al., 2015; Keen et al., 2017) – some of the more commonly used algorithms for early applications of human speech recognition (Muda et al., 2010; Dahake and Shaw, 2016) – can be used for the automated detection of terrestrial animal signals from long-term recordings. Many different automated detection approaches for terrestrial animals using these early machine-learning models have been developed (Kalan et al., 2015; Zeppelzauer et al., 2015; Keen et al., 2017). Given the diversity of signal types and acoustic environments, no single detection algorithm performs well across all signal types and recording environments.

A summary of existing automated detection/classification approaches

Python and R are the two most popular open-source programming languages for scientific research (Scavetta and Angelov, 2021). Although Python has surpassed R in overall popularity, R remains an important and complementary language, especially in the life sciences (Lawlor et al., 2022). An analysis of 30 ecology journals indicated that in 2017 over 58% of ecological studies utilized the R programming environment (Lai et al., 2019). Although we could not find a more recent assessment, we are certain that R remains an important tool for ecologists and conservationists. Therefore, automated detection/classification workflows in R may be more accessible to ecologists already familiar with the R programming environment. Already, many existing R packages can be used for importing, visualizing, and manipulating sound files. For example, "seewave" (Sueur et al., 2008) and "tuneR" (Ligges et al., 2016) are some of the more commonly used packages for reading in sound files, visualizing spectrograms and extracting features.

An early workflow and R package "flightcallr" used random forest classification to classify bird calls, but the detection of candidate signals using band-limited energy summation was done using an external program, Raven Pro (Ross and Allen, 2014). One of the first R packages that provided a complete automated detection/classification of acoustic signals workflow in R was "monitoR," which provides functions for detection using spectrogram cross-correlation and bin template matching (Katz et al., 2016b). In spectrogram cross-correlation, the detection and classification steps are combined. The R package "warbleR" has functions for visualization and detection of acoustic signals using band-limited energy summation, all done in R (Araya-Salas and Smith-Vidaurre, 2017).

There has been an increase in the use of deep learning—a subfield of machine listening that utilizes neural network architecture—for the combined automated detection/classification of acoustic signals. Target species include North Atlantic right whales (*Eubalaena glacialis,* Shiu et al., 2020), fin whales (*Balaenoptera physalus*, Madhusudhana et al., 2021), North American and European bird species (Kahl et al., 2021), multiple forest birds and mammals in the Pacific Northwest (Ruff et al., 2021), chimpanzees (*Pan troglodytes*, Anders et al., 2021), high frequency and ultrasonic mouse lemur (*Microcebus murinus*) calls (Romero-Mujalli et al., 2021) and Hainan gibbon (*Nomascus hainanus*) vocalizations (Dufourq et al., 2021). See Table 1 for a summary of existing approaches that use R or Python for the automated detection of acoustic signals from terrestrial PAM data. Note that the only applications for gibbons are on a single species, the Hainan gibbon.

Recently, a workflow was developed that provided a graphical interface through a Shiny application and RStudio for the automated detection of acoustic signals, with the automated detection and classification done using a deep convolutional neural network (CNN) implemented in Python (Ruff et al., 2021). Another R package utilizes deep learning for the automated detection of bat echolocation calls; this package also relies on CNNs implemented in Python (Silva et al., 2022). Deep learning approaches are promising, but they often require large amounts of training data, which can be challenging to obtain, particularly for rare animals or signals (Anders et al., 2021). In addition, training deep learning models may require extensive computational power and specialized hardware (Dufourq et al., 2022); effective training of deep learning models also generally requires a high level of domain knowledge (Hodnett et al., 2019).

Feature extraction

An often necessary step for classification of acoustic signals (unless using deep learning or spectrogram cross-correlation) is feature extraction, wherein the digital waveform is reduced to a meaningful number of informative acoustic features. Traditional approaches relied on manual feature extraction from the spectrogram, but this method requires substantial effort from human observers, which means it is not optimal for automated approaches. Early automated approaches utilized feature sets such as Mel-frequency cepstral coefficients; MFCCs (Heinicke et al., 2015), a feature extraction method originally designed for human speech applications (Han et al., 2006; Muda et al., 2010). Despite their relative simplicity, MFCCs can be used to effectively distinguish between female Northern grey gibbon individuals (Clink et al., 2018a), terrestrial and underwater soundscapes (Dias et al., 2021), urban soundscapes (Noviyanti et al., 2019), and even the presence or absence of queen bees in a bee hive (Soares et al., 2022). Although the use of MFCCs as features for distinguishing between individuals in other gibbon species has been limited, the many documented cases of vocal individuality across gibbon species (Haimoff and Gittins, 1985; Haimoff and Tilson, 1985; Sun et al., 2011; Wanelik et al., 2012; Feng et al., 2014) indicate that MFCCs will most likely be effective features for discriminating individuals of other gibbon species. There are numerous other options for feature extraction, including automated generation of spectro-temporal features for sound events (Sueur et al., 2008; Ross and Allen, 2014) and calculating a set of acoustic indices (Huancapaza Hilasaca et al., 2021).

Other approaches rely on spectrogram images and treat sound classification as an image classification problem (Lucio et al., 2015; Wäldchen and Mäder, 2018; Zottesso et al., 2018). For many of the current deep learning approaches, the input for the classification is the spectrogram, which can be on the linear or Mel-frequency scale (reviewed in Stowell, 2022). An approach that has gained traction in recent years is the use of embeddings, wherein a pre-trained convolutional neural network (CNN), for example, using 'Google's AudioSet' dataset (Gemmeke et al., 2017), is used to create a set of informative, representative features. A common way to do this is to remove the final classification layer from the pre-trained network, which leaves a high-dimensional feature representation of the acoustic data (Stowell, 2022). This approach has been used successfully in numerous ecoacoustic applications (Sethi et al., 2020, 2022; Heath et al., 2021).

Training, validation, and test datasets

When doing automated detection of animal calls, the number and diversity of training data samples must be taken into consideration to minimize false positives (where the system falsely classifies the signal as the signal of interest) and false negatives (e.g., missed opportunities), where the system fails to detect the signal of interest. To avoid overfitting — a phenomenon that occurs when model performance is not generalizable to data that was not included in the training dataset — it is essential to separate data into training, validation, and test sets (Heinicke et al., 2015; Mellinger et al., 2016). The training dataset is the sample of data that was used to fit the model, the validation set is used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters, and the test dataset is the sample of data used to provide an unbiased evaluation of a final model fit. Some commonly used metrics include precision (the proportion of detections that are true detections) and recall (the proportion of actual calls that are successfully detected; Mellinger et al., 2016). Often, these metrics are converted to false alarm rates, such as the rate of false positives per hour, which can help guide decisions about the detection threshold. In addition, when doing automated detection and classification, it is common to use a threshold (such as the probability assigned to a classification by a machine learning algorithm) to make decisions about rejecting or accepting a detection (Mellinger et al., 2016). Varying these thresholds will result in changes to false-positive and the proportion of missed calls. These can be plotted with receiver operating curves (ROC; Swets, 1964) or detection error tradeoff curves (DET; Martin et al., 1997).

PAM of gibbons

Gibbons are pair-living, territorial small apes that regularly emit species- and sex-specific long-distance vocalizations that can be heard

TABLE 1 Summary of existing approaches that use R or Python for the automated detection/classification of acoustic signals from terrestrial PAM data.

Signal type	Training data recording location	Detection/ classification approach	R?	Python?	Open source?	Citation	Repository?
Nocturnal flight calls of multiple avian species	Six locations in New York State, USA	BLED detector in external program + RF	Y	N	Y	Ross and Allen (2014)	Package on R forge (Ross, 2013)
Four primate species	Taï National Park, Côte d'Ivoire	Speaker segmentation + SVM or Gaussian Mixture Models	Y	N	N	Heinicke et al. (2015)	Code availability not indicated in publication
Two northeastern songbird species	10 sites in Vermont and New York, USA	Binary point matching or spectrogram cross-correlation	Y	N	Y	Katz et al. (2016a,b)	Package on CRAN (Hafner and Katz, 2018)
Forest elephants	Three sites in Gabon and one in the Central African Republic	CNNs	N	Y	N	Bjorck et al. (2019)	Code availability not indicated in publication
Two frog species	Temperate N. America and Panama	Measure the presence of periodic structure based on the power spectral density	Y	Y	Y	Lapp et al. (2021)	Python and R implementations on GitHub
No signals specified	~	Binary point matching or spectrogram cross-correlation + SVM, RF, others	Y	N	Y	Balantic and Donovan (2020)	Package on Gitlab
Chimpanzees	Taï National Park, Côte d'Ivoire	Convolutional recurrent neural networks	N	Y	Y	Anders et al. (2021)	Package on GitHub
984 bird species	North America and Europe	Deep artificial neural networks	N	Y	Y	Kahl et al. (2021)	Source code on GitHub
12 bird species and 2 small mammal species	Forested landscapes of Oregon and Washington, USA	CNNs	Y	Y	Y	Ruff et al. (2021)	Code and data on Zenodo (Ruff et al., 2020)
Hainan gibbons	Hainan, China	CNNs	N	Y	Y	Dufourq et al. (2021)	Code available on GitHub; training data on Zenodo (Dufourq et al., 2020)
Bat echolocation calls and two owl species	Europe	CNNs	Y	Y	Y	Silva et al. (2022)	Package on CRAN (Silva, 2022)
Hainan gibbons, black-and-white ruffed lemurs and two bird species	Hainan, China; Ranomafana National Park, Madagascar; Mount Mulanje Biosphere Reserve, Malawi and Intaka Island Nature Reserve in Cape Town, South Africa	Pretrained CNNs (e.g., transfer learning)	N	Y	Y	Dufourq et al. (2022)	Code available on GitHub
60 species of katydids	Barro Colorado Island, Panama	CNNs	N	Y	Y	Madhusudhana et al. (2019)	Code available on Zenodo (Madhusudhana, 2021)

Repositories are linked if they have an associated digital object identifier (DOI) or are available *via* package development web sites such as the Comprehensive R Archive Network (CRAN). Otherwise, availability as indicated in associated publications is shown.

>1 km in a dense forest (Mitani, 1984, 1985; Geissmann, 2002; Clarke et al., 2006). All but one of the approximately 20 gibbon species are classified as Endangered or Critically Endangered, making them an important target for conservation efforts (IUCN, 2022). Gibbons are often difficult to observe visually in the forest canopy but relatively easy to detect acoustically (Mitani, 1985), which makes them ideal candidates for PAM. Indeed, many early studies relied on human observers listening to calling gibbons to estimate group density using fixed-point counts (Brockelman and Srikosamatara, 1993; Hamard et al., 2010; Phoonjampa et al., 2011; Kidney et al., 2016). To date, relatively few gibbon species have been monitored using PAM, including the Hainan gibbon in China (Dufourq et al., 2021), yellow-cheeked gibbons in Vietnam (Vu and Tran, 2019, 2020), and Northern grey gibbons (*Hylobates funereus*) on Malaysian Borneo (Clink et al., 2020b). However, this will undoubtedly

change over the next few years with increased interest and accessibility of equipment and analytical tools needed for effective PAM of gibbon species across Southeast Asia.

Most gibbon species have two types of long-distance vocalizations. Male solo is the term used for male vocalizations emitted while vocalizing alone, and duets are the coordinated vocal exchange between the adult male and female of the pair (Cowlishaw, 1992, 1996). Gibbons generally call in the early morning, with male gibbon solos starting earlier than the duets (Clink et al., 2020b). In the current paper, we focused our analysis on a call type in the female contribution to the duet, known as the great call, for two reasons. First, the structure of the great call is highly stereotyped, individually distinct (Terleph et al., 2015; Clink et al., 2017), of longer duration than other types of gibbon vocalizations, and the males tend to be silent during the female great call, which facilitates better automated detection. Second, most acoustic density estimation techniques focus on duets, as females rarely sing if they are not in a mated pair (Mitani, 1984). In contrast, males will solo whether in a mated pair or drifters (Brockelman and Srikosamatara, 1993), which means automated detection of the female call will be more relevant for density estimation (Kidney et al., 2016) using PAM. Northern grey gibbon females have been shown to emit individually distinct calls (Clink et al., 2017, 2018a), and these calls can be discriminated well using both supervised and unsupervised methods (Clink and Klinck, 2021).

Individual vocal signatures and PAM

A major hurdle in the implementation of many PAM applications is the fact that individual identity is unknown, as data are collected in the absence of a human observer. In particular, density estimation using PAM data would greatly benefit from the ability to infer the number of individuals in the survey area from acoustic data (Stevenson et al., 2015). The location of the calling animal can infer individual identity. Still, precise acoustic localization that relies on the time difference of arrival (TDOA) of a signal at multiple autonomous recording units can be logistically and analytically challenging (Wijers et al., 2021). Another way that individual identity can be inferred from acoustic data is through individually distinct vocal signatures. Individual vocal signatures have been identified across a diverse range of taxonomic groups (Darden et al., 2003; Gillam and Chaverri, 2012; Kershenbaum et al., 2013; Favaro et al., 2016). Most studies investigating individual signatures use supervised methods, wherein the identity of the calling individual is known, but see Sainburg et al. (2020) for unsupervised applications on individual vocal signatures. Identifying the number of individuals based on acoustic differences from PAM data remains a challenge, as unsupervised approaches must be used since the data are, by definition, collected in the absence of human observers (Clink and Klinck, 2021; Sadhukhan et al., 2021).

Overview of the automated detection/classification workflow

This workflow complements existing R packages for acoustic analysis, such as tuneR (Ligges et al., 2016), seewave (Sueur et al., 2008), warbleR (Araya-Salas and Smith-Vidaurre, 2017), and monitoR (Katz et al., 2016b), and contributes functionalities for automated detection and classification using support vector machine, SVM (Meyer et al., 2017) and random forest, RF (Liaw and Wiener, 2002) algorithms. Automated detection of signals in this workflow follows nine main steps: (1) Create labeled training, validation, and test datasets; (2) identify potential sound events using a band-limited energy detector; (3) data reduction and feature extraction of sound events using Mel-frequency cepstral coefficients; MFCCs (Han et al., 2006; Muda et al., 2010); (4) train machine learning algorithms on the training dataset (5) classify the sound events in the validation dataset using trained machine learning algorithms and calculate performance metrics on the validation dataset to find optimal settings; (6) use a manually labeled test dataset to evaluate model performance; (7) run the detector/classifier over the entire dataset (once the optimal settings have been identified); (8) verify all detections and remove false positives; and (9) use the validated output from the detector/classifier for inference (Figure 1).

When training the system, it is important to use data that will not be used in the subsequent testing phase, as this may artificially inflate accuracy estimates (Heinicke et al., 2015). Creating labeled datasets and subsequent validation of detections to remove false positives requires substantial input and investment by trained analysts; this is the case for all automated detection approaches, even those that utilize sophisticated deep learning approaches. In addition, automated approaches generally require substantial investment in modifying and tuning parameters to identify optimal settings. Therefore, although automated approaches substantially reduce processing time relative to manual review, they still require high levels of human investment throughout the process.

Objectives

We have three main objectives with this manuscript. Although more sophisticated methods of automated detection that utilize deep learning approaches exist (e.g., Dufourq et al., 2021, 2022; Wang et al., 2022), these methods generally require substantial training datasets and are not readily available for users of the R programming environment (R Core Team, 2022). However, see (Silva et al., 2022) for a comprehensive deep-learning R package that relies heavily on Python. We aim to provide an open-source, step-by-step workflow for the automated detection and classification of Northern grey gibbon (*H. funereus*; hereafter gibbons) female calls using readily available machine learning algorithms in the R programming environment. The results of our study will provide an important benchmark for automated detection/classification applications for gibbon female great calls. We also test whether a post-processing step that utilizes unsupervised clustering can help improve the performance of our system, namely if this approach can help further differentiate between true and false positives. Lastly, as there have been relatively few studies of gibbons that utilize automated detection methods to address a well-defined research question (but see Dufourq et al., 2021 for an example on Hainan gibbons), we aimed to show how PAM can be used to address two different research questions. Specifically, we aim to answer the questions: (1) can we use unsupervised approaches to estimate how many female individuals (or clusters) there are in our study area, and (2) can this approach be used to investigate temporal patterns of gibbon calling activity? We utilized affinity propagation clustering to estimate the number of females (or clusters) in our dataset (Dueck, 2009). This unsupervised clustering algorithm has been shown to be useful for identifying the number of gibbon females in a labeled dataset (Clink and Klinck, 2021). To investigate temporal patterns of calling activity, we compared estimates derived from our automated system to those obtained using manual annotations from LTSAs by a human observer (Clink et al., 2020b).

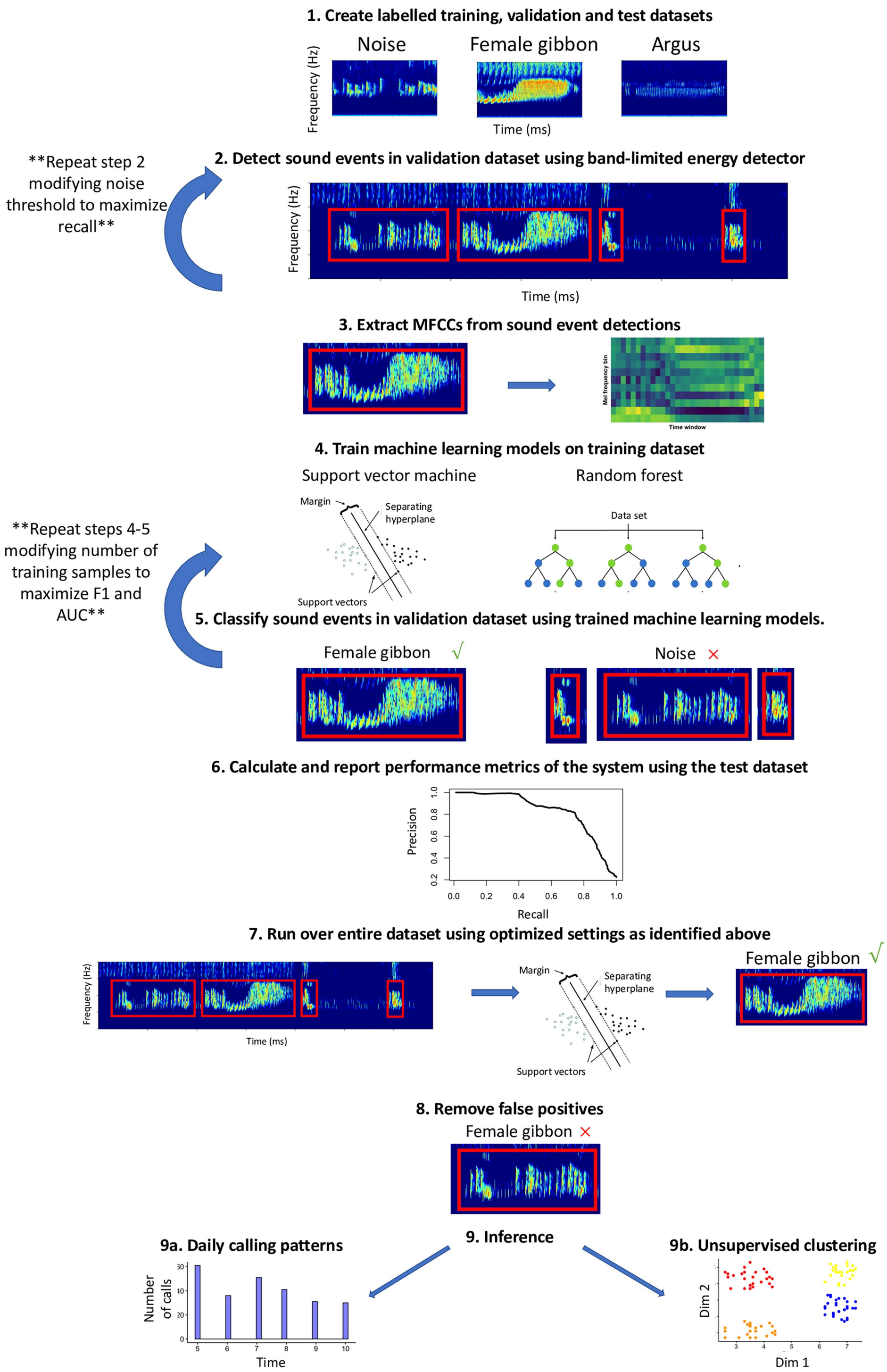


FIGURE 1
Schematic of automated detection/classification workflow presented in the current study. See the text for details about each step.

Materials and methods

Data collection

Acoustic data were collected using first-generation Swift recorders (Koch et al., 2016) developed by the K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. The sensitivity of the used microphones was −44 (+/−3) dB re 1 V/Pa. The microphone's frequency response was not measured but is assumed to be flat (+/− 2 dB) in the frequency range 100 Hz to 7.5 kHz. The analog signal was amplified by 40 dB and digitized (16-bit resolution) using an analog-to-digital converter (ADC) with a clipping level of −/+ 0.9 V. Recordings were saved as consecutive two-hour Waveform Audio File Format (WAV) with a size of approximately 230 MB. We recorded using a sampling rate of 16 kHz, giving a Nyquist frequency of 8 kHz, which is well above the range of the fundamental frequency of Northern grey gibbon calls (0.5 to 1.6 kHz). We deployed eleven Swift autonomous recording units spaced on a 750-m grid encompassing an area of approximately 3 km^2 in the Danum Valley Conservation Area, Sabah, Malaysia (4°57′53.00″N, 117°48′18.38″E) from February 13–April 21, 2018. We attached recorders to trees at approximately 2-m height and recorded continuously over a 24-h period.

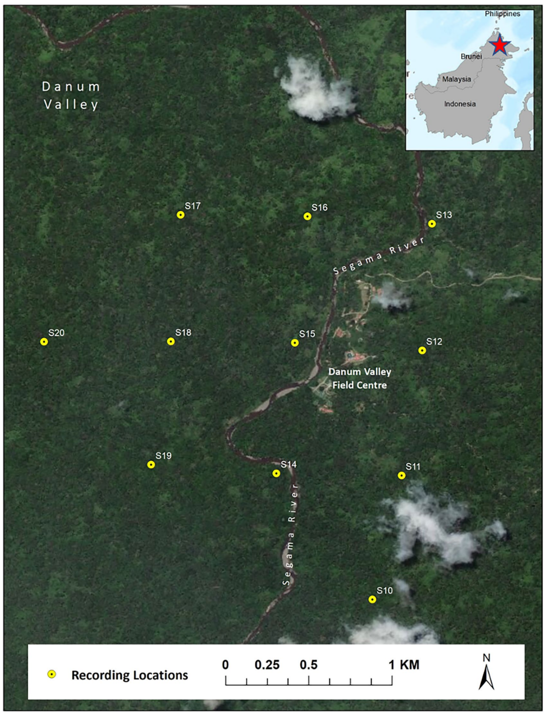


FIGURE 2
Map of recording locations of Swift autonomous recording units in Danum Valley Conservation Area, Sabah, Malaysia.

Source height (Darras et al., 2016) and presumably recorder height can influence the detection range of the target signal, along with the frequency range of the signal, levels of ambient noise in the frequency range of interest, topography, and source level of the calling animal (Darras et al., 2018). Given the monetary and logistical constraints for placing recorders in the canopy, we opted to place the recorders at a lower height. Our estimated detection range is approximately 500 meters using the settings described below (Clink and Klinck, 2019), and future work investigating the effect of recorder height on detection range will be informative. Danum Valley Conservation Area encompasses approximately 440 km^2 of lowland dipterocarp forest and is considered 'aseasonal' as it does not have distinct wet and dry seasons like many tropical forest regions (Walsh and Newbery, 1999). Gibbons are less likely to vocalize if there was rain the night before, although rain appears to have a stronger influence on male solos than coordinated duets (Clink et al., 2020b). The reported group density of gibbons in the Danum Valley Conservation Area is ~4.7 per km^2 (Hanya and Bernard, 2021), and the home range size of two groups was reported as 0.33 and 0.34 km^2 (33 and 34 ha; Inoue et al., 2016).

We limited our analysis to recordings taken between 06:00 and 11:00 local time, as gibbons tend to restrict their calling to the early morning hours (Mitani, 1985; Clink et al., 2020b), which resulted in a total of over 4,500 h of recordings for the automated detection. See Clink et al. (2020b) for a detailed description of the study design and Figure 2 for a study area map. On average, the gibbon duets at this site are 15.1 min long (range = 1.6–55.4 min) (Clink et al., 2020b). The duets are comprised of combinations of notes emitted by both the male and female, often with silent gaps of varying duration between the different components of the duet. The variability of note types and silent intervals in the duet would make training an automated detector/classifier system to identify any component of the duet a challenge (especially in the absence of a lot of training data). In addition, focusing on a certain call type within the longer vocalization is the established approach for automated detection/classification of gibbon vocalizations (Dufourq et al., 2021). Therefore, our automated detection/classification approach focused on the female great call. See Figure 3 for a representative

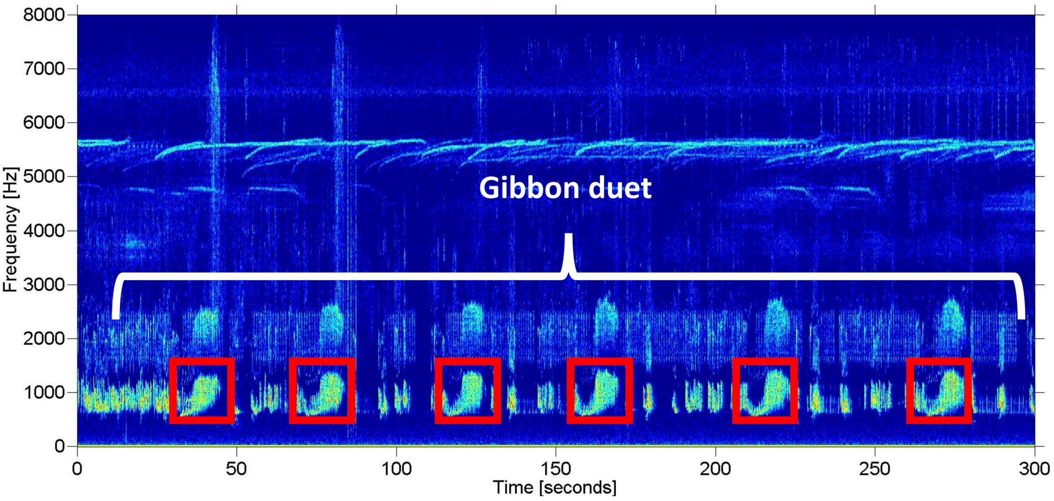


FIGURE 3
Representative spectrogram of a Northern grey gibbon duet recorded in Danum Valley Conservation Area, Sabah, Malaysia. The white bracket indicates a portion of the gibbon duet (also known as a bout), and the red boxes indicate unique great calls emitted by the gibbon female. The spectrogram was created using the Matlab-based program Triton (Wiggins, 2003).

spectrogram of a Northern grey gibbon duet and female great calls within the duet.

Creating a labeled training dataset

It is necessary to validate automated detection and classification systems using different training and test datasets (Heinicke et al., 2015). We randomly chose approximately 500 h of recordings to use for our training dataset and used a band-limited energy detector (settings described below) to identify potential sounds of interest in the gibbon frequency range, which resulted in 1,439 unique sound events. The subsequent sound events were then annotated by a single observer (DJC) using a custom-written function in R to visualize the spectrograms into the following categories: great argus pheasant (*Argusianus argus*) long and short calls (Clink et al., 2021), helmeted hornbills (*Rhinoplax vigil*), rhinoceros hornbills (*Buceros rhinoceros*), female gibbons and a catch-all "noise" category. For simplicity of training the machine learning algorithms, we converted our training data into two categories: "female gibbon" or "noise," and subsequently trained binary classifiers, although the classifiers can also deal with multi-class classification problems. The binary noise class contained all signals that were not female gibbon great calls, including great argus pheasants and hornbills. To investigate how the number of training data samples influences our system's performance, we randomly subset our training data into batches of 10, 20, 40, 80, 160, 320, and 400 samples of each category (female gibbon and noise) over 10 iterations each. We were also interested to see how the addition of high-quality focal recordings influenced the performance, so we added 60 female gibbon calls collected during focal recordings from previous field seasons (Clink et al., 2018b) to a set of training data. We compared the performance of the detection/classification system using random iterations to the training dataset containing all the training data samples ($n = 1{,}439$) and the dataset with the female calls added.

Sound event detection

Detectors are commonly used to isolate potential sound events of interest from background noise (Delacourt and Wellekens, 2000; Davy and Godsill, 2002; Lu et al., 2003). In this workflow, we identified potential sound events based on band-limited energy summation (Mellinger et al., 2016). For the band-limited energy detector (BLED), we first converted the 2-h recordings to a spectrogram (made with a 1,600-point (100 ms) Hamming window (3 dB bandwidth = 13 Hz), with 0% overlap and a 2,048-point DFT) using the package "seewave" (Sueur et al., 2008). We filtered the spectrogram to the frequency range of interest (in the case of Northern grey gibbons 0.5–1.6 kHz). For each non-overlapping time window, we calculated the sum of the energy across frequency bins, which resulted in a single value for each 100 ms time window. We then used the "quantile" function in base R to calculate the threshold value for signal versus noise. We ran early experiments using different quantile values and found that using the 15th quantile gave the best recall for our signal of interest. We then considered any events which lasted for 5 s or longer to be detections. Note that settings for the band-limited energy detector, MFCCs, and machine learning algorithms can be modified; we modified the detector and MFCC settings as independent steps in early experiments. We found in early experiments that modifying the quantile values and the duration of the detections influenced the performance of our system, so we suggest practitioners adopting this method experiment with modifying these settings to fit their system.

Supervised classification

We were interested in testing the performance of secondary classifiers —support vector machine (SVM) or random forest (RF) — for classifying our detected sound events. To train the classifiers, we used the training datasets outlined above and calculated Mel-frequency cepstral coefficients (MFCCs) for each of the labeled sound events using the R package "tuneR" (Ligges et al., 2016). We calculated MFCCs focusing on the fundamental frequency range of female gibbon calls (0.5–1.6 kHz). We focused on the fundamental frequency range because harmonics are generally not visible in the recordings unless the animals were very close to the recording units. As the duration of sound events is variable, and machine learning classification approaches require feature vectors of equal length, we averaged MFCCs over time windows. First, we divided each sound event into 8 evenly spaced time windows (with the actual length of each window varying based on the duration of the event) and calculated 12 MFCCs along with the delta coefficients for each time window (Ligges et al., 2016). We appended the duration of the event onto the MFCC vector, resulting in a vector for each sound event of length 177. We then used the E1071 package (Meyer et al., 2017) to train a SVM and the "randomForest" package (Liaw and Wiener, 2002) to train a RF, respectively. Each algorithm assigned each sound event to a class ("female gibbon" or "noise") and returned an associated probability. For SVM, we set "cross = 25," meaning that we used 25-fold cross-validation, set the kernel to "radial," and used the "tune" parameter to find optimal settings for the cost and gamma parameters. For the random forest algorithm, we used the default settings apart from setting the number of trees = 10,000.

Validation and test datasets

We annotated our validation and test datasets using a slightly different approach than we used for the training data. We did this because our system utilizes a band-limited energy detector. If we simply labeled the resulting clips (like we did with the training data), our performance metrics would not account for the detections that were missed initially by the detector. Therefore, to create our test and validation datasets, one observer (DJC) manually annotated 48 randomly chosen hours of recordings taken from different recorders and times across our study site using spectrograms created in Raven Pro 1.6. Twenty-four hours were used for validation, and the remaining 24 h were used as a test dataset to report the final performance metrics of the system. For each sound file, we identified the begin and end time of any female gibbon vocalization. We also labeled calls as high quality (wherein the full structure of the call was visible in the spectrogram and there were no overlapping bird calls or other background noises) or low quality (wherein the call was visible in the spectrogram, but the full structure was not, or there was overlapping with another calling animal/noise). As the detector isolates sound events based on energy in a certain frequency band, sometimes the start time of the detection does not align exactly with the annotated start time of the call; therefore, when calculating the performance metrics we considered sound events that started 4 s before the annotations or 2 s after the annotations to be a match.

We evaluated our system using five different metrics using the R package 'ROCR' (Sing et al., 2005) to calculate precision, recall, and false alarm rate. We were interested to see how the performance of our classifiers varied when we changed the probability threshold, so we calculated the area under the precision-recall curves, which shows the trade-off between the rate of false-positives and false-negatives at different probability thresholds. We calculated the area under the receiver operating characteristic curve (AUC) for each machine learning algorithm and training dataset configuration. We also calculated the F1 score, as it integrates both precision and recall information into the metric.

We used a model selection approach to test for the effects of training data and machine learning algorithm on our performance metrics (AUC), so we created as series of two linear models using the R package "lme4" (Bates et al., 2017). The first model we considered, the null model, had only AUC as the outcome, with no predictor variables. The second model, which we considered the full model, contained the machine learning algorithm (SVM or RF) and training data category as predictors. We used the Akaike information criterion (AIC) to compare the fit of the two models to our data, implemented in the "bbmle" package (AICctab adjusted for small sample sizes; Bolker, 2014). We chose the settings that maximized AUC and the F1-score for the subsequent analysis of the full dataset (described below).

Verification workflow

The optimal detector/classifier settings for our two main objectives were slightly different. For our first objective, wherein we wanted to compare patterns of vocal activity based on the output of our automated detector to patterns identified using human-annotated datasets (Clink et al., 2020b), we aimed to maximize recall while also maintaining an acceptable number of false positives. In early tests, we found that using a smaller quantile threshold (0.15) for the BLED detector improved recall. One observer (IK) manually verified all detections using a custom function in R that allows observers to quickly view spectrograms and verify detections. Although duet bouts contain many great calls, we considered instances where at least one great call was detected during each hour as the presence of a duet. We then compared our results to those identified using a human observer and calculated the percent of annotated duets the automated system detected. To compare the two distributions, we used a Kolmogorov–Smirnov test implemented using the 'ks.test' function in the R version 4.2.1 programming environment (R Core Team, 2022). We first converted the times to "Unix time" (the number of seconds since 1970-01-01 00:00:00 UTC; Grolemund and Wickham, 2011) so that we had continuous values for comparison. We used a non-parametric test as we did not assume a normal distribution of our data.

For the objective wherein we used unsupervised clustering to quantify the number of females (clusters) in our dataset, we needed higher quality calls in terms of signal-to-noise ratio (SNR) and overall structure. This is because the use of MFCCs as features for discriminating among individuals is highly dependent on SNR (Spillmann et al., 2017). For this objective, we manually omitted all detections that did not follow the species-specific structure with longer introductory notes that transition into rapidly repeating trill notes and only used detections with a probability >0.99 as assigned by the SVM (Clink et al., 2017).

Unsupervised clustering

We used unsupervised clustering to investigate the tendency to cluster in: (A) the verified detections containing true and false positives after running the detector/classifier over our entire dataset: and (B) female calls that follow the species-specific structure of the great call wherein different clusters may reflect different individuals. We used affinity propagation clustering, a state-of-the-art unsupervised approach (Dueck, 2009) that has been used successfully in a few bioacoustics applications, including anomaly detection in a forest environment (Sethi et al., 2020) and clustering of female gibbon calls with known identity (Clink and Klinck, 2021). Our previous work showed that out of three unsupervised algorithms compared, affinity propagation clustering returned a number of clusters that matched the number of known female individuals in our dataset most closely (Clink and Klinck, 2021). Input preferences for the affinity propagation clustering algorithm can vary the number of clusters returned. We initially used an adaptive approach wherein we varied the input preference from 0.1 to 1 in increments of 0.1 (indicated by "q" in the "APCluster" R package; Bodenhofer et al., 2011) and calculated silhouette coefficients using the "cluster" package (Maechler et al., 2019). We found that the optimal q identified in this manner led to an unreasonably high number of clusters for the true/false positives, so we set $q = 0.1$, resulting in fewer clusters.

We input an MFCC vector for each sound event into the affinity propagation clustering algorithm. For the true/false positives, we calculated the MFCCs slightly differently than outlined above, as fewer features resulted in better clustering. Instead of creating a standardized number of time windows for each event, we calculated MFCCs for each sound event using the default settings (wintime = 0.025, hoptime = 0.01, and numcep = 12). We then took the mean and standard deviation for each Mel-frequency band and the delta coefficients, resulting in 48 unique values for each sound event. We also included the duration of the signal. For the true and false positive detections, we used normalized mutual information (NMI) as an external validation measure implemented in the 'aricode' package (Chiquet and Rigaill, 2019). NMI provides a value between 0 and 1, with 1 indicating a perfect match between two sets of labels (Xuan et al., 2010). For clustering of the high-quality female calls, we used the adaptive approach to find the optimal value of q. We used the standard number of MFCC windows approach as outlined above.

To visualize clustering in our dataset, we used a uniform manifold learning technique (UMAP) implemented in the R package 'umap' (Konopka, 2020). UMAP is a data reduction and visualization approach that has been used to visualize differences in forest soundscapes (Sethi et al., 2020), taxonomic groups of neotropical birds (Parra-Hernández et al., 2020), and female gibbon great calls (Clink and Klinck, 2021).

Data availability

A tutorial, annotated code, and all data needed to recreate figures presented in the manuscript are available on GitHub.[1] Access to raw sound files used for training and testing can be granted by request to the corresponding author.

---

[1] https://github.com/DenaJGibbon/Workflow-for-automated-detection-and-classification-gibbon-calls

TABLE 2 Summary of precision, recall, F1, and area under the curve (AUC) calculated using the validation dataset for random subsets of training data compared to the full training dataset and the full dataset augmented with female great calls.

Training data	Algorithm	Precision (mean±sd)	Recall (mean±sd)	F1 (mean±sd)	AUC (mean±sd)
$n = 10$	RF	0.96 ± 0.07	0.3 ± 0.18	0.45 ± 0.2	0.76 ± 0.01
	SVM	0.95 ± 0.08	0.41 ± 0.22	0.58 ± 0.22	0.79 ± 0.01
$n = 20$	RF	0.97 ± 0.04	0.32 ± 0.18	0.48 ± 0.21	0.74 ± 0.05
	SVM	0.96 ± 0.08	0.43 ± 0.22	0.59 ± 0.22	0.73 ± 0.31
$n = 40$	RF	1 ± 0.03	0.35 ± 0.18	0.52 ± 0.21	0.72 ± 0.03
	SVM	0.96 ± 0.05	0.52 ± 0.22	0.68 ± 0.22	0.73 ± 0.05
$n = 80$	RF	1 ± 0.03	0.37 ± 0.18	0.53 ± 0.21	0.73 ± 0.03
	SVM	0.95 ± 0.05	0.59 ± 0.18	0.73 ± 0.17	0.77 ± 0.03
$n = 160$	RF	1 ± 0.04	0.41 ± 0.2	0.58 ± 0.21	0.74 ± 0.02
	SVM	0.94 ± 0.06	0.68 ± 0.14	0.77 ± 0.11	0.77 ± 0.01
$n = 320$	RF	1 ± 0.04	0.46 ± 0.2	0.63 ± 0.22	0.76 ± 0.01
	SVM	0.93 ± 0.06	0.71 ± 0.11	0.79 ± 0.08	0.81 ± 0.01
$n = 400$	RF	1 ± 0.04	0.48 ± 0.2	0.65 ± 0.21	0.76 ± 0
	SVM	0.92 ± 0.06	0.71 ± 0.13	0.79 ± 0.11	0.82 ± 0.01
All	RF	1 ± 0.02	0.34 ± 0.19	0.51 ± 0.22	0.76 ± NA
	SVM	**0.94 ± 0.05**	**0.71 ± 0.12**	**0.8 ± 0.09**	**0.83 ± NA**
All + F	RF	1 ± 0.02	0.39 ± 0.19	0.56 ± 0.21	0.76 ± NA
	SVM	0.94 ± 0.05	0.72 ± 0.17	0.8 ± 0.16	0.83 ± NA

Precision, recall, and F1 values reported are for probability thresholds >0.50. Performance metrics were calculated using the 'ROCR' package (Sing et al., 2005). These metrics were used to determine which settings resulted in the best performance of the system. The bold indicates the best performing settings that were used for subsequent analysis of our entire dataset.

Results

Training data and algorithm influence performance

The classification accuracy of SVM for the training dataset containing all samples was 98.82%, and the accuracy of the RF was 97.85%. We found that the number of training data samples and the selected machine learning algorithm substantially influenced the performance of our detector/classifier using the validation dataset (Table 2). Using an AIC model selection approach, we found that the model with AUC as an outcome and with the machine learning algorithm and training data category as predictors performed much better than the null model ($\Delta AICc = 11.2$; 100% of model weight). When using AUC as the metric, we found that SVM performed slightly better than RF, and performance normalized when the number of training samples was greater than $n = 160$ (Figure 4). We also found that the model with F1 score as an outcome and machine learning algorithm and training data category as predictors performed much better than the null model ($\Delta AICc = 34{,}730.6$; 100% of model weight; Figure 4). Again, SVM performed better than RF, but in this case, the training dataset that contained all the samples ($n = 433$ female calls and $n = 1{,}006$ noise events) or all the samples plus extra female calls performed better (Figure 4). There were noticeable differences in the performance of the two algorithms regarding F1 score across different probability thresholds (Figure 5). SVM had a higher performance at higher probability thresholds, whereas performance for RF had the highest F1 value when the probability threshold was 0.60. We decided to use the SVM algorithm with all the training samples for our full analysis. We used the 24-h test dataset to calculate the final performance metrics of our system. We found that the highest F1 score (0.78) was when the probability threshold was 0.90, precision was 0.88, and recall was 0.71.

Comparison of an automated system to human annotations

We used the SVM algorithm and all training data samples to run over our full dataset resulting in 4,771 detections, of which 3,662 were true positives and 1109 were false positives (precision = 0.77). A histogram showing the distributions of automatically detected calls and manually annotated calls is shown in Figure 6. A Kolmogorov–Smirnov test indicated that the two distributions were not significantly different $D = 0.07$, $p > 0.05$.

Unsupervised clustering

We used unsupervised clustering to investigate the tendency to cluster in true/false positives and high-quality female calls. For our first aim, we used affinity propagation clustering to differentiate between true and false positives after we used our detection/classification system. We did not find that affinity propagation clustering effectively separated false positives, as the NMI score was close to zero ($NMI = 0.03$). Although there were only two classes in our dataset (true and false positives), the clustering results indicated 53 distinct clusters. Supervised

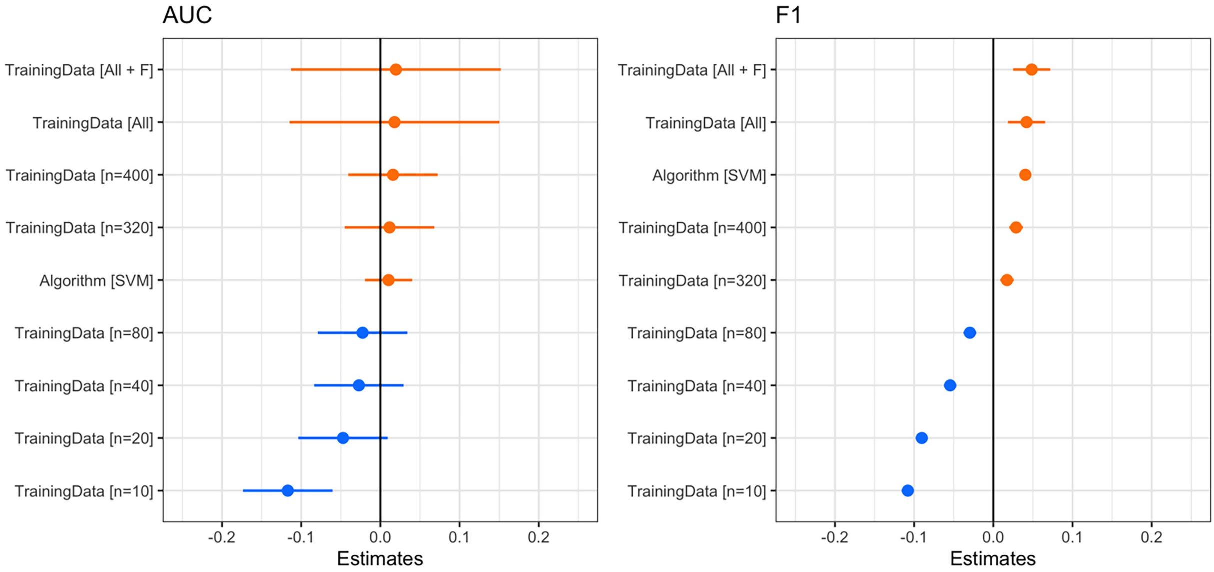


FIGURE 4
Coefficient plots from the linear model with AUC (left) or F1 score (right) as the outcome and training data category and machine learning algorithm as predictors. Using AIC, we found that both models performed substantially better than the null model. For both coefficient plots, the reference training data category is *n*=160. We considered predictors to be reliable if the confidence intervals did not overlap zero. For AUC (left), training data samples smaller than *n*=160 had a slightly negative impact on AUC, whereas a larger number of training data samples had a slightly positive impact. Note that the confidence intervals overlap zero, so these can be interpreted only as trends. The use of the SVM algorithm had a slightly positive effect on AUC. For the F1 score (right), the number of training samples had a reliable effect on the F1 score. When samples were less than *n*=160, the F1 score was lower. When there were more samples, the F1 score was higher. SVM also had a reliably positive effect on the F1 score.

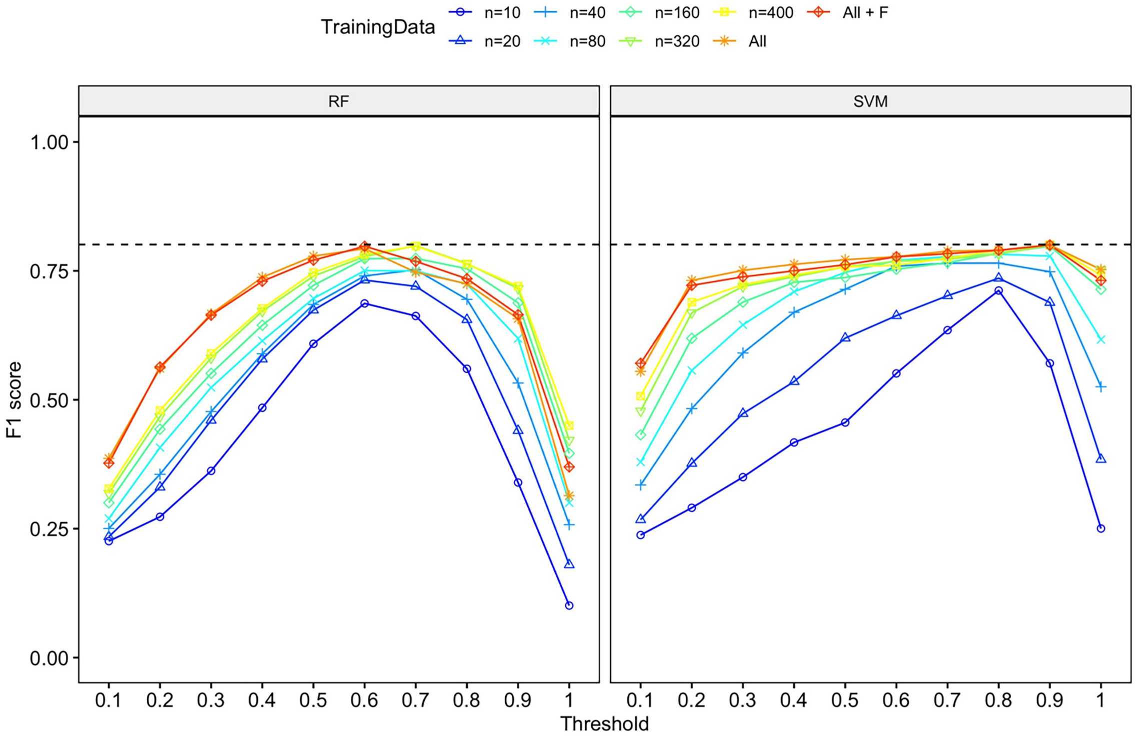


FIGURE 5
F1-score for each machine learning algorithm (RF or SVM), probability threshold category, and training data category. Both algorithms had comparable performance in terms of F1 score, although the probability threshold with the highest F1 score differed. The dashed line indicates the highest F-1 score (0.80) for both algorithms on the validation dataset.

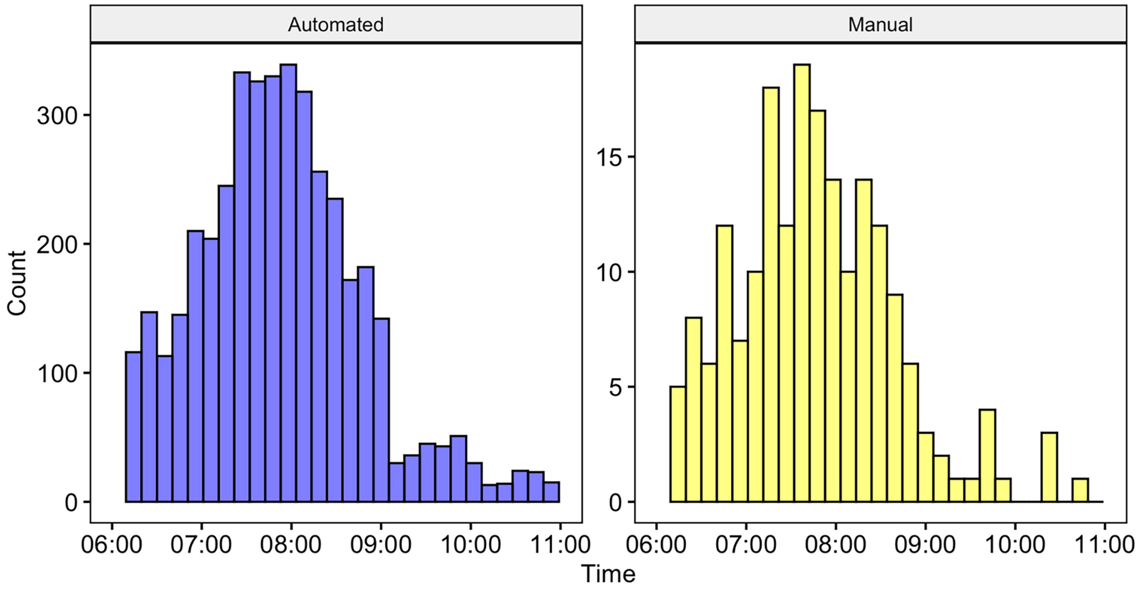


FIGURE 6
Histogram showing the number of calls detected by time using the automated system (left) and manually annotated by a human observer (right). Note that the differences in axes are due to the detections for the automated system being at the call level, whereas the annotations were at the bout level (and bouts are comprised of multiple calls). There was no statistically significant difference between the two distributions (see text for details).

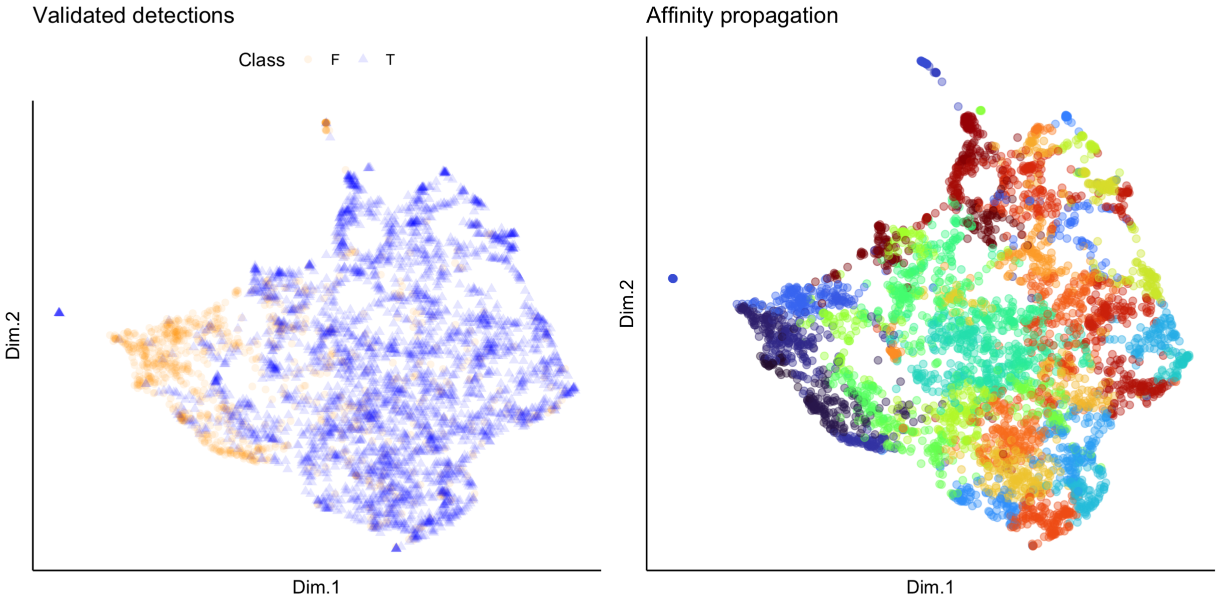


FIGURE 7
UMAP projections indicating validated detections (left) and cluster assignment by affinity propagation clustering (right). Each point represents a detection, and the colors in the plot on the left indicate whether the detection was a true (T; indicated by the blue triangles) or false (F; indicated by the orange circles) positive. The colors in the plot on the right indicate which of the 53 clusters returned by the affinity propagation clustering algorithm the detection was assigned to.

classification accuracy using SVM for true and false positives was ~95%. UMAP projections of the true and false positive detections are shown in Figure 7. For our second aim, we used affinity propagation clustering to investigate the tendency to cluster in the high-quality female calls detected by our system ($n = 194$). Using adaptive affinity propagation clustering, we found that setting $q = 0.2$ resulted in the highest silhouette coefficient (0.18) and returned ten unique clusters. UMAP projections of female calls are shown in Figure 8. Histograms indicating the number of calls from each recorder assigned to each cluster by the affinity propagation algorithm are shown in Figure 9.

Discussion

We show that using open-source R packages, a detector and classifier can be developed with an acceptable performance that exceeds that of

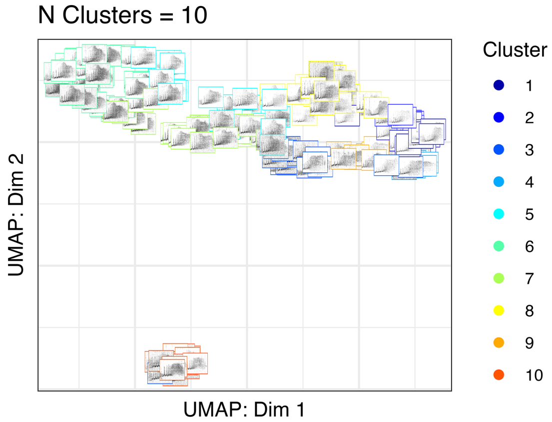


FIGURE 8
UMAP projections of Northern grey gibbon female calls. The location of each spectrogram indicates the UMAP projection of the call, and the border color indicates cluster assignment by affinity propagation clustering.

previously published automated detector/classifiers for primate calls [e.g., Diana monkey F1 based on reported precision and recall = 65.62 (Heinicke et al., 2015)]. However, the performance of this system (maximum F1 score = 0.78) was below some reported deep learning approaches [e.g., F1 score = 90.55 for Hainan gibbons (Dufourq et al., 2021), F1 score = ~87.5 for owl species (Ruff et al., 2021), F1 score = 87.0 for bats (Silva et al., 2022)]. In addition, we found that temporal patterns of calling based on our automated system matched those of the human annotation approach. We also tested whether using an unsupervised approach (affinity propagation clustering) could help further distinguish true and false positive detections but found that the clustering results ($n = 53$ clusters) did not differentiate true and false positives in any meaningful way. Visual inspection of the false positives indicated that many of them were overlapping with great argus pheasants, or were other parts of the gibbon duet or solo. A majority of the false positives were male solo phrases, and these vocalization types contain rapidly repeating notes in the same frequency range as the female gibbon call. Lastly, we applied unsupervised clustering to a reduced dataset of validated detections of female calls that followed the species-specific call structure and found evidence for ten unique clusters. Inspection of spatial patterns of distribution of the clusters indicates that the clusters do not correlate with female identity.

Calls versus bouts

Our analysis focused on one call type within the gibbon duet: the female great call. We did this for practical reasons, as female calls tend to be stereotyped and follow a species- and sex-specific pattern. In addition, females rarely call alone, which means the presence of the female call can be used to infer the presence of a pair of gibbons. Also, most acoustic survey methods focus on the duet for the reasons described above, and generally, only data on the presence or absence of a duet bout at a particular time and location are needed (Brockelman and Srikosamatara, 1993; Kidney et al., 2016). When calculating the performance of our automated detection/classification system, we focused on the level of the call, as this is a common way to evaluate the performance of automated systems (Dufourq et al., 2020). Finally, when comparing temporal patterns of calling behavior, we compared to an existing dataset of annotations at the level of the duet bout. We did this because annotating duet bouts using LTSAs is much more efficient than annotating each individual call for the entire dataset. However, for certain applications such as individual vocal signatures, the analysis necessarily focuses on individual calls within a bout.

Comparison of ML algorithms

We found that SVM performed slightly better than RF in most metrics reported (except precision). However, RF had a comparable classification accuracy to SVM on the training dataset (SVM accuracy = 98.82% and RF accuracy = 97.85% for all training data samples). This reduced performance can be attributed to the substantially lower recall of RF relative to SVM, despite RF having higher precision in many cases (data summarized in Table 2). The precision of SVM decreased slightly as we increased the number of training samples, which may be due to increased variability in the training data samples that influenced the algorithm's precision. We did not see that the precision of RF decreased with an increased number of training samples, but RF recall remained low regardless of the amount of training data. These patterns are reflected in differences in the F1 scores across probabilities for both algorithms.

The tolerable number of false positives, or the minimum tolerable recall of the system, will depend heavily on the research question. For example, when modeling occupancy, it may be important that no calls are missed, and hence, a higher recall would be desirable. But, for studies that focus more on the behavioral ecology of the calling animals (Clink et al., 2020a,b), it may be important for the detector to identify calls with a low amount of false positives but less important if the detector misses many low signal-to-noise calls. Therefore, in some cases where high precision is desired but recall is less important, RF may be a better choice. It is also possible that tuning the RF (as we did with SVM) may result in better performance. However, we did not do this as it is generally agreed that RF works well using default values of the hyperparameters (Probst et al., 2019).

Influence of training data

We found that the AUC and F1 metrics normalized when using 160 samples of training data or more for each of the two classes (gibbon female and noise). However, using all training data or data augmented with female calls resulted in better F1 scores. The training datasets that contained all the samples and added females were unbalanced and contained many more noise samples than female calls. Including more diverse noise samples lead to better performance in this system, and both RF and SVM handle unbalanced datasets effectively. It is important to note that although we found performance normalized when training with 160 calls or more, this number does not account for the additional number of calls needed for validation and training. Therefore, the total number of calls or observations to effectively train and subsequently evaluate the performance of the system will be >160 calls. We realize that compiling a dataset of 160 or more calls for rare sound events from elusive species may be unrealistic. We found that our in our system including as few as 40 calls allowed for acceptable performance (F1 score for SVM = 0.70), so the approach could be potentially used successfully with a much smaller training dataset.

In addition, our training, validation, and test datasets came from different recording units, times of day, and multiple territories of different gibbon groups. Including 40 calls from the same recorder and same individual would presumably not be as effective as including calls from different individuals and recording locations. A full discussion of the effective preparation of datasets for machine learning is out of the scope of the present paper, but readers are urged to think carefully about the preparation of acoustic datasets for automated detection and aim to include samples from a diverse number of recording locations, individuals and time of day. Transfer learning which utilizes pre-trained convolutional neural networks for different classification problems than the model was originally trained, provides another alternative for small datasets, with transfer learning providing up to an 82% F1 score with small datasets (Dufourq et al., 2022). Future work that compares the approach presented here with transfer learning will be highly informative.

Unsupervised clustering to distinguish true/false positives

We did not find that affinity propagation clustering helped further differentiate true and false positives in our dataset, despite being able to differentiate between the two classes using supervised methods with ~95% accuracy. As noted above, many of the false positives were phrases from male solos, and these phrases are highly variable in note order and note sequence (Clink et al., 2020a), which may have led to the high number of clusters observed. The NMI score was close to zero, indicating a lack of accordance between the unsupervised cluster assignments and the true labels. These types of unsupervised approaches have been fruitful in distinguishing among many different types of acoustic signals, including soundscapes (Sethi et al., 2020), bird species (Parra-Hernández et al., 2020), and gibbon individuals (Clink and Klinck, 2021). We extracted MFCCs for all sound events focusing on the relevant frequency range for female gibbon great calls. As detections were based on band-limited energy summation in this frequency range, extracting MFCCs in this frequency range was a logical choice. We did early experiments where we summarized the extracted MFCCs in different ways and slightly modified the frequency range. We did not find that these early experiments led to better separation of true and false positives. Therefore, we conclude that the use of MFCCs and affinity propagation clustering is not an effective way to differentiate between true and false positives in our dataset. It is possible that using different features may have led to different results, and embeddings from convolutional neural networks as features (e.g., Sethi et al., 2020) or the use of low dimentional latent space projections learned from the spectrograms (Sainburg et al., 2020) are promising future directions.

Unsupervised clustering of validated gibbon female calls

The ability to distinguish between individuals based on their vocalizations is important for many different PAM applications, and population density estimation in particular (Augustine et al., 2018, 2019). The home range size of two gibbon pairs in our population was previously reported to be about 0.34 km^2 (34 ha; Inoue et al., 2016), but within gibbon populations, the home range size can vary substantially (Cheyne et al., 2019), making it difficult to know exactly how many pairs were included in our study area. In another study, gibbon group density was reported as 4.7 groups per km^2; the discrepancy between this value and home range size estimates provided by Inoue et al. (2016) is presumably due to the fact that the studies were measuring different parameters (density vs. home range) and the fact that home ranges can overlap, even in territorial species. Therefore, based on conservative estimates of gibbon density and home range size, up to 12 pairs may occur in our 3 km^2 study area.

Our unsupervised approach using affinity propagation clustering on high-quality female calls returned ten unique clusters. We showed that affinity propagation clustering consistently returned a similar number of clusters to the actual number of individuals in a different dataset (Clink and Klinck, 2021). However, an inspection of the histograms in Figure 9 shows that some clusters appear to have strong spatial patterns (e.g., only appearing on a few recorders in close spatial proximity), whereas others appear on many recorders. In some cases, the same clusters appear on recorders that are >1.5 km apart — a presumably larger distance than the width of a gibbon home range — therefore, it seems unlikely that these clusters are associated with female identity. When using unsupervised approaches, it is common practice to assign each cluster to the class that contains the highest number of observations, and we showed affinity propagation clustering reliably returned a number of clusters that matched the number of individuals in the dataset, but often 'misclassified' calls to the wrong cluster/individual (Clink and Klinck, 2021). Importantly, our previous work was done on high-quality, focal recordings with a substantial amount of preprocessing to ensure the calls were comparable (e.g., did not contain shorter introductory notes or overlap with the male). In the present study, we manually screened calls to ensure they followed the species-specific structure and were relatively high-quality, but the limitations of PAM data (collected using an omnidirectional, relatively inexpensive microphone, and at variable distances to the calling animals) may preclude effective unsupervised clustering of individuals.

We conclude that more work needs to be done before we can reliably use unsupervised methods to estimate the number of individuals in a study area. Our current ability to utilize these approaches to return the number of individuals reliably is presently limited, especially because there is not a lot of information regarding the stability of individual signatures over time; but see (Feng et al., 2014). Future work that utilizes labeled training datasets collected using PAM data to train classifiers that can subsequently predict new individuals (e.g., an approach similar to that presented in; Sadhukhan et al., 2021) will help further our ability to identify unknown individuals from PAM data.

Generalizability of the system

Gibbon female calls are well-suited for automated detection and classification as they are loud and highly stereotyped, and gibbon females tend to call often. During a particular calling bout, they emit multiple calls, allowing for ample training data. Although gibbon female calls are individually distinct (Clink et al., 2017, 2018a), the differences between individuals were not sufficient to

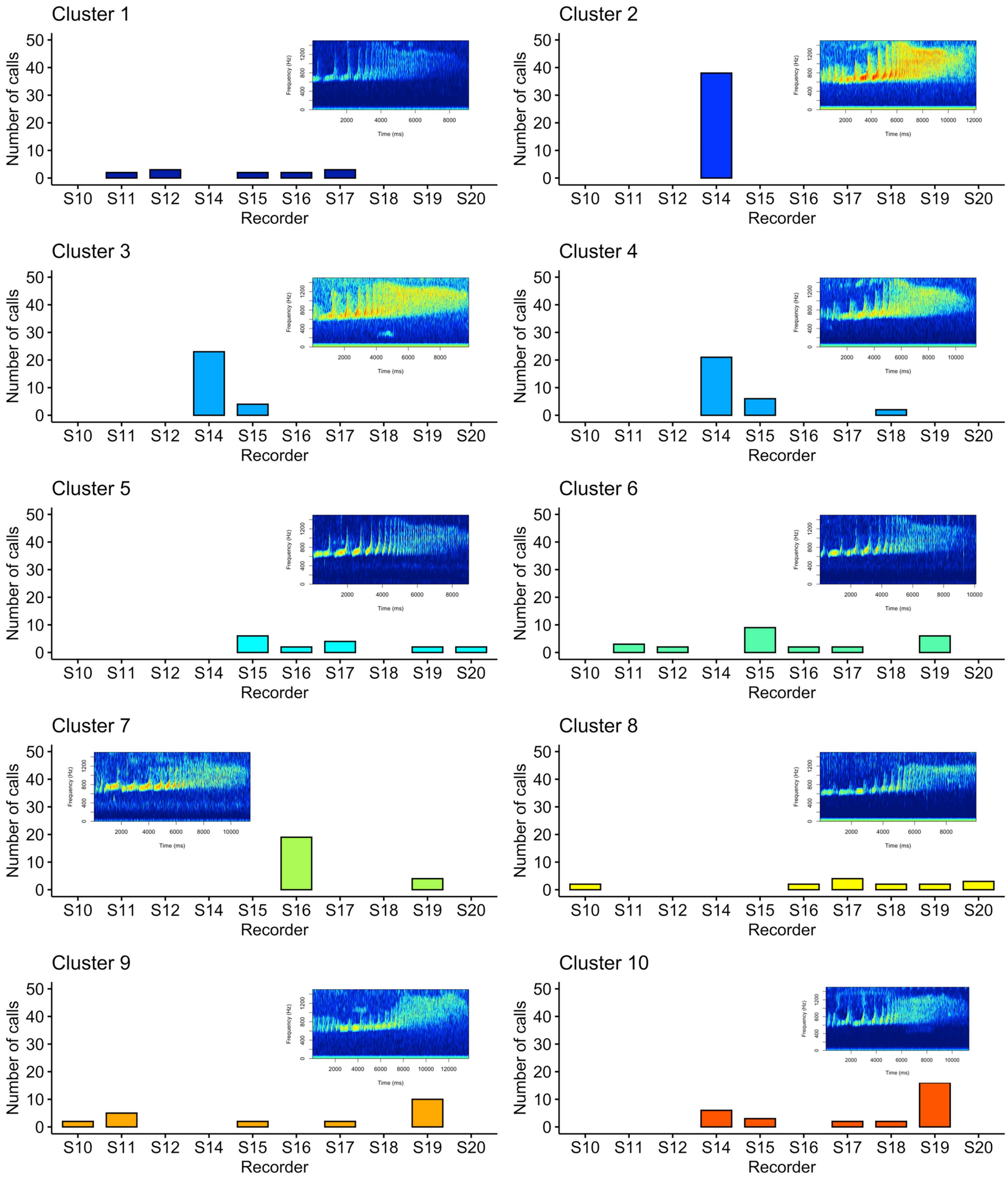


FIGURE 9
Histograms showing the number of calls assigned to each cluster by the affinity propagation algorithm. Each panel indicates one of the clusters as assigned by affinity propagation clustering, the *x*-axis indicates the associated recording unit where the call was detected, and the *y*-axis indicates the number of calls for each cluster and recorder. The spectrograms shown exemplify each cluster assigned by the affinity propagation clustering algorithm.

preclude detection and classification using our system. Importantly, the fact that gibbon female calls tend to be of longer duration (> 6-s) than many other signals in the frequency range meant that the duration of the signal could be used as an effective metric to reject nonrelevant signals. The generalizability of our methods to other systems/datasets will depend on a variety of conditions, in particular, the signal-to-noise ratio of the call(s) of interest, type and variability of background noise, the amount of stereotypy in the calls of interest, and the amount of training data that can be obtained to train the system. Future applications that apply this approach to other gibbon species, or compare this approach with deep learning techniques, will be important next steps to determine the utility and effectiveness of automated detection approaches for other taxa.

Future directions

Due to the three-step design of our automated detection, classification, and unsupervised clustering approach, modifying the system at various stages should be relatively straightforward. In particular, using MFCCs as features was a logical approach given how well MFCCs work to distinguish among gibbon calls [this paper and Clink et al. (2018a)]. However, it is possible that using different types of feature sets may result in even better performance of the automated system. As mentioned above, the use of embeddings from pre-trained convolutional neural networks is a possibility. In addition, the supervised classification algorithms included in our approach were not optimized; the RF algorithm, in particular, was implemented using the default values set by the algorithm developers.

Therefore, further tuning and optimization of the algorithms may also influence the performance. Lastly, this approach was developed using training, validation, and test data from one site (Danum Valley Conservation Area). Future work investigating the performance of this system in other locations with (presumably) different types of ambient noise will be informative.

Conclusion

Here we highlight how the open-source R-programming environment can be used to process and visualize acoustic data collected using autonomous recorders that are often programmed to record continuously for long periods of time. Even the most sophisticated machine learning algorithms are never 100% accurate or precise and will return false positives or negatives (Bardeli et al., 2010; Heinicke et al., 2015; Keen et al., 2017), which is also the case with human observers, but this is rarely quantified statistically (Heinicke et al., 2015). We hope this relatively simple automated detection/classification approach will serve as a useful foundation for practitioners interested in automated acoustic analysis methods. We also show that unsupervised approaches need further work and refinement before they can be reliably used to distinguish between different data classes recorded using autonomous recording units. Given the importance of being able to distinguish among individuals for numerous types of PAM applications, this should be a high-priority area for future research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Institutional approval was provided by Cornell University (IACUC 2017-0098).

Author contributions

DC, AA, and HK conceived the ideas and designed the methodology. DC and IK annotated and validated data. DC and IK led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding

DC acknowledges the Fulbright ASEAN Research Award for U.S. Scholars for providing funding for the field research.

Acknowledgments

The authors thank the two reviewers who provided valuable feedback that greatly improved the manuscript. We thank the makers of the packages “randomforest,” “e1017,” “seewave,” “signal,” and “tuneR,” on which this workflow relies extensively. We thank Yoel Majikil for his assistance with data collection for this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

References

Anders, F., Kalan, A. K., Kühl, H. S., and Fuchs, M. (2021). Compensating class imbalance for acoustic chimpanzee detection with convolutional recurrent neural networks. *Eco. Inform.* 65:101423. doi: 10.1016/j.ecoinf.2021.101423

Araya-Salas, M., and Smith-Vidaurre, G. (2017). warbleR: an R package to streamline analysis of animal acoustic signals. *Methods Ecol. Evol.* 8, 184–191. doi: 10.1111/2041-210X.12624

Augustine, B. C., Royle, J. A., Kelly, M. J., Satter, C. B., Alonso, R. S., Boydston, E. E., et al. (2018). Spatial capture–recapture with partial identity: an application to camera traps. *Ann. Appl. Stat.* 12, 67–95. doi: 10.1214/17-AOAS1091

Augustine, B. C., Royle, J. A., Murphy, S. M., Chandler, R. B., Cox, J. J., and Kelly, M. J. (2019). Spatial capture–recapture for categorically marked populations with an application to genetic capture–recapture. *Ecosphere* 10:e02627. doi: 10.1002/ecs2.2627

Balantic, C., and Donovan, T. (2020). AMMonitor: remote monitoring of biodiversity in an adaptive framework with r. *Methods Ecol. Evol.* 11, 869–877. doi: 10.1111/2041-210X.13397

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K. H., and Frommolt, K. H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recogn. Lett.* 31, 1524–1534. doi: 10.1016/j.patrec.2009.09.014

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. (2017). lme4: Linear Mixed-Effects Models Using “Eigen” and S4. R package version 1.1-13. Available at: http://keziamanlove.com/wp-content/uploads/2015/04/StatsInRTutorial.pdf https://cran.r-project.org/web/packages/lme4/lme4.pdf (Accessed January 23, 2023).

Bjorck, J., Rappazzo, B. H., Chen, D., Bernstein, R., Wrege, P. H., and Gomes, C. P. (2019). Automatic detection and compression for passive acoustic monitoring of the african forest elephant. *Proc. AAAI Conf. Artific. Intellig.* 33, 476–484. doi: 10.1609/aaai.v33i01.3301476

Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). APCluster: an R package for affinity propagation clustering. *Bioinformatics* 27, 2463–2464. doi: 10.1093/bioinformatics/btr406

Bolker, B. M. (2014). bbmle: tools for general maximum likelihood estimation. Available at: http://cran.stat.sfu.ca/web/packages/bbmle/ (Accessed January 23, 2023).

Brockelman, W. Y., and Srikosamatara, S. (1993). Estimation of density of gibbon groups by use of loud songs. *Am. J. Primatol.* 29, 93–108. doi: 10.1002/ajp.1350290203

Cheyne, S. M., Capilla, B. R., K, A., Supiansyah, , Adul, , Cahyaningrum, E., et al. (2019). Home range variation and site fidelity of Bornean southern gibbons [*Hylobates albibarbis*] from 2010–2018. *PLoS One* 14, e0217784–e0217713. doi: 10.1371/journal.pone.0217784

Chiquet, J., and Rigaill, G. (2019). Aricode: efficient computations of standard clustering comparison measures. Available at: https://cran.r-project.org/package=aricode (Accessed January 23, 2023).

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One* 1:e73. doi: 10.1371/journal.pone.0000073

Clink, D. J., Bernard, H., Crofoot, M. C., and Marshall, A. J. (2017). Investigating individual vocal signatures and small-scale patterns of geographic variation in female bornean gibbon (*Hylobates muelleri*) great calls. *Int. J. Primatol.* 38, 656–671. doi: 10.1007/s10764-017-9972-y

Clink, D. J., Crofoot, M. C., and Marshall, A. J. (2018a). Application of a semi-automated vocal fingerprinting approach to monitor Bornean gibbon females in an experimentally fragmented landscape in Sabah, Malaysia. *Bioacoustics* 28, 193–209. doi: 10.1080/09524622.2018.1426042

Clink, D. J., Grote, M. N., Crofoot, M. C., and Marshall, A. J. (2018b). Understanding sources of variance and correlation among features of Bornean gibbon (*Hylobates muelleri*) female calls. *J. Acoust. Soc. Am.* 144, 698–708. doi: 10.1121/1.5049578

Clink, D. J., Groves, T., Ahmad, A. H., and Klinck, H. (2021). Not by the light of the moon: investigating circadian rhythms and environmental predictors of calling in Bornean great argus. *PLoS One* 16:e0246564. doi: 10.1371/journal.pone.0246564

Clink, D. J., Hamid Ahmad, A., and Klinck, H. (2020a). Brevity is not a universal in animal communication: evidence for compression depends on the unit of analysis in small ape vocalizations. *R. Soc. Open Sci.* 7. doi: 10.1098/rsos.200151

Clink, D. J., Hamid Ahmad, A., and Klinck, H. (2020b). Gibbons aren't singing in the rain: presence and amount of rainfall influences ape calling behavior in Sabah, Malaysia. *Sci. Rep.* 10:1282. doi: 10.1038/s41598-020-57976-x

Clink, D. J., and Klinck, H. (2019). A case study on Bornean gibbons highlights the challenges for incorporating individual identity into passive acoustic monitoring surveys. *J. Acoust. Soc. Am.* 146:2855. doi: 10.1121/1.5136908

Clink, D. J., and Klinck, H. (2021). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods Ecol. Evol.* 12, 328–341. doi: 10.1111/2041-210X.13520

Cowlishaw, G. (1992). Song function in gibbons. *Behaviour* 121, 131–153. doi: 10.1163/156853992X00471

Cowlishaw, G. (1996). Sexual selection and information content in gibbon song bouts. *Ethology* 102, 272–284. doi: 10.1111/j.1439-0310.1996.tb01125.x

Dahake, P. P., and Shaw, K. (2016). Speaker dependent speech emotion recognition using MFCC and support vector machine. in International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 1080–1084.

Darden, S. K., Dabelsteen, T., and Pedersen, S. B. (2003). A potential tool for swift fox (*Vulpes velox*) conservation: individuality of long-range barking sequences. *J. Mammal.* 84, 1417–1427. doi: 10.1644/BEM-031

Darras, K., Furnas, B., Fitriawan, I., Mulyani, Y., and Tscharntke, T. (2018). Estimating bird detection distances in sound recordings for standardizing detection ranges and distance sampling. *Methods Ecol. Evol.* 9, 1928–1938. doi: 10.1111/2041-210X.13031

Darras, K., Pütz, P., Rembold, K., and Tscharntke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biol. Conserv.* 201, 29–37. doi: 10.1016/j.biocon.2016.06.021

Davy, M., and Godsill, S. (2002). Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing Orlando, FL: IEEE. 1313–1316.

Deichmann, J. L., Acevedo-Charry, O., Barclay, L., Burivalova, Z., Campos-Cerqueira, M., d'Horta, F., et al. (2018). It's time to listen: there is much to be learned from the sounds of tropical ecosystems. *Biotropica* 50, 713–718. doi: 10.1111/btp.12593

Delacourt, P., and Wellekens, C. J. (2000). DISTBIC: a speaker-based segmentation for audio data indexing. *Speech Comm.* 32, 111–126. doi: 10.1016/S0167-6393(00)00027-3

Dias, F. F., Pedrini, H., and Minghim, R. (2021). Soundscape segregation based on visual analysis and discriminating features. *Eco. Inform.* 61:101184. doi: 10.1016/j.ecoinf.2020.101184

Dueck, D. (2009). Affinity propagation: Clustering data by passing messages. Toronto, ON, Canada: University of Toronto, 144.

Dufourq, E., Batist, C., Foquet, R., and Durbach, I. (2022). Passive acoustic monitoring of animal populations with transfer learning. *Eco. Inform.* 70:101688. doi: 10.1016/j.ecoinf.2022.101688

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., et al. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sens. Ecol. Conserv.* 7, 475–487. doi: 10.1002/rse2.201

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., et al. (2020). Automated detection of Hainan gibbon calls for passive acoustic monitoring. doi: 10.5281/zenodo.3991714.

Favaro, L., Gili, C., Da Rugna, C., Gnone, G., Fissore, C., Sanchez, D., et al. (2016). Vocal individuality and species divergence in the contact calls of banded penguins. *Behav. Process.* 128, 83–88. doi: 10.1016/j.beproc.2016.04.010

Feng, J.-J., Cui, L.-W., Ma, C.-Y., Fei, H.-L., and Fan, P.-F. (2014). Individuality and stability in male songs of cao vit gibbons (*Nomascus nasutus*) with potential to monitor population dynamics. *PLoS One* 9:e96317. doi: 10.1371/journal.pone.0096317

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/S1464793101005826

Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., et al. (2017). Audio set: an ontology and human-labeled dataset for audio events. in 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), Piscataway, NJ IEEE 776–780.

Gibb, R., Browning, E., Glover-Kapfer, P., and Jones, K. E. (2018). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10, 169–185. doi: 10.1111/2041-210X.13101

Gillam, E. H., and Chaverri, G. (2012). Strong individual signatures and weaker group signatures in contact calls of Spix's disc-winged bat, *Thyroptera tricolor*. *Anim. Behav.* 83, 269–276. doi: 10.1016/j.anbehav.2011.11.002

Grolemund, G., and Wickham, H. (2011). Dates and times made easy with lubridate. *J. Stat. Softw.* 40, 1–25. doi: 10.18637/jss.v040.i03

Hafner, S. D., and Katz, J. (2018). monitoR: acoustic template detection in R. Available at: http://www.uvm.edu/rsenr/vtcfwru/R/?Page=monitoR/monitoR.htm (Accessed January 23, 2023).

Haimoff, E., and Gittins, S. (1985). Individuality in the songs of wild agile gibbons (*Hylobates agilis*) of Peninsular Malaysia. *Am. J. Primatol.* 8, 239–247. doi: 10.1002/ajp.1350080306

Haimoff, E., and Tilson, R. (1985). Individuality in the female songs of wild kloss' gibbons (*Hylobates klossii*) on Siberut Island, Indonesia. *Folia Primatol.* 44, 129–137. doi: 10.1159/000156207

Hamard, M., Cheyne, S. M., and Nijman, V. (2010). Vegetation correlates of gibbon density in the peat-swamp forest of the Sabangau catchment, Central Kalimantan, Indonesia. *Am. J. Primatol.* 72, 607–616. doi: 10.1002/ajp.20815

Han, W., Chan, C.-F., Choy, C.-S., and Pun, K.-P. (2006). An efficient MFCC extraction method in speech recognition. in 2006 IEEE International Symposium on Circuits and Systems, Piscataway, NJ IEEE 4.

Hanya, G., and Bernard, H. (2021). Interspecific encounters among diurnal primates in Danum Valley, Borneo. *Int. J. Primatol.* 42, 442–462. doi: 10.1007/s10764-021-00211-9

Heath, B. E., Sethi, S. S., Orme, C. D. L., Ewers, R. M., and Picinali, L. (2021). How index selection, compression, and recording schedule impact the description of ecological soundscapes. *Ecol. Evol.* 11, 13206–13217. doi: 10.1002/ece3.8042

Heinicke, S., Kalan, A. K., Wagner, O. J. J., Mundry, R., Lukashevich, H., and Kühl, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods Ecol. Evol.* 6, 753–763. doi: 10.1111/2041-210X.12384

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., and Rogers, A. (2018). AudioMoth: evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods Ecol. Evol.* 9, 1199–1211. doi: 10.1111/2041-210X.12955

Hodnett, M., Wiley, J. F., Liu, Y. H., and Maldonado, P. (2019). *Deep Learning with R for Beginners: Design Neural Network Models in R 3.5 Using TensorFlow, Keras, and MXNet*. Birmingham Packt Publishing Ltd.

Huancapaza Hilasaca, L. M., Gaspar, L. P., Ribeiro, M. C., and Minghim, R. (2021). Visualization and categorization of ecological acoustic events based on discriminant features. *Ecol. Indic.* 126:107316. doi: 10.1016/j.ecolind.2020.107316

Inoue, Y., Sinun, W., and Okanoya, K. (2016). Activity budget, travel distance, sleeping time, height of activity and travel order of wild east Bornean Grey gibbons (*Hylobates funereus*) in Danum Valley conservation area. *Raff. Bull. Zool.* 64, 127–138.

IUCN (2022). *The IUCN Red List of Threatened Species. The IUCN Red List of Threatened Species*. Available at: https://www.iucnredlist.org ().

Kahl, S., Wood, C. M., Eibl, M., and Klinck, H. (2021). BirdNET: a deep learning solution for avian diversity monitoring. *Eco. Inform.* 61:101236. doi: 10.1016/j.ecoinf.2021.101236

Kalan, A. K., Mundry, R., Wagner, O. J. J., Heinicke, S., Boesch, C., and Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecol. Indic.* 54, 217–226. doi: 10.1016/j.ecolind.2015.02.023

Kalan, A. K., Piel, A. K., Mundry, R., Wittig, R. M., Boesch, C., and Kühl, H. S. (2016). Passive acoustic monitoring reveals group ranging and territory use: a case study of wild chimpanzees (pan troglodytes). *Front. Zool.* 13:34. doi: 10.1186/s12983-016-0167-8

Katz, J., Hafner, S. D., and Donovan, T. (2016a). Assessment of error rates in acoustic monitoring with the R package monitoR. *Bioacoustics* 25, 177–196. doi: 10.1080/09524622.2015.1133320

Katz, J., Hafner, S. D., and Donovan, T. (2016b). Tools for automated acoustic monitoring within the R package monitoR. *Bioacoustics* 25, 197–210. doi: 10.1080/09524622.2016.1138415

Keen, S. C., Shiu, Y., Wrege, P. H., and Rowland, E. D. (2017). Automated detection of low-frequency rumbles of forest elephants: a critical tool for their conservation. *J. Acoust. Soc. Am.* 141, 2715–2726. doi: 10.1121/1.4979476

Kershenbaum, A., Sayigh, L. S., and Janik, V. M. (2013). The encoding of individual identity in dolphin signature whistles: how much information is needed? *PLoS One* 8:e77671. doi: 10.1371/journal.pone.0077671

Kidney, D., Rawson, B. M., Borchers, D. L., Stevenson, B. C., Marques, T. A., and Thomas, L. (2016). An efficient acoustic density estimation method with human detectors applied to gibbons in Cambodia. *PLoS One* 11:e0155066. doi: 10.1371/journal.pone.0155066

Koch, R., Raymond, M., Wrege, P., and Klinck, H. (2016). SWIFT: a small, low-cost acoustic recorder for terrestrial wildlife monitoring applications. in North American Ornithological Conference Washington, DC, 619.

Konopka, T. (2020). Umap: uniform manifold approximation and projection. Available at: https://cran.r-project.org/package=umap (Accessed January 23, 2023).

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., and Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere* 10:e02567. doi: 10.1002/ecs2.2567

Lapp, S., Wu, T., Richards-Zawacki, C., Voyles, J., Rodriguez, K. M., Shamon, H., et al. (2021). Automated detection of frog calls and choruses by pulse repetition rate. *Conserv. Biol.* 35, 1659–1668. doi: 10.1111/cobi.13718

Lawlor, J., Banville, F., Forero-Muñoz, N.-R., Hébert, K., Martínez-Lanfranco, J. A., Rogy, P., et al. (2022). Ten simple rules for teaching yourself R. *PLoS Comput. Biol.* 18:e1010372. doi: 10.1371/journal.pcbi.1010372

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

Ligges, U., Krey, S., Mersmann, O., and Schnackenberg, S. (2016). {tuneR}: analysis of music. Available at: https://r-forge.r-project.org/projects/tuner/

Lu, L., Zhang, H.-J., and Li, S. Z. (2003). Content-based audio classification and segmentation by using support vector machines. *Multimed. Syst.* 8, 482–492. doi: 10.1007/s00530-002-0065-0

Lucio, D. R., Maldonado, Y., and da Costa, G. (2015). Bird species classification using spectrograms. in 2015 Latin American Computing Conference (CLEI), Arequipa, Peru 1–11.

Madhusudhana, S. (2021). shyamblast/Koogu. Zenodo. doi: 10.5281/zenodo.5781423.

Madhusudhana, S., Shiu, Y., Klinck, H., Fleishman, E., Liu, X., Nosal, E.-M., et al. (2021). Improve automatic detection of animal call sequences with temporal context. *J. R. Soc. Interface* 18:20210297. doi: 10.1098/rsif.2021.0297

Madhusudhana, S. K., Symes, L. B., and Klinck, H. (2019). A deep convolutional neural network based classifier for passive acoustic monitoring of neotropical katydids. *J. Acoust. Soc. Am.* 146:2982. doi: 10.1121/1.5137323

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). Cluster: cluster analysis basics and extensions.

Markolf, M., Zinowsky, M., Keller, J. K., Borys, J., Cillov, A., and Schülke, O. (2022). Toward passive acoustic monitoring of lemurs: using an affordable open-source system to monitor Phaner vocal activity and density. *Int. J. Primatol.* 43, 409–433. doi: 10.1007/s10764-022-00285-z

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., et al. (2013). Estimating animal population density using passive acoustics. *Biol. Rev.* 88, 287–309. doi: 10.1111/brv.12001

Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. (1997). The DET curve in assessment of detection task performance. *Proc. Eurospeech* 4, 1895–1898.

Mellinger, D. K., Roch, M. A., Nosal, E.-M., and Klinck, H. (2016). "Signal processing" in *Listening in the Ocean*. eds. W. W. Au and M. O. Lammers (Berlin: Springer), 359–409.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2017). e1071: Misc functions of the Department of Statistics.

Mielke, A., and Zuberbühler, K. (2013). A method for automated individual, species and call type recognition in free-ranging animals. *Anim. Behav.* 86, 475–482. doi: 10.1016/j.anbehav.2013.04.017

Mitani, J. C. (1984). The behavioral regulation of monogamy in gibbons (*Hylobates muelleri*). *Behav. Ecol. Sociobiol.* 15, 225–229. doi: 10.1007/BF00292979

Mitani, J. C. (1985). Gibbon song duets and intergroup spacing. *Behaviour* 92, 59–96. doi: 10.1080/0141192032000137321

Muda, L., Begam, M., and Elamvazuthi, I. (2010). Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *J. Comput.* 2, 2151–9617. doi: 10.48550/arXiv.1003.4083

Noviyanti, A., Sudarsono, A. S., and Kusumaningrum, D. (2019). Urban soundscape prediction based on acoustic ecology and MFCC parameters. *AIP Conf. Proc.* 2187:050005. doi: 10.1063/1.5138335

Parra-Hernández, R. M., Posada-Quintero, J. I., Acevedo-Charry, O., and Posada-Quintero, H. F. (2020). Uniform manifold approximation and projection for clustering taxa through vocalizations in a neotropical passerine (rough-legged tyrannulet, *Phyllomyias burmeisteri*). *Animals* 10:1406. doi: 10.3390/ani10081406

Pérez-Granados, C., and Schuchmann, K.-L. (2021). Passive acoustic monitoring of the diel and annual vocal behavior of the Black and Gold Howler Monkey. *Am. J. Primatol.* 83:e23241. doi: 10.1002/ajp.23241

Phoonjampa, R., Koenig, A., Brockelman, W. Y., Borries, C., Gale, G. A., Carroll, J. P., et al. (2011). Pileated gibbon density in relation to habitat characteristics and post-logging forest recovery. *Biotropica* 43, 619–627. doi: 10.1111/j.1744-7429.2010.00743.x

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdiscipl. Rev.* 9:e1301. doi: 10.1002/widm.1301

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (Accessed January 23, 2023).

Romero-Mujalli, D., Bergmann, T., Zimmermann, A., and Scheumann, M. (2021). Utilizing DeepSqueak for automatic detection and classification of mammalian vocalizations: a case study on primate vocalizations. *Sci. Rep.* 11:24463. doi: 10.1038/s41598-021-03941-1

Ross, J. C. (2013). Flightcallr: Classify Night Flight Calls Based on Acoustic Measurements. Available at: https://R-Forge.R-project.org/projects/flightcallr/ (Accessed January 23, 2023).

Ross, J. C., and Allen, P. E. (2014). Random Forest for improved analysis efficiency in passive acoustic monitoring. *Eco. Inform.* 21, 34–39. doi: 10.1016/j.ecoinf.2013.12.002

Ruff, Z. J., Lesmeister, D. B., Appel, C. L., and Sullivan, C. M. (2020). Convolutional neural network and R-Shiny app for identifying vocalizations in Pacific Northwest forests. doi: 10.5281/zenodo.4092393.

Ruff, Z. J., Lesmeister, D. B., Appel, C. L., and Sullivan, C. M. (2021). Workflow and convolutional neural network for automated identification of animal sounds. *Ecol. Indic.* 124:107419. doi: 10.1016/j.ecolind.2021.107419

Sadhukhan, S., Root-Gutteridge, H., and Habib, B. (2021). Identifying unknown Indian wolves by their distinctive howls: its potential as a non-invasive survey method. *Sci. Rep.* 11:7309. doi: 10.1038/s41598-021-86718-w

Sainburg, T., Thielk, M., and Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Comput. Biol.* 16:e1008228. doi: 10.1371/journal.pcbi.1008228

Scavetta, R. J., and Angelov, B. (2021). *Python and R for the Modern Data Scientist*. California, USA: O'Reilly Media, Inc.

Sethi, S. S., Ewers, R. M., Jones, N. S., Orme, C. D. L., and Picinali, L. (2018). Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device. *Methods Ecol. Evol.* 9, 2383–2387. doi: 10.1111/2041-210X.13089

Sethi, S. S., Ewers, R. M., Jones, N. S., Sleutel, J., Shabrani, A., Zulkifli, N., et al. (2022). Soundscapes predict species occurrence in tropical forests. *Oikos* 2022:e08525. doi: 10.1111/oik.08525

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., et al. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proc. Natl. Acad. Sci.* 117, 17049–17055. doi: 10.1073/pnas.2004702117

Shiu, Y., Palmer, K. J., Roch, M. A., Fleishman, E., Liu, X., Nosal, E.-M., et al. (2020). Deep neural networks for automated detection of marine mammal species. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-020-57549-y

Shonfield, J., and Bayne, E. M. (2017). Autonomous recording units in avian ecological research: current use and future applications. *Avian Conserv. Ecol.* 12:art14. doi: 10.5751/ACE-00974-120114

Silva, B. (2022). soundClass: Sound Classification Using Convolutional Neural Networks. Available at: https://CRAN.R-project.org/package=soundClass (Accessed January 23, 2023).

Silva, B., Mestre, F., Barreiro, S., Alves, P. J., and Herrera, J. M. (2022). soundClass: an automatic sound classification tool for biodiversity monitoring using machine learning. *Methods Ecol. Evolut.* 13, 2356–2362. doi: 10.1111/2041-210X.13964

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., Sing, T., Sander, O., et al. (2005). ROCR: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941. doi: 10.1093/bioinformatics/bti623

Soares, B. S., Luz, J. S., de Macêdo, V. F., Silva, R. R. V. E., De Araújo, F. H. D., and Magalhães, D. M. V. (2022). MFCC-based descriptor for bee queen presence detection. *Expert Syst. Appl.* 201:117104. doi: 10.1016/j.eswa.2022.117104

Spillmann, B., van Schaik, C. P., Setia, T. M., and Sadjadi, S. O. (2017). Who shall I say is calling? Validation of a caller recognition procedure in Bornean flanged male orangutan (*Pongo pygmaeus*) long calls. *Bioacoustics* 26, 109–120. doi: 10.1080/09524622.2016.1216802

Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M., and Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array. *Methods Ecol. Evol.* 6, 38–48. doi: 10.1111/2041-210X.12291

Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* 10:e13152. doi: 10.7717/peerj.13152

Sueur, J., Aubin, T., and Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* 18, 213–226. doi: 10.1080/09524622.2008.9753600

Sugai, L. S. M., and Llusia, D. (2019). Bioacoustic time capsules: using acoustic monitoring to document biodiversity. *Ecol. Indic.* 99, 149–152. doi: 10.1016/j.ecolind.2018.12.021

Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., and Llusia, D. (2019). Terrestrial passive acoustic monitoring: review and perspectives. *Bioscience* 69, 15–25. doi: 10.1093/biosci/biy147

Sun, G.-Z., Huang, B., Guan, Z.-H., Geissmann, T., and Jiang, X.-L. (2011). Individuality in male songs of wild black crested gibbons (*Nomascus concolor*). *Am. J. Primatol.* 73, 431–438. doi: 10.1002/ajp.20917

Swets, J. A. (1964). *Signal Detection and Recognition by Human Observers: Contemporary Readings*. New York, NY: Wiley.

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2015). Lar gibbon (*Hylobates lar*) great call reveals individual caller identity. *Am. J. Primatol.* 77, 811–821. doi: 10.1002/ajp.22406

Vu, T. T., and Tran, L. M. (2019). An application of autonomous recorders for gibbon monitoring. *Int. J. Primatol.* 40, 169–186. doi: 10.1007/s10764-018-0073-3

Vu, T. T., and Tran, D. V. (2020). Using autonomous recorders and bioacoustics to monitor the globally endangered wildlife in the Annamite mountain landscape: a case study with crested argus in Song Thanh Nature Reserve. *J. Nat. Conserv.* 56:125843. doi: 10.1016/j.jnc.2020.125843

Wäldchen, J., and Mäder, P. (2018). Machine learning for image based species identification. *Methods Ecol. Evol.* 9, 2216–2225. doi: 10.1111/2041-210X.13075

Walsh, R. P., and Newbery, D. M. (1999). The ecoclimatology of Danum, Sabah, in the context of the world's rainforest regions, with particular reference to dry periods and their impact. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 354, 1869–1883. doi: 10.1098/rstb.1999.0528

Wanelik, K. M., Azis, A., and Cheyne, S. M. (2012). Note- , phrase- and song-specific acoustic variables contributing to the individuality of male duet song in the Bornean southern gibbon (*Hylobates albibarbis*). *Primates* 54, 159–170. doi: 10.1007/s10329-012-0338-y

Wang, Y., Ye, J., and Borchers, D. L. (2022). Automated call detection for acoustic surveys with structured calls of varying length. *Methods Ecol. Evolut.* 13, 1552–1567. doi: 10.1111/2041-210X.13873

Wiggins, S. (2003). Autonomous acoustic recording packages (ARPs) for long-term monitoring of whale sounds. *Mar. Technol. Soc. J.* 37, 13–22. doi: 10.4031/002533203787537375

Wijers, M., Loveridge, A., Macdonald, D. W., and Markham, A. (2021). CARACAL: A versatile passive acoustic monitoring tool for wildlife research and conservation. *Bioacoustics* 30, 41–57.

Xuan, N., Julien, V., Wales, S., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* 11, 2837–2854.

Zeppelzauer, M., Hensman, S., and Stoeger, A. S. (2015). Towards an automated acoustic detection system for free-ranging elephants. *Bioacoustics* 24, 13–29. doi: 10.1080/09524622.2014.906321

Zottesso, R. H. D., Costa, Y. M. G., Bertolini, D., and Oliveira, L. E. S. (2018). Bird species identification using spectrogram and dissimilarity approach. *Eco. Inform.* 48, 187–197. doi: 10.1016/j.ecoinf.2018.08.007

Zwart, M. C., Baker, A., McGowan, P. J. K., and Whittingham, M. J. (2014). The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars. *PLoS One* 9:e102770. doi: 10.1371/journal.pone.0102770

frontiers | Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fevo.2023.1145205

Check for updates

OPEN ACCESS

EDITED BY
Patrice Adret,
Universidad Autónoma Gabriel René Moreno,
Bolivia

REVIEWED BY
Alba Garcia De La Chica,
University of Buenos Aires, Argentina
Valeria Torti,
University of Turin, Italy

*CORRESPONDENCE
Allison R. Lau
✉ alljones@ucdavis.edu

RECEIVED 15 January 2023
ACCEPTED 10 April 2023
PUBLISHED 25 May 2023

CITATION
Lau AR, Cunningham AD and Bales KL (2023) Pairing status and stimulus type predict responses to audio playbacks in female titi monkeys.
Front. Ecol. Evol. 11:1145205.
doi: 10.3389/fevo.2023.1145205

Pairing status and stimulus type predict responses to audio playbacks in female titi monkeys

Allison R. Lau[1,2]*, Ashley D. Cunningham[3] and Karen L. Bales[1,2,3,4]

[1]Graduate Group in Animal Behavior, University of California, Davis, Davis, CA, United States, [2]California National Primate Research Center, University of California, Davis, Davis, CA, United States, [3]Department of Psychology, University of California, Davis, Davis, CA, United States, [4]Department of Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, CA, United States

Some paired primates use complex, coordinated vocal signals to communicate within and between family groups. The information encoded within those signals is not well understood, nor is the intricacy of individuals' behavioral and physiological responses to these signals. Considering the conspicuous nature of these vocal signals, it is a priority to better understand paired primates' responses to conspecific calls. Pair-bonded titi monkeys (*Plecturocebus cupreus*) sing duets comprised of the male and female's long call. Here, we use a playback study to assess female titi monkeys' responses to different vocal stimuli based on the subject's pairing status. Six adult female titi monkeys participated in the study at two timepoints—pre-pairing and post-pairing. At each timepoint, subjects underwent three distinct playbacks—control recording, male solo vocalization, and pair duet. Behaviors such as locomotion and vocalizations were scored during and after the playback, and cortisol and androgen values were assessed via a plasma blood sample. Female titi monkeys attended more to social signals compared to the control, regardless of pairing status. However, in the time immediately following any playback type, female titi monkeys trilled more and spent a greater proportion of time locomoting during pre-pairing timepoints (compared to post-pairing). Female titi monkeys' behavioral responses to social audio stimuli, combined with subjects' increases in cortisol and androgens as paired individuals, imply female titi monkeys attend and respond to social signals territorially.

KEYWORDS

long call, duet, pair bonding, hormones, titi monkey

Introduction

Many social animals use vocal signals to communicate with conspecifics (Silk, 2007). Frequently, research studies quantify the variation of and identify the mechanisms underlying social vocal signals (Fishbein et al., 2021). However, listeners' perception and interpretation of these vocal signals represents fertile grounds for additional study to illuminate what aspects of vocal variation are meaningful to conspecifics. Primates across the order frequently use vocal communication and display wide variety in call structure, meaning, and function (Bradbury and Vehrencamp, 1998, pp. 658–665). Combined with contextual knowledge, such as relationship with the caller or listener, vocal variation can be understood by conspecifics and can simplify social interactions by, for instance, communicating the motivational state (i.e., aggressiveness or passiveness) of another (Cheney and Seyfarth, 2018). Generally, primates have a highly adaptive ability to produce vocal variation within a wide range of social situations (Cheney and Seyfarth, 2018).

Playback studies—in which researchers broadcast pre-recorded vocal signals—have been used to identify what information vocalizations signal to conspecifics based upon the behavioral responses of listening individuals (Fischer et al., 2013). While playback studies have been used across the animal kingdom, we focus here on the non-human primate literature, to better place our study in context. For example, calls can encode information about physiological and emotional states during social conflict (e.g., male chacma baboons (*Papio ursinus*) and chimpanzees (*Pan troglodytes*; Fischer et al., 2013). Many species use calls to evoke intended behavioral responses from group members, such as red-fronted lemurs (*Eulemur fulvus rufus*) and white sifakas (*Propithecus verreauxi*) using alarm calls to initiate group fleeing behavior in the presence of predators (Fischer et al., 2013). In wild rhesus macaques (*Macaca mulatta*), wild Japanese macaques (*Macaca fuscata*), captive cotton-top tamarins (*Saguinus oedipus*), and captive coppery titi monkeys (*Plecturocebus cupreus*) vocalizations can provide cues about sex, reproductive status, and group membership (Ghazanfar and Hauser, 2001; Clink et al., 2019). Female rhesus macaques use vocalizations to distinguish kin from non-kin, as shown by the greater amount of time spent orienting toward vocalizations made by kin (Ghazanfar and Hauser, 2001). Overall, playback studies demonstrate the wide range of information that can be communicated via vocal signals in primates.

Primates live in a wide range of social groupings, spanning from solitary to multi-level fission/fusion societies (Terborgh and Janson, 1986; Kappeler and van Schaik, 2002). These varying social groups have emerged across primate evolution due to a combination of ecological, social, and physiological pressures that provide rich opportunity to investigate sociality (Kappeler and van Schaik, 2002; Dunbar and Shultz, 2021). Amongst this wide variety, pair living remains one of the least common social groupings in mammals, especially primates (Kleiman, 1977; Lukas and Clutton-Brock, 2013), and one that necessitates unique communication (Singletary and Tecot, 2020). Some or all species of gibbons (Marshall and Marshall, 1976; Palombit, 1994; Geissmann, 2002), siamangs (Geissmann, 2002), titi monkeys (Robinson, 1979a,b), owl monkeys (Depeine et al., 2008), and tarsiers (Nietsch, 1999; Clink et al., 2022) live in adult female/male pairs and communicate using specialized vocal signals (Singletary and Tecot, 2020). Given the rarity of this social organization, it is of considerable importance to better understand the communication processes reinforcing pair living.

Playback studies have been used to investigate the meaning of vocal signals in some pair-living primates (Robinson, 1981; Fichtel and Hilgartner, 2013; Caselli et al., 2015; Garcia de la Chica et al., 2021). For example, researchers played back calls of unknown, single owl monkeys (*Aotus azarae*) to owl monkey pairs and found the location of the playback did not influence resulting behaviors, but paired owl monkeys—both the adult male and female—reacted more to unfamiliar male calls than female calls with greater movement toward the playback and more vocalizations, revealing the tendency for both mates to defend their partner (Garcia de la Chica et al., 2021). Beyond owl monkeys, the remaining pair-living primates (e.g., gibbons, siamangs, titi monkeys, and tarsiers) all participate in highly coordinated vocal interactions—often called duets—in which the adult female and male emit sex-specific vocal contributions (Marshall and Marshall, 1976; Robinson, 1979a,b; Nietsch, 1999; De Gregorio et al., 2022). However, the fine-scale social behaviors of these species are difficult to study in the wild (Bossuyt, 2002; Caselli et al., 2014, 2015).

Copious studies of titi monkeys—both in the wild and in captivity—have illuminated the strong and selective pair bonds that mated adult titi monkeys (*Plecturocebus* spp.) form with each other (Bales et al., 2017). Pair bonds are enduring socio-emotional attachments characterized by a suite of behaviors including preference for one's mate over an opposite-sex stranger, proximity maintenance, and separation distress; for an extensive definition and review, see Bales et al. (2021). Titi monkeys duet every morning in species-typical, stereotyped duets, communicating social information both with their mate and with conspecifics. Functionally, titi monkey's duets serve as territorial signals, allowing groups to claim occupancy and reinforce boundaries when threatened (Robinson, 1979b, 1981). Titi monkeys approach neighboring groups when conspecifics duet near territorial boundaries, providing evidence of joint territorial defense by the adult female and male (*Plecturocebus cupreus*; Robinson, 1979b, 1981; *Callicebus nigrifrons*; Caselli et al., 2014, 2015). Titi monkeys mate guard and display agonistic behaviors toward strangers (Fernandez-Duque et al., 2000). Coppery titi monkey (*Plecturocebus cupreus*; previously classified as *Callicebus moloch,* then *Callicebus cupreus,* and ultimately *Plecturocebus cupreus* and hereafter referred to as "titi monkeys") pairs duet together as early as the first day of pairing (Müller and Anzenberger, 2002). The aforementioned behaviors—territorial defense, mate guarding, agonism toward strangers, and vocal duetting—represent a suite of behaviors commonly attributed to titi monkeys' general territorial defense (Robinson, 1981; Caselli et al., 2015; Mercier et al., 2020).

Titi monkeys' expansive vocal repertoires have been studied in multiple species and in multiple contexts (observation and experimentation, both in wild and captive settings). Trills are often used in the context of separation wherein individuals cannot access a group member (Moynihan, 1966). Additionally, infants most commonly use trills as a generalized vocalization that elicits care from parents (Robinson, 1979a; Lau et al., 2020). Peeps are commonly used as a proximity-seeking call and are used by subadult and adult individuals (Arias del Razo et al., 2022a). Titi monkey long calls (referred to as a "long call" when vocalized alone, but a "duet contribution" when two titi monkeys coordinate their long calls to form a "duet") are one of the most conspicuous and species-typical behaviors performed by titi monkeys. Titi monkey duets, in particular, are understudied both in captivity and in the wild. What is known about this particularly conspicuous behavior indicates wide variation in titi monkey duet features and the potential for these duets to carry information about caller identity. In the wild, titi monkey duets were longer in duration during inter-group encounters as opposed to spontaneous duets (Dolotovskaya and Heymann, 2022). Adult male and female's contributions to the duet are individually identifiable (Lau et al., 2020). However, cross-sectional evidence shows that titi duet contributions do have a degree of plasticity, as pair mates converge with their partner in their note rate over time (Clink et al., 2019). Within that plasticity, there are also limited impacts of heritability in titi monkey's duet contributions (Clink et al., 2022). While most studies of coppery titi monkey vocal variance have occurred in captivity, audibly and visually, titi monkey duets from captivity are indistinguishable from those in the wild (Robinson, 1979a; Lau et al., 2020). This previous work demonstrates the ability of titi monkey calls to carry a wide range of information that is both statistically identifiable and, to some extent, behaviorally relevant to titi monkeys. The present study aims to investigate another element of titi monkey communication by observing responses to social playbacks in a controlled, captive setting. This project serves the secondary function of validating the retention and use of previously observed wild titi monkey behavior in a captive population for the first time.

Physiologically, multiple hormones may be involved in social and territorial behaviors in this species. Cortisol, a steroid hormone produced by the adrenal gland, plays an important role in social behavior. In closely related, pair-living owl monkeys (*Aotus azarae*), females' cortisol levels are high during gestation, and both males and females have lower cortisol during periods of intensive infant care (Corley et al., 2021). The activational effects of cortisol serve a variety of functions and the nuance of cortisol's affects has recently reframed the importance of interpreting cortisol results with full consideration for the context in which it is investigated (Epel et al., 2018). Titi monkeys display robust responses to dexamethasone challenge of the adrenocortical system (Mendoza and Moberg, 1985). Additionally, titi monkeys are quite responsive to novelty in that titi monkeys require far less novelty than closely related squirrel monkeys (*Saimiri sciureus*) to evoke a cortisol response (Hennessy et al., 1995). Titi monkey infants show increased cortisol when separated from their parents (Hoffman et al., 1995) and in adulthood, titi monkeys' cortisol levels are higher when separated from one's mate compared to a non-separation period of identical duration (Arias del Razo et al., 2022a). In addition to the impacts of cortisol, androgens in titi monkeys have been studied in the social contexts of puberty (Arias del Razo et al., 2020), jealousy (Maninger et al., 2017), and, importantly for the present study, separation from one's pair mate (Arias del Razo et al., 2022a). The previous research of the role of both cortisol and androgens in titi monkey social behavior indicates that many titi monkey social situations will likely involve activation of adrenocortical and androgen systems. Given what little is known about the role of cortisol and androgens in the social vocalizations of this species, we aimed to investigate the physiological impacts of social vocal communication within the present study.

While previous work has identified the variance and function of titi monkey duet contributions, few studies to date have assessed how titi monkeys' behavioral and physiological responses to vocal stimuli vary. In black-fronted titi monkeys, three pairs responded with vocal and approach responses to all conspecific playback stimuli (male solo, female solo, and duet), but not to the control (Caselli et al., 2015). Additionally, black-fronted titi monkeys did not respond differentially to solos and duets and appeared to use a joint territorial defense approach in responding to any conspecific vocalizations (Caselli et al., 2015). While we expect similar behavioral responses in coppery titi monkeys, the present study expands on Caselli and colleagues' previous work and expands our knowledge into the captive setting. Due to funding limitations and the focus on female pair-bonding behavior in the Bales Laboratory, we chose to focus our attention on female titi monkeys for this project. In this study, we assessed titi monkey females' responses to unfamiliar male solo vocalizations and duets of unfamiliar female/male pairs. Titi monkeys' social behaviors are fairly subtle, and individuals display species-wide neophobia—both of which make studying intricate social responses in the field quite challenging (Bossuyt, 2002). To date, no study has looked at titi monkeys' responses to playbacks in captivity. It is unknown what information titi monkeys receive when listening to social vocalizations. For this study, we utilized the breeding colony of coppery titi monkeys (*Plecturocebus cupreus*) at the California National Primate Research Center (CNPRC) in Davis, California. The CNPRC facility allowed us to perform playback studies with experimental control and fine-scale observations that are impossible in the field. We chose to focus on female titi monkeys based upon limited resources and the unique role of the female titi monkey in parenting and maintaining pair proximity (Dolotovskaya et al., 2020b).

While this study was inherently investigatory in nature, we did pose a hypothesis and a few corresponding predictions prior to the study. First, we hypothesized that captive female titi monkeys' behavioral and physiological states when hearing playbacks reflect the known territorial responses of titis based upon pairing status. We predicted female titi monkeys would have higher cortisol levels during social stimuli playbacks compared to the control, regardless of pairing status. We also predicted female titi monkeys would have higher androgens during post-pairing duet stimulus playback and solo stimulus playback compared to the post-pairing control playback and all pre-pairing stimuli. Behaviorally, we predicted a greater number of vocalizations, locomotion, and time spent orienting to the stimuli in response to duet and solo playbacks compared to the control, regardless of pairing status. We expected all behavioral and physiological changes in social playback responses compared to the control playback to be higher for duet playbacks than solo playbacks.

Methods

Subjects

All coppery titi monkeys (*Plecturocebus cupreus*) used for this project were captive born at the CNPRC. The titi monkeys were housed indoors in stainless steel enclosures measuring 1.2 × 1.2 × 2.1 m (volume = 3.024 m^3) or 1.2 × 1.2 × 1.8 m (volume = 2.592 m^3). Cage height depended upon their location in the CNPRC. All rooms were maintained at 21°C on a 12-h light cycle with lights on from 06:00 to 18:00. Subjects were fed a diet of monkey chow, carrots, bananas, apples, and rice cereal twice a day. Subjects were offered one Spanish peanut during daily health checks as a reward for presenting their abdomen and digits for inspection. Water was available *ad libitum* and additional oat foraging enrichment was provided twice daily. Subjects were housed in natal family groups that varied in composition during the pre-pairing portion of the study. During the post-pairing portion of the study, subjects lived in female/male pairs. All groups were in acoustic contact with other titi monkey pairs both within their room and with animals in other rooms but had minimal visual contact with animals outside their own cage. This housing situation is the same as described in previous studies of this colony (Mendoza and Mason, 1986a; Tardif et al., 2006).

For this study, we chose our focal subjects (n = 6 females) from available, unpaired females living with either one parent, one same-sex sibling, or both parents and a sibling. At the start of the playback study, females ages ranged from 1.89–3.64 years of age, for a mean ± SD age of 2.64 ± 0.74 years.

Study design

Testing occurred at two testing timepoints: once approximately 1 month before the focal subject was paired with their pair mate, and again approximately 8 months post-pairing (Figure 1). The post-pairing timepoint was originally scheduled at 6 months post-pairing because 4–6 months post-pairing is the timeframe in which a strong behavioral preference, as well as associated neurological changes, are displayed towards one's partner (Rothwell et al., 2020; Arias del Razo et al., 2022b). However, COVID-19 pandemic-related delays necessitated the delay of the post-pairing sessions. As such, females were tested 8 months post-pairing.

Prior to testing, we relocated females to a private testing room to eliminate acoustic contact with other animals. Female subjects were accompanied to the testing room with either their family (pre-pairing timepoint) or their pair mate (post-pairing timepoint). Following the relocation, we gave subjects two undisturbed days to habituate to the testing room (Bales et al., 2017). Following the habituation period, testing occurred over three days during which the three playback tests were counter-balanced across subjects: (1) control recording of monkey room ambient acoustics (no animal vocalizations), (2) unfamiliar, unpaired male solo vocalization, (3) unfamiliar, paired male and female duet (Figure 1). We chose the control recording, which included the sound of hoses and the air conditioning system, as these sounds were familiar, but recorded in a room that did not have titi monkeys in it. Thus, this control recording represented a familiar sound type, but a novel recording of it. Similarly, we chose solo male recordings and duet recordings as all titi monkeys are familiar with these social signals. However, we chose recordings from unfamiliar animals to assess responses to the type of call, not the individual who was calling. We focused on male solo vocalizations and duets as they represent two distinct pairing statuses (unpaired and paired) and are signals that an opposite sex individual (our focal female subjects) pays specific attention to either when seeking a mate or defending their territory (Robinson, 1981).

Male solos and pair duets were recorded at the CNPRC. We confirmed stimuli were unknown to test subjects by ensuring recordings were either (1) recorded before the focal subject was born or (2) recorded in a room other than the room in which the focal subject lived. We used male solo recordings from monkeys who were an average age of 14.8 years old ± standard deviation of 4.3 years (range 11.7–20.9 years old). These male monkeys were not paired and living alone at the time of recording. They vocalized solos alone, without any other monkeys vocalizing. We did not edit the monkeys' calls. We used duet recordings from pairs that had been together for 6.6 ± 6.1 years (range 1.9–15.2 years). There is evidence that titi monkeys' duet contributions change with age—individuals' pulse note rate of repetition decreases with age and the total duration of the pulse note duet contribution increases in overall duration (Clink et al., 2019). Additionally, there is evidence that individuals' pulse duration decreases with increasing pair tenure, and that pair mates become more similar to each other in their pulse rate with pair tenure (Clink et al., 2019). As such, we aimed to capture a wide range of ages and pair tenure in the various stimulus recordings used. However, we were limited by the availability of solo males and the need to ensure stimulus individuals were strangers to our focal subjects.

All playbacks were broadcast at species-typical amplitude, measured by sound pressure level (SPL) meter (approximately 110 dB). We pre-recorded playback vocalizations from titi monkey individuals both unrelated to our subject females and from different housing

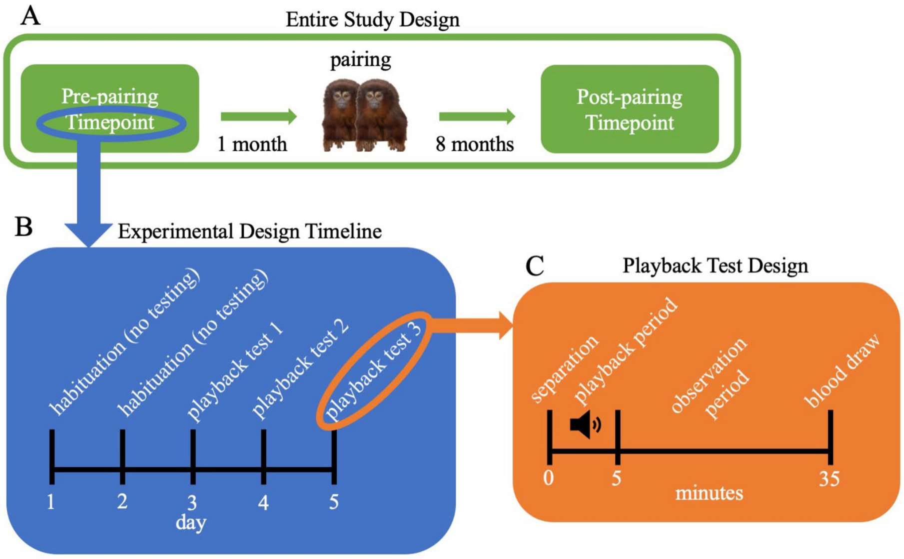


FIGURE 1
Schematic of the playback study design. **(A)** Entire Study Design: adult female titi monkey subjects underwent the pre-pairing playback timepoint while still living with their family. One month after the pre-pairing timepoint, female titi monkey subjects underwent pairing with an unfamiliar adult male. Eight months following pairing with one's mate, female subjects underwent the same playback study at the post-pairing timepoint. **(B)** Pairing Timepoint Design: during each testing timepoint, titi monkey subjects were given 2 days to habituate to the testing room, followed by 3 days of playback tests. **(C)** Playback Test Design: for each playback test, females were first separated from their cage mates (family or pair mate depending on the timepoint). The 5-min playback period then commenced, followed by a 30-min observation period of silence. At the end of the 35-min test, subjects were handled for a blood draw.

areas, ensuring the playback stimuli were unfamiliar. Each subject heard a unique exemplar for each playback type at each timepoint to avoid pseudo-replication and ensure appropriate sampling of the colony population (Kroodsma, 1989, 1990).

At the beginning of each test, we removed all family members (during the pre-pairing timepoint) or the pair mate (during the post-pairing timepoint) from the cage, leaving the female subject alone. We separated the focal subject from her home social setting to ensure all responses to the playback stimuli were individually driven and not impacted by the behavior of other animals in the group. While the separation paradigm itself does introduce a degree of social distress (Arias del Razo et al., 2022a), separation occurred before all playback stimulus types. As such, the control stimulus serves as the reference level for the solo and duet stimuli and provides a comparison with which to observe the behavioral and physiological impacts of social playbacks beyond the impacts of separation alone. The playback recording was broadcast for 5 min—the average duration of indoor-housed coppery titi monkey duets—followed by an additional 30-min observation period. In total, each playback test lasted 35 min. The methods for the separation paradigm and subsequent blood draw (detailed below) followed temporary separation protocols developed for previous projects in this lab with the addition of the acoustic playback (Figure 1; Arias del Razo et al., 2022a).

Female cycling

Our hormonal outcome measures, androgens and cortisol, can vary based on the levels of circulating estrogen and progesterone (Van Goozen et al., 1997). To assess reproductive status, we collected urine samples three times weekly during the first morning urination (0530–0600 h), with a maximal interval of 3 days between collection of successive samples for any given individual while our female monkeys were participating in this study. Urine sample collection began 2 weeks prior to the start of the playback study. We collected an average of 13.14 ± standard error of 0.88 samples per individual (range: 9–15). Following collection, samples were aliquoted into 2 ml cryo tubes and stored at −80°C until assay. Titi monkey reproductive cycles are, on average, 17 days long (Valeggia et al., 1999). As such, we assayed 1 month's worth of samples per subject for urinary estrogen (E1C) and pregnanediol (PdG) conjugates to identify reproductive cycling (or lack thereof). E1C and PdG were assayed at the UC Davis Endocrinology Laboratory using an enzyme-immunoassay described in detail elsewhere (Valeggia et al., 1999; Conley et al., 2022). Inter-assay Coefficients of Variation (CVs) were 0.88% and intra-assay CVs were 3.73%.

Ovulation was assumed to have occurred if PdG concentrations were > 100 ng/mg Cr in two consecutive samples that together totaled >400 ng/mg Cr, and were defined as luteal phases (Conley et al., 2022). Given recent evidence that female titi monkeys begin regularly reproductively cycling around 2.5 years of age but can have intermittent cycles earlier—and can begin cycling while in the natal family group or once paired—cycling information was included in all behavioral and physiological variables' initial models (Conley et al., 2022).

For inclusion in our models, females were coded either as non-cycling (based on urinary assay), cycling (based on urinary assay or a previous pregnancy), or pregnant (based on a positive ultrasound). The three reproductive statuses—non-cycling, cycling, or pregnant—were coded 0, 1, and 2, respectively.

Hormonal response to playbacks

All playback experiments occurred at the same time of day (1,330 h) to eliminate the potential confounding effects of circadian cortisol and androgen rhythms (Place and Nichols, 1991; Smith and French, 1997). At the end of each 35-min test, a 0.5 ml blood sample was collected via femoral venipuncture to assess androgen and cortisol levels. Samples were collected 41.10 ± 0.32 min (mean ± standard error) from the start of the audio playback and 3.52 ± 0.30 min from the start time of handling. Blood samples were immediately placed on ice and, within 5 min, centrifuged at 4°C for 15 min. We extracted plasma and stored samples at −80°C. Plasma cortisol and androgens were assayed at the UC Davis Endocrinology Laboratory using an enzyme immunoassay previously validated both chemically and biologically for titi monkeys and described in detail elsewhere (Witczak et al., 2021; Conley et al., 2022; Arias del Razo et al., 2022a). Inter-assay CV was 2.5% for cortisol (intra-assays CVs were 9.3 and 9.4% for the two plates) and the intra-assay CV was 13.6% for the single androgen plate. All hormone measures were natural log-transformed prior to all analyses so that the data met the assumptions of normality.

Behavioral scoring

Behavioral measures were recorded to assess the female's response to each playback type (Mendoza and Mason, 1986a; Fernandez-Duque et al., 2000) during two periods: (1) the playback period in which the females listened to an audio stimulus and (2) the observation period in which females were observed for 30 min immediately after the audio stimulus. We separated all behavioral analyses into these two periods (playback and observation) to illuminate the immediate and following impacts of the playback on behavior. The 35-min test was filmed to enable later behavioral scoring. The percent of time locomoting was scored from video recordings of each test using the DVRecorder module of Behavior Tracker.[1] Orientation to the stimulus was scored in real time using the Recorder module of Behavior Tracker (see footnote 1).

We recorded the subjects' vocalizations during and after the playback using a Marantz PND 660 recorder and a Marantz directional condenser microphone (Marantz, Kanagawa, Japan) to enable accurate classification of quickly repeated, intricate calls. After testing, calls were identified and scored from spectrograms using Raven Pro 1.6 Sound Analysis Software (K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2022, Ithaca, NY). We generated spectrograms with a 512-point (11.6 ms) Hann window (3 dB bandwidth = 124 Hz), with 75% overlap, and a 1,024-point discrete Fourier transform, resulting in time and frequency measurement precision of 2.9 ms and 43.1 Hz (Lau et al., 2020). We did not down-sample the original sound files. One observer (ARL) scored all the

---

[1] www.behaviortracker.com

subjects' vocal output, which included peeps, trills, alarm calls, long call introduction notes, long calls, and the latency to vocalize (Figure 2).

For the full ethogram and descriptive statistics of all outcome variables, see Table 1 and Table 2, respectively.

Model selection

We conducted all data analyses using R programming language and environment (R Core Team, 2022). We used backwards model selection of linear mixed effects models (*lmm* function) from the *nlme* package (Pinheiro, 2009; Pinheiro et al., 2017) to assess how pairing status (unpaired or paired), reproductive status (not cycling, cycling, or pregnant), and playback type (ambient control, male solo, or duet) predicted behavioral and physiological measures (cortisol, androgens, percent time orienting to the stimulus, percent time locomoting, peeps, and trills). Subject served as a random effect for all models due to known variability in titi monkey behavior (vocal behavior: Lau et al., 2020; pair affiliation: Rothwell et al., 2020; parenting behavior: Karaskiewicz et al., 2021).

We built each initial model with our three fixed effects and random effect included. As we worked through backwards model selection, we removed each fixed effect one at a time and compared each model to the initial model using the *anova* function (R Core Team, 2022). We used the resulting log likelihood ratio and p value to assess model fit, using a standard threshold of $p \leq 0.05$ as our criteria for retaining or excluding fixed effects.

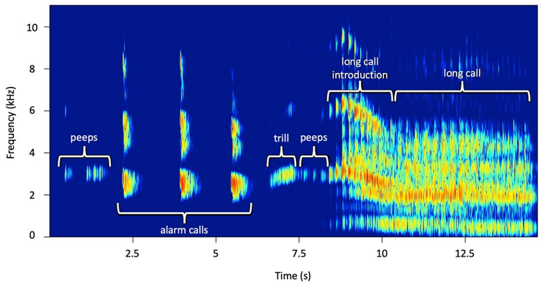


FIGURE 2
Representative spectrogram of female titi monkey vocal responses during the playback experiment. Peeps, alarm calls, trills, long call introductions, and long calls are displayed within one spectrogram.

Regardless of how much they contributed to each model, pairing status and playback type were retained as predictors in all final models to fully account for the experimental paradigm of the study. Reproductive status remained as a fixed effect in all models in which reproductive status contributed significantly to the final model. Given the known variation in titi monkey vocal behavior based upon female reproductive status (Dolotovskaya and Heymann, 2022), we retained reproductive status in all vocal behavior models. Regardless of the random effect's contribution to overall variance, we retained the random effect of subject in all models. We examined a quantile-quantile plot of the residuals of each final model to assess goodness of fit. We report the results of the final model for all behavioral and physiological outcome measures.

For the outcome variable latency to vocalize, we used a survival/event time model because one subject (during the post-pairing, solo stimulus playback) did not vocalize during entire the 35-min test. We fitted two Cox Proportional Hazards regression models using the *coxph* function of the *survival* library (Therneau, 2019). The first was a null model—the second model added fixed effects of reproductive status, stimulus, and pairing status. We used Akaike's Information Criterion to compare the second model to the first.

All figures presented below were created in R programming language and environment (R Core Team, 2022) using the *ggplot2* (Wickham et al., 2016) and *cowplot* (Wilke et al., 2019) packages.

Post-hoc comparisons

Following backwards model selection, we chose to run contrast comparisons to determine the difference between the three levels of predictor variables that had three levels (reproductive status and stimulus type) and that were statistically significant predictors in the respective model. We used the *glht* function from the *multcomp* package (Hothorn et al., 2016) to perform Tukey's Honest Significant Difference (HSD) test, allowing us to compare the means of each level for our three-level predictor variables.

TABLE 1 Ethogram of all behaviors scored during the playback study.

Behavior	Definition
% Time Locomoting	Time subject spends moving one or more limbs over the total duration of time, previously defined in(Arias del Razo et al., 2022a).
% Time Orienting to Stimulus	Time subject spends orienting to the audio stimulus over the total duration of time, previously defined in (Lau et al., 2021). Orientation required visual orientation toward the stimulus. The movement of the head required to shift attention from elsewhere to the stimulus did not necessarily involve locomotion of the limbs (as defined above).
Peeps	Single note, short duration vocalizations, previously defined in (Arias del Razo et al., 2022a).
Trills	Vocalization of modulating frequency emitted with one breath, previously defined in (Moynihan, 1966; Robinson, 1979a and Lau et al., 2020).
Long Call Intros	Vocalizations preceding a long call. Long call intros include a single high frequency note followed immediately by a low frequency note, previously defined in (Robinson, 1979a).
Long Calls	Vocalizations including high frequency (chirps and pulses) and low frequency components, previously defined in (Robinson, 1979a and Lau et al., 2020).
Alarm Calls	Large bandwidth, single note vocalizations, previously defined in (Cäsar et al., 2012). Alarm call structure varied by individual but most closely resembled call A and call B alarm calls uttered by Callicebus nigrifrons (Cäsar et al., 2012).
Latency to Vocalize	The time (sec) to first vocalization in response to the playback.

TABLE 2 List of behavioral and physiological measures collected during the playback study.

Variable	Scored via	Stimulus type	Mean±SD, N	Range	Stimulus type	Mean±SD, N	Range
Cortisol (ng cort/mL plasma)	Plasma	Control	960.17 ± 295.25, N = 6	522.2–1,307.0	Control	5,415.87 ± 5595.18, N = 6	794.6–15,001.6
		Solo	987.80 ± 358.86, N = 6	565.8–1,588.2	Solo	5268.17 ± 6031.98, N = 6	601.6–15,941.4
		Pair	847.80 ± 215.12, N = 6	573.8–1,177.6	Pair	5,679.43 ± 6197.84, N = 6	854.6–15,891.8
Testosterone (pg T/mL plasma)	Plasma	Control	284.33 ± 40.18, N = 6	144.7–820.4	Control	560.02 ± 226.71, N = 6	144.7–820.4
		Solo	235.36 ± 59.62, N = 5	279.4–887.8	Solo	592.43 ± 197.44, N = 6	279.4–887.8
		Pair	225.68 ± 45.16, N = 6	243.4–928.7	Pair	599.78 ± 239.23, N = 6	243.4–928.7
Latency to vocalize	Spectrogram	Control	68.56 ± 53.66, N = 6	23.1–169.2	Control	345.76 ± 634.01, N = 6	37.3–1632.3
		Solo	294.56 ± 267.9, N = 6	42.5–641.2	Solo	332.42 ± 315.91, N = 5	27.2–712.5
		Pair	958.97 ± 381.73, N = 6	37.7–1080.9	Pair	336.40 ± 377.61, N = 6	48.5–959.0

			Playback period		Observation period			Playback period		Observation period	
			Mean ± SD, N	Range	Mean ± SD, N	Range		Mean ± SD, N	Range	Mean ± SD, N	Range
% Time locomoting	Video	Control	0.32 ± 0.22, N = 6	0.08–0.64	0.31 ± 0.17, N = 6	0.17–0.58	Control	0.27 ± 0.29, N = 6	0.01–0.76	0.26 ± 0.23, N = 6	0.03–0.66
		Solo	0.35 ± 0.30, N = 6	0.0–0.76	0.28 ± 0.24, N = 6	0.0–0.67	Solo	0.13 ± 0.15, N = 6	0.00–0.37	0.15 ± 0.08, N = 6	0.02–0.23
		Pair	0.39 ± 0.18, N = 6	0.12–0.64	0.41 ± 0.18, N = 6	0.21–0.64	Pair	0.30 ± 0.25, N = 6	0.0–0.59	0.25 ± 0.17, N = 6	0.09–0.53
% Time orienting to stimulus	Live	Control	0.07 ± 0.07, N = 6	0.0–0.17	0.01 ± 0.02, N = 6	0.0–0.05	Control	0.02 ± 0.01, N = 6	0.0–0.03	0.01 ± 0.01, N = 6	0.0–0.02
		Solo	0.18 ± 0.17, N = 6	0.0–0.43	0.02 ± 0.02, N = 6	0.0–0.05	Solo	0.14 ± 0.11, N = 6	0.04–0.30	0.02 ± 0.02, N = 6	0.0–0.05
		Pair	0.12 ± 0.08, N = 6	0.02–0.22	0.02 ± 0.02, N = 6	0.0–0.05	Pair	0.14 ± 0.13, N = 6	0.02–0.35	0.01 ± 0.01, N = 6	0.0–0.03
Peeps (count)	Spectrogram	Control	51 ± 70.89, N = 6	4–192	107.17 ± 132.82, N = 6	6–372	Control	23.83 ± 21.79, N = 6	0–062	74.17 ± 96.16, N = 6	5–267
		Solo	1.5 ± 2.81, N = 6	0–7	41.17 ± 31.77, N = 6	2–71	Solo	10.50 ± 23.78, N = 6	0–59	68.00 ± 69.72, N = 6	0–199
		Pair	2.5 ± 3.39, N = 6	0–9	149.83 ± 167.66, N = 6	38–482	Pair	13.67 ± 16.45, N = 6	0–35	135.67 ± 81.52, N = 6	46–250
Trills (count)	Spectrogram	Control	29.5 ± 28.81, N = 6	6–78	49.67 ± 48.16, N = 6	0–111	Control	13.83 ± 21.90, N = 6	0–57	21.33 ± 37.37, N = 6	0–97
		Solo	2.5 ± 2.66, N = 6	0–7	31.17 ± 35.12, N = 6	0–95	Solo	1.33 ± 2.42, N = 6	0–6	18.17 ± 33.84, N = 6	0–86
		Pair	3.33 ± 3.93, N = 6	0–11	55.33 ± 43.14, N = 6	1–101	Pair	5.17 ± 6.91, N = 6	0–18	28.00 ± 43.05, N = 6	2–115
Long call intros (count)	Spectrogram	Control	0.33 ± 0.82, N = 6	0–2	4.17 ± 10.21, N = 6	0–25	Control	8.17 ± 13.51, N = 6	0–32	141.33 ± 306.00, N = 6	0–762
		Solo	0 ± 0, N = 6	0–0	0 ± 0, N = 6	0–0	Solo	1.33 ± 2.37, N = 6	0–8	98.50 ± 227.83, N = 6	0–563
		Pair	0 ± 0, N = 6	0–0	7.12 ± 12.14, N = 6	0–30	Pair	0 ± 0, N = 6	0–0	51.67 ± 62.63, N = 6	0–142
Long calls (count)	Spectrogram	Control	0 ± 0, N = 6	0–0	0 ± 0, N = 6	0–0	Control	0.67 ± 1.63, N = 6	0–4	1.33 ± 2.16, N = 6	0–5
		Solo	0 ± 0, N = 6	0–0	0 ± 0, N = 6	0–0	Solo	0.67 ± 0.41, N = 6	0–1	3 ± 6, N = 6	0–15
		Pair	0 ± 0, N = 6	0–0	0 ± 0, N = 6	0–0	Pair	0 ± 0, N = 6	0–0	3 ± 3.46, N = 6	0–8
Alarm calls (count)	Spectrogram	Control	0 ± 0, N = 6	0–0	0.17 ± 0.41, N = 6	0–1	Control	0 ± 0, N = 6	0–0	39.17 ± 95.94, N = 6	0–235
		Solo	0 ± 0, N = 6	0–0	0.17 ± 0.41, N = 6	0–1	Solo	0.17 ± 0.41	0–1	0.5 ± 1.22, N = 6	0–3
		Pair	0 ± 0, N = 6	0–0	0.5 ± 0.55, N = 6	0–1	Pair	0 ± 0, N = 6	0–0	0.83 ± 1.60, N = 6	0–4

Mean and standard deviation and ranges are provided for all variables across all timepoints and stimulus types.

Results

Reproductive status was retained in the final models for all vocal behaviors (see Supplementary material for reproductive status results). All final models included subject as a random effect, regardless of how much subject contributed to the model fit. All final model results are presented in Table 3. All significant behavioral and physiological outcome measures are included in the text as boxplot visualizations; all others are available upon request. The descriptive statistics for all variables are displayed by predictor in Table 2. Model results are presented in Table 3.

We ran four post-hoc Tukey's HSD tests across the entire project. We only performed Tukey's HSD on models for which a three-level predictor variable contributed significantly to the overall model variance. In three of our final models, playback stimulus type contributed to total variance (models for percent time orienting during the playback period, peeps during the playback period, and trills during the playback period). In one of our final models, reproductive status contributed to the total variance (model for peeps during the observation period). Full results of post-hoc comparisons are presented below.

Physiological responses

We successfully collected plasma blood samples from all subjects at all timepoints. However, for one test timepoint, there was enough volume to assay for cortisol, but not androgens. As such, results for androgens represent 35 samples while results for cortisol represent a full 36 samples. Female subjects had higher cortisol (conditional $R^2 = 0.3487$, $t_{(28)} = 5.378$, $p < 0.0001$; Figure 3A) and androgens (partial $R^2 = 0.\ 0.5430$, $t_{(27)} = 8.741$, p < 0.0001; Figure 3B) in the post-pairing timepoints than in the pre-pairing timepoints, regardless of stimulus type (Table 3).

Behavioral responses

We successfully captured the behavioral responses of all 6 subjects, across 2 pairing statuses, in response to three playback stimulus types. This resulted in a total of 36 observations per outcome variable across the study.

During the playback period, titi monkeys' percent time orienting to the direction of the audio stimuli varied based on playback stimulus type (partial $R^2 = 0.1009$, $t_{(28)} = 2.222$, $p = 0.0270$; Table 3). Our Tukey's HSD post-hoc test indicated that our subjects spent a lower percent of time orienting to the stimulus during the control playback period compared to the solo ($p < 0.001$) or duet ($p = 0.0276$) conditions; Figure 4A]. However, there was no significant difference in the means for the solo and paired playbacks [($p = 0.5353$), nor were there differences based upon pairing status (partial $R^2 = 0.0033$, $t_{(28)} = -0.892$, $p = 0.3591$).

Additionally, during the observation period, our subjects' percent time spent orienting to the stimuli did not vary based on pairing status (partial $R^2 = 0.0282$, $t_{(28)} = -1.218$, $p = 0.2131$) or playback stimulus type (partial $R^2 = 0.0192$, $t_{(28)} = 0.699$, $p = 0.4711$) (Figure 4B).

Within the playback period, subjects' percent time locomoting was not strongly predicted by pairing status (partial $R^2 = 0.0663$, $t_{(28)} = -1.642$, $p = 0.0968$) or stimulus type ($R^2 = 0.0073$, $t_{(28)} = 0.315$, $p = 0.5744$) (Figure 4C). However, during the observation period, titi monkeys spent a greater proportion of time locomoting during all the pre-pairing tests as compared to the post-pairing tests ($R^2 = 0.0942$, $t_{(28)} = -2.073$, $p = 0.0385$; Figure 4D), but playback stimulus did not predict locomotor behavior (partial $R^2 = 0.0080$, $t_{(28)} = 0.604$, $p = 0.5330$).

Vocal behaviors we scored included peeps, trills, alarm calls, long call introduction notes, and long calls (Table 1; Figure 2; Robinson, 1979a,b). Due to highly skewed data and few individuals vocalizing some vocal types, we were unable to run models for alarm calls, long call introduction notes, and long calls. However, descriptive statistics of these outcome variables are available in Table 2, along with the raw data in our Supplementary material.

During the playback period, the number of peeps vocalized was predicted by playback stimulus type (partial $R^2 = 0.1271$, $t_{(28)} = -2.213$, $p = 0.0256$), but not reproductive status (partial $R^2 = 0.0157$, $t_{(28)} = 0.677$, $p = 0.4206$) or pairing status (partial $R^2 = 0.0127$, $t_{(28)} = -0.639$, $p = 0.4681$). Our Tukey's HSD post-hoc test indicated that our subjects peeped more in response to the control playback compared to the solo playback ($p = 0.0333$) and the duet playback ($p = 0.0421$; Figure 4E). However, there was not a significant difference in the number of peeps in response to the solo and pair playbacks ($p = 0.9958$).

Additionally, during the playback period, the number of trills was also predicted by playback stimulus type (partial $R^2 = 0.1737$, $t_{(28)} = -2.630$, $p = 0.0080$) but not pairing status (partial $R^2 = 0.0073$, $t_{(28)} = -0.5953$, $p = 0.5695$) or reproductive status (partial $R^2 = 0.0007$, $t_{(28)} = -0.164$, $p = 0.8616$). Our Tukey's HSD post-hoc test indicated that our subjects trilled more in response to the control playback compared to the solo playback ($p < 0.001$) and the duet playback ($p = 0.0029$; Figure 4G), but there was not a significant difference in the number of trills in response to the solo and duet playback types ($p = 0.8985$).

In the observation period, pairing status predicted subjects' number of trills, in that subjects vocalized more trills pre-pairing compared to post-pairing (partial $R^2 = -0.1419$, $t_{(28)} = -2.351$, $p = 0.0373$; Figure 4H). Stimulus type (partial $R^2 = -0.0119$, $t_{(28)} = 0.530$, $p = 0.5799$) nor reproductive status (partial $R^2 = -0.1335$, $t_{(28)} = 1.255$, $p = 0.2857$) predicted trill behavior.

Reproductive status predicted only one behavior in this study: number of peeps during the observation period (partial $R^2 = 0.1202$, $t_{(28)} = -2.054$, $p = 0.0330$) (Figure 5). Pairing status (partial $R^2 = 0.0494$, $t_{(28)} = 1.434$, $p = 0.1406$) nor playback stimulus type (partial $R^2 = 0.0416$, $t_{(28)} = -1.327$, $p = 0.1729$) predicted peep behavior in the observation period (Figure 4F). Our Tukey's HSD post-hoc test indicated our subjects vocalized more peeps when reproductively cycling ($p = 0.0217$) or pregnant ($p = 0.0433$) compared to non-cycling. However, there was not a significant difference in the number of trills vocalized between cycling and pregnant females ($p = 0.6520$).

Finally, for latency to vocalize, one of our 6 females did not vocalize during the post-pairing, solo playback stimulus test, resulting in a total of 35 latencies to vocalize and one censored observation. Our Cox Proportional Hazards models' AIC values for our first (null) model and second model were 191.44 and 192.60, respectively, indicating that our null model had a slightly better fit. However, the difference in AIC values was relatively small (1.16), suggesting that both models may provide a reasonable fit to the data. Broadly speaking, reproductive status, playback type, nor pairing status influenced female titi monkeys' latency to vocalize (Figure 4I).

TABLE 3 Results of the linear mixed-effects models assessing physiological and behavioral responses to different stimulus types during two different pairing statuses.

Model		Estimate	s.e.	df	t-value	LLR	p value	Partial R^2	Marginal R^2	Conditional R^2
Cortisol ~ PairingStatus + Stimulus										
	Intercept	2.962	0.137	28	21.652				0.3492	0.5949
	PairingStatus	0.53	0.099	28	5.378	21.286	**<0.0001**	0.3487		
	Stimulus	−0.012	0.06	28	−0.193	0.04	0.8417	0.0004		
Random effects (adjusted repeatability of subject = 0.412)										
	Subject	0.247				7.863	0.005			
	Residual	0.296								
Testosterone ~ PairingStatus + Stimulus										
	Intercept	2.406	0.061	27	39.134				0.5456	0.7742
	PairingStatus	0.337	0.039	27	8.741	39.278	**<0.0001**	0.5430		
	Stimulus	−0.014	0.023	27	−0.620	0.411	0.5217	0.0027		
Random effects (adjusted repeatability of subject = 0.534)										
	Subject	0.122				13.026	0.0003			
	Residual	0.114								
Playback period										
% Time Orienting ~ PairingStatus + Stimulus										
	Intercept	0.084	0.040	28	2.137				0.1055	0.3819
	PairingStatus	−0.027	0.030	28	−0.892	0.841	0.3591	0.0033		
	Stimulus	0.042	0.019	28	2.222	4.873	**0.027**	0.1009		
Random effects (adjusted repeatability of subject = 0.342)										
	Subject	0.067				5.483	0.0192			
	Residual	0.092								
% Time Locomotion ~ PairingStatus + Stimulus										
	Intercept	0.328	0.077	28	4.277				0.0736	0.1733
	PairingStatus	−0.120	0.073	28	−1.642	2.758	0.0968	0.0663		
	Stimulus	0.024	0.045	28	0.544	0.315	0.5744	0.0073		
Random effects (adjusted repeatability of subject = 0.134)										
	Subject	0.086				0.833	0.3615			
	Residual	0.219								
Peeps ~ PairingStatus + Stimulus + ReproductiveStatus										
	Intercept	29.270	12.139	27	2.411				0.1439	0.1439
	PairingStatus	−10.293	16.101	27	−0.639	0.526	0.4681	0.0127		
	Stimulus	−14.917	6.741	27	−2.213	4.983	**0.0256**	0.1271		
	ReproductiveStatus	7.959	11.750	27	0.677	0.621	0.4206	0.0157		
Random effects (adjusted repeatability of subject = 0.036)										
	Subject	6.390				<0.001	0.9998			
	Residual	33.024								
Trills ~ PairingStatus + Stimulus + ReproductiveStatus										
	Intercept	20.931	5.736	27	3.649				0.1959	0.1959
	PairingStatus	−4.111	7.648	27	−0.5953	0.324	0.5695	0.0073		
	Stimulus	−8.708	3.311	27	−2.630	7.043	**0.0080**	0.1737		
	ReproductiveStatus	−0.889	5.408	27	−0.164	0.030	0.8616	0.0007		

(Continued)

TABLE 3 (Continued)

Model		Estimate	s.e.	df	*t*-value	LLR	*p* value	Partial R^2	Marginal R^2	Conditional R^2
Random effects (adjusted repeatability of subject = <0.001)										
	Subject	0.001				<0.001	0.9998			
	Residual	15.983								
Observation period										
% Time Orienting ~ PairingStatus + Stimulus										
	Intercept	0.016	0.006	28	2.730				0.0381	0.3509
	PairingStatus	−0.005	0.004	28	−1.218	1.550	0.2131	0.0282		
	Stimulus	0.002	0.003	28	0.699	0.519	0.4711	0.0192		
Random effects (adjusted repeatability of subject = 0.359)										
	Subject	0.010				6.007	0.0142			
	Residual	0.013								
% Time Locomoting ~ PairingStatus + Stimulus										
	Intercept	0.315	0.063	28	5.010				0.1022	0.2635
	PairingStatus	−0.116	0.056	28	−2.073	4.281	**0.0385**	0.0942		
	Stimulus	0.021	0.034	28	0.604	0.389	0.5330	0.0080		
Random effects (adjusted repeatability of subject = 0.210)										
	Subject	0.086				2.096	0.1477			
	Residual	0.167								
Peeps ~ PairingStatus + Stimulus + ReproductiveStatus										
	Intercept	112.368	38.146	27	2.946				0.1630	0.2175
	PairingStatus	71.263	49.700	27	1.434	2.171	0.1406	0.0494		
	Stimulus	26.042	19.624	27	1.327	1.857	0.1729	0.0416		
	ReproductiveStatus	−78.041	37.989	27	−2.054	4.546	**0.0330**	0.1202		
Random effects (adjusted repeatability of subject = 0.113)										
	Subject	34.270				0.318	0.5729			
	Residual	96.138								
Trills ~ PairingStatus + Stimulus + ReproductiveStatus										
	Intercept	32.935	17.942	27	1.836				−0.0452	0.5120
	PairingStatus	−41.629	17.706	27	−2.351	4.338	**0.0373**	−0.1419		
	Stimulus	3.083	5.822	27	0.530	0.3064	0.5799	−0.0119		
	ReproductiveStatus	18.740	14.937	27	1.255	1.140	0.2857	−0.1335		
Random effects (adjusted repeatability of subject = 0.580)										
	Subject	33.498				9.768	0.0018			
	Residual	28.524								

The Model column indicates the statistical model tested (written in the form of independent variable ~ dependent variables). Bolded values indicate the *p* value was significant at $p < 0.05$. All models included subject as a random effect. For this table, s.e. indicates the standard error of the corresponding parameter estimate. df indicates the degrees of freedom. LLR indicates the log-likelihood ratio. The adjusted repeatability of the random effect represents the proportion of variance due to the random effect over the total variance not explained by fixed effects. A smaller value of adjusted repeatability represents higher overall repeatability and thus higher reliability. The reference levels for our predictor variables were, respectively: pairing status (ref. level: unpaired), stimulus (ref. level: control), and reproductive status (non-cycling). The full R script for these analyses is presented as Supplementary material.

Discussion

Generally, the findings of the present study are consistent with what is currently known about titi monkey social behavior. This project is the first to validate the use of vocal playbacks in the captive setting, providing evidence that titi monkeys do respond to social acoustic stimuli in a manner consistent with expectations for their species and social status. Across all outcome variables, the response to solo and duet playback stimuli did not differ significantly. As such, we focus on differences between control and social (solo and duet) playback stimuli.

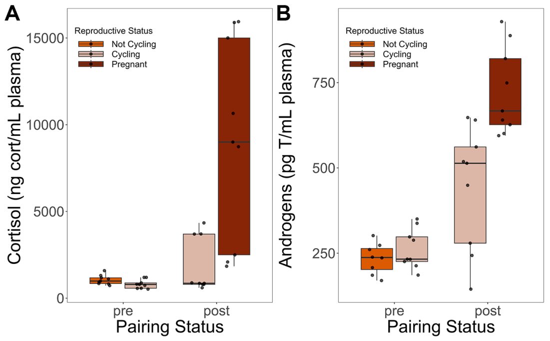


FIGURE 3
Cortisol **(A)** and androgen **(B)** boxplots of female titi monkey plasma hormone levels according to pairing status and reproductive status. Each box delineates the 1st and 3rd interquartile (25% and 75%) with the median as the 2nd interquartile (50%). The whiskers represent the data "range." Data points above and below the whiskers are outliers. For both cortisol **(A)** and androgens **(B)**, females had higher values post-pairing than pre-pairing. This effect existed regardless of stimulus type or reproductive status. For descriptive statistics, see Table 2. For model results, see Table 3.

Notably, a few findings emerged from this project peripheral to our initial predictions. We will first discuss the physiological responses to playbacks, behavioral responses to the playbacks, and then report interesting side notes, before discussing the limitations of and future recommendations following this study.

Physiological responses to playbacks

Female titi monkeys had higher androgen and cortisol levels post-pairing compared to pre-pairing. This difference existed irrespective of female reproductive status (removed from the final model) or playback type (included in the final model). Though reproductive status was removed from the final model due to backwards model selection, graphs of the cortisol values do indicate that pregnancy and cycling generally increase cortisol levels (Figure 3). However, the fact that cortisol and androgens are higher at post-pairing timepoints compared to pre-pairing timepoints may indicate a territorial response to the playback paradigm. Given the titi monkey's unique parental care system in which the father contributes significantly to infant care (Mendoza and Mason, 1986b), combined with the fact that female titi

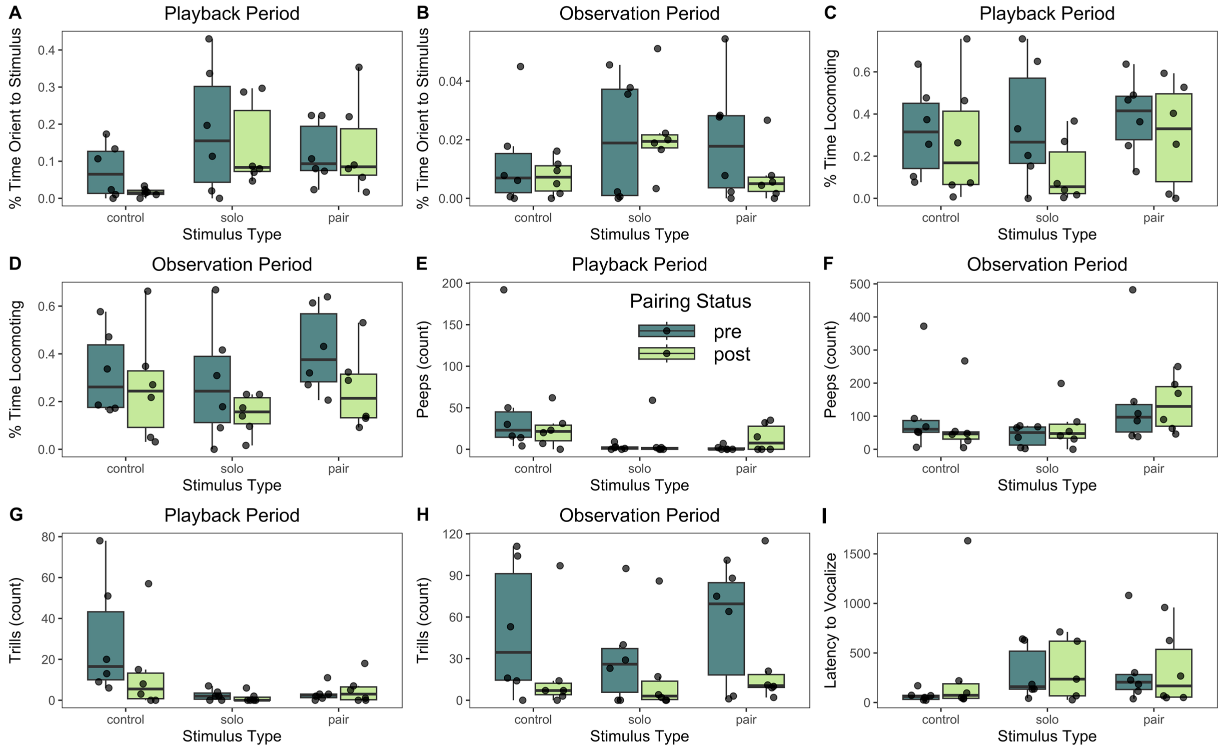


FIGURE 4
Boxplots of all behavioral outcomes modeled in this study, according to pairing status and stimulus type. Each box delineates the 1st and 3rd interquartile (25% and 75%) with the median as the 2nd interquartile (50%). The whiskers represent the data "range." Data points above and below the whiskers are outliers. **(A)** During the playback period, percent time spent orienting to the stimulus varied based upon playback stimulus type. **(B)** During the observation period, percent time orienting did not vary significantly based upon any of our predictor variables. **(C,D)** The percent time spent locomoting did not vary based upon any of our predictor variables for either the playback period **(C)** or the observation period **(D)**. **(E,F)** Playback stimulus type predicted the number of peeps emitted during the playback period **(E)**, but not during the observation period **(F)**. **(G)** The number of trills vocalized during the playback period varied based on stimulus type in that female titi monkeys vocalized more trills during the control playback compared to the two social playbacks. **(H)** During the observation period, the number of trills vocalized was predicted by pairing status in that titi monkey females vocalized more trills pre-pairing compared to post-pairing. **(I)** The latency to vocalize was not predicted by any of our predictors. For descriptive statistics, see Table 2. For model results, see Table 3.

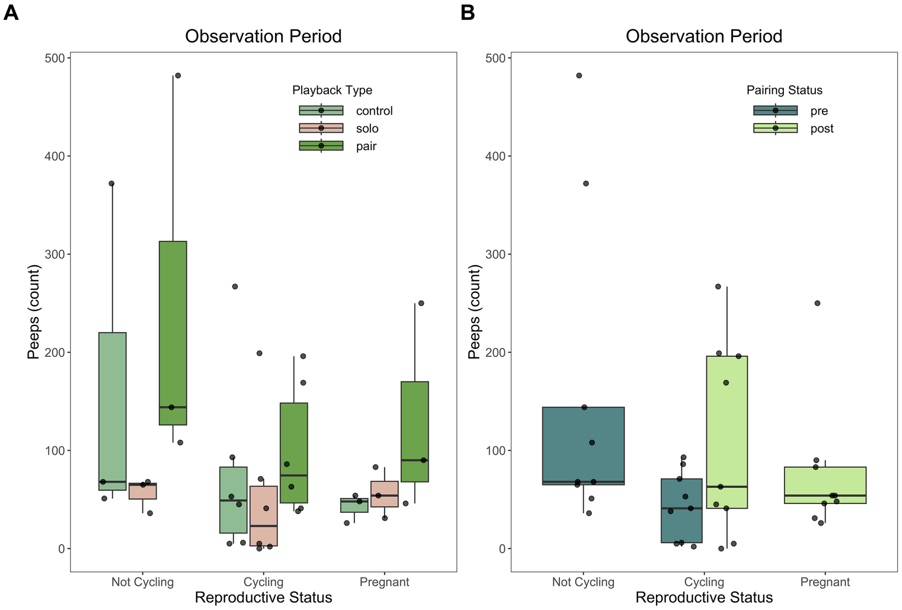


FIGURE 5
The number of peeps was predicted by reproductive status during the observation period. Female titi monkeys vocalized more peeps when not reproductively cycling as compared to reproductively cycling or pregnant females. This effect held regardless of playback stimulus type **(A)** or pairing status **(B)**. For descriptive statistics, see Table 2. For model results, see Table 3.

monkeys actively maintain proximity with their mates (Dolotovskaya et al., 2020b), the higher androgen levels observed in post-pairing females in this study may reflect a reversal of traditional sex roles in this species. Or, at the very least, an equivalent contribution of both male and female titi monkeys to territorial responses. Female titi monkeys may respond behaviorally (Robinson, 1981; present study) and endocrinologically (present study) to territorial threats in a manner similar to male individuals of other species (Ord, 2021). The findings of this study are supported by the behavioral results of a simulated intruder test in which paired adult male and female titi monkeys responded with agonistic behaviors when viewing themselves in a mirror (Mercier et al., 2020). This study provided evidence of territorial behavior in female titi monkeys, including back arching and tail-lashing (Mercier et al., 2020). Future studies should investigate the role of androgens in both male and female titi monkeys' responses to territorial intrusions.

Behavioral responses to playbacks

The playback stimulus type (control vs. solo vs. duet) predicted behavior during the playback period (5-min playback), but not the observation period (30 min following the playback). Specifically, female titi monkeys vocalized a greater number of trills and peeps during the control playbacks as opposed to the social playbacks (solo and duet). Additionally, subjects spent a greater proportion of time orienting to the direction of the playback audio during the social playbacks (solo and duet) as opposed to the control stimuli. Together, these results imply titi monkey females are actively listening (not vocalizing as much) and assessing (looking in the direction of) social signals (solo and duet) as compared to the control playback. Alternatively, or in conjunction, female titi monkeys may be vocalizing more in response to the control playback due to a lack of acoustic competition (i.e., if no other monkeys are vocalizing, the subject may vocalize more). Subjects' vocal responses during the control playback correspond with typical titi monkey responses to separation from their mate or family members in previous separation paradigms (Mendoza and Mason, 1986a; Hoffman et al., 1995; Arias del Razo et al., 2022a). These results partially support our initial hypothesis that titi monkey females would respond differently to control playbacks versus social playbacks, but the lack of a distinctly different response to the male solos or pair duets does not allow us to speculate on what information titi monkey females do or do not perceive within these unfamiliar calls. This result may be a reflection of titi monkeys' generalized neophobic responses to unfamiliar stimuli, as seen previously in neophobia (Hennessy et al., 1995) and novel object presentation studies (Lau et al., 2021).

Pairing status (pre-pairing vs. post-pairing) predicted vocal and locomotor behavior during the observation period (30 min following the playback). Titi monkeys trilled more in the pre-pairing conditions than the post-pairing conditions. Trill vocalizations are typically uttered by infant and juvenile titi monkeys more often than adults and are commonly thought of as "infant" vocalizations (Lau et al., 2020; Savidge and Bales, 2020) as trill vocalizations typically elicit reunion behaviors from parents (Hoffman et al., 1995). Based upon the younger age of our females during their pre-pairing timepoint and status as unpaired females within their natal groups, the larger number of trills pre-pairing compared to post-pairing fits the pre-existing knowledge of titi vocal behavior at different developmental stages and social situations.

In addition to their vocal responses, titi monkey females also spent more time locomoting in the pre-pairing observation periods as compared to the post-pairing observation periods. In the wild, unpaired titi monkeys occupy either their parents' territories or exist as a floater without a territory prior to finding a mate (Dolotovskaya et al., 2020a). As such, a withdraw response (locomotion) as opposed to defense via long calling (Robinson, 1981) is consistent with pre-pairing females' lack of a territory that is theirs to defend. Given the laboratory nature of this study, titi monkeys are unable to show species-typical withdrawal or fleeing behavior that would likely occur in a wild setting.

Taken together, female titi monkeys' response following a playback (during the observation period) is determined by females' pairing status more so than the content of the individual playbacks.

Notes of interest

Reproductive status was excluded from all models except one. The final model for number of peeps during the observation period retained reproductive status as a fixed effect. In this study, titi monkey females vocalized more peeps during the observation period while non-cycling compared to cycling or pregnant females. Peeps are used primarily as contact calls or general arousal signals (Robinson, 1979a,b; Arias del Razo et al., 2022a) While this single result alone is not enough to fully assess the impacts of reproductive status on titi monkey vocal behavior, the greater number of peeps uttered by non-cycling females may suggest that cycling and pregnant females spend more time attending to the environment while non-cycling females may employ a strategy of soliciting their family group. However, the limited sample size of 6 individuals does not allow for any truly conclusive assertions about titi monkey vocal behavior regarding reproductive status.

Limitations and future directions

While this project was originally designed with a target sample size of 9 individuals, we were restricted to only 6 individuals due to COVID-19 pandemic-related issues. We recommend additional experiments to bolster the findings presented here. This study was also limited to female animals as part of a larger project assessing female pair bonding. Projects that include male titi monkeys will allow for comparisons between the sexes.

Additionally, while this study was conducted at a consistent time of day to control for daily hormone fluctuations, future studies may find interesting behavioral variation in response to playbacks at different times of day. Temporal fluctuations of behavior in this species have not yet been investigated.

The duet playbacks used in this study were broadcast from one speaker. Previous work in avian studies indicates that multi-speaker playbacks simulate a more realistic duet playback (Douglas and Mennill, 2010). Separating each sound source from a titi monkey duet recording is very difficult given substantial overlap between male and female contributions. However, it would be possible to artificially create a duet by broadcasting two solo songs simultaneously in a stereo playback design (I.e., male song from speaker A and female song from speaker B). To make the playback realistic, each song would have to be edited to ensure accurate coordination of male and female song phrases when triggering the playback. This method would constitute an ideal, unfamiliar duet stimulus. Future studies should attempt this method.

One possible confounding factor is the nested separation study occurring within this playback study. Adult titi monkeys' attention and anxiety-related behaviors are impacted by the removal of a pair mate from the enclosure (Savidge and Bales, 2020). While a separation from the subject's family or mate (depending on pairing status) occurred for all playback tests, the overall impacts of separation cannot be disentangled from the impacts of each playback stimulus type. The results found here may have been stronger if separation did not occur, as the separation paradigm induces physiological and behavioral arousal (Arias del Razo et al., 2022a). However, by separating females from their family/mate, we were able to ensure that the results found here were not confounded by idiosyncratic behavior of the family/mate and were individually driven. The results presented here suggest that beyond the effects of separation, social playbacks do alter behavior and physiology of the listener. Future studies should aim to replicate this study and compare individuals' responses to those of paired males and females listening to playbacks together as the joint pair response to playbacks will further illuminate social communication patterns in this species.

Conclusion

In summary, we found evidence that female titi monkeys attend to social signals by vocalizing less and orienting more in the direction of the playback than control recordings while the playback is occurring regardless of pairing status. However, in the time immediately following any playback type, female's pairing status predicts vocal and locomotor responses irrespective of playback type. Namely, female titi monkeys trill more pre-pairing and long call more post-pairing, as well as spend a greater proportion of time locomoting at pre-pairing timepoints. Future studies should aim to understand male titi monkeys' responses to different acoustic signals as well as those of paired monkeys listening to playbacks in tandem.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This study was approved by the IACUC of the University of California, Davis. This study met all legal requirements of the United States as well as guidelines set by the American Society of Primatologists for the ethical treatment of non-human primates. This study was carried out in compliance with the ARRIVE guidelines.

Author contributions

AL and KB designed the study. AL carried out data collection. AL and AC processed the data, ran all statistical analyses, and composed the first draft of this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Institutes of Health (grant numbers HD092055 and OD011107), the Good Nature Institute, and the American Society of Primatologists Small Grant.

Acknowledgments

We gratefully acknowledge National Institutes of Health grant HD092055 and OD011107, as well as the Good Nature Institute. We thank the American Society of Primatologists for awarding the Small Grant to AL in support of this project. Many thanks to Rebecca Cotterman and Al Conley for their expertise in endocrinology. We thank the husbandry staff at the CNPRC for providing excellent care to all subjects. We thank Damien Caillaud and Gail Patricelli for early insight on the design of this study. We thank many individuals from the Bales Lab who helped with data collection, including Lynea Witczak, Madison Dufek, Amber Wright, Chloe Karaskiewicz, John Paulus, Alexander Baxter, Logan Savidge, and Mathieu Legrand. We appreciate comments on early drafts of this manuscript from Chloe Karaskiewicz and help with Figure 1 from Danny Lau. Our thanks go to the titi monkeys who participated in this study, and our final thanks go to Chuck Snowdon for his years of support and enthusiasm for this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2023.1145205/full#supplementary-material

References

Arias del Razo, R., Berger, T., Conley, A. J., Freeman, S. M., Goetze, L. R., Jacob, S., et al. (2020). Effects of chronic intranasal oxytocin on behavior and cerebral glucose uptake in juvenile titi monkeys. *Psychoneuroendocrinology* 113:104494. doi: 10.1016/j.psyneuen.2019.104494

Arias del Razo, R., Vazquez, M. D. L. V., Turcanu, P., Legrand, M., Floch, M., Weinstein, T. A., et al. (2022b). Long term effects of chronic intranasal oxytocin on adult pair bonding behavior and brain glucose uptake in titi monkeys (*Plecturocebus cupreus*). *Horm. Behav.* 140:105126. doi: 10.1016/j.yhbeh.2022.105126

Arias del Razo, R., Vazquez, M. D. L. V., Turcanu, P., Legrand, M., Lau, A. R., Weinstein, T. A., et al. (2022a). Effects of chronic and acute intranasal oxytocin treatments on temporary social separation in adult Titi monkeys (*Plecturocebus cupreus*). *Front. Behav. Neurosci.* 16:877631 doi: 10.3389/fnbeh.2022.877631

Bales, K. L., Ardekani, C. S., Baxter, A., Karaskiewicz, C. L., Kuske, J. X., Lau, A. R., et al. (2021). What is a pair bond? *Horm. Behav.* 136:105062. doi: 10.1016/j.yhbeh.2021.105062

Bales, K. L., Del Razo, R. A., Conklin, Q. A., Hartman, S., Mayer, H. S., Rogers, F. D., et al. (2017). Focus: comparative medicine: Titi monkeys as a novel non-human primate model for the neurobiology of pair bonding. *Yale J. Biol. Med.* 90, 373–387.

Bossuyt, F. (2002). Natal dispersal of titi monkeys (*Callicebus moloch*) at Cocha Cashu, Manu National Park, Peru. *Am. J. Phys. Anthropol.* 34:47.

Bradbury, J. W., and Vehrencamp, S. L. (1998). *Principles of animal communication*. Sinauer Associates Inc, Sunderland, MA.

Cäsar, C., Byrne, R., Young, R. J., and Zuberbühler, K. (2012). The alarm call system of wild black-fronted titi monkeys, *Callicebus nigrifrons*. *Behav. Ecol. Sociobiol.* 66, 653–667. doi: 10.1007/s00265-011-1313-0

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., and Setz, E. Z. (2014). Vocal behavior of black-fronted titi monkeys (*Callicebus nigrifrons*): acoustic properties and behavioral contexts of loud calls. *Am. J. Primatol.* 76, 788–800. doi: 10.1002/ajp.22270

Caselli, C. B., Mennill, D. J., Gestich, C. C., Setz, E. Z., and Bicca-Marques, J. C. (2015). Playback responses of socially monogamous black-fronted titi monkeys to simulated solitary and paired intruders. *Am. J. Primatol.* 77, 1135–1142. doi: 10.1002/ajp.22447

Cheney, D. L., and Seyfarth, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proc. Natl. Acad. Sci.* 115, 1974–1979. doi: 10.1073/pnas.1717572115

Clink, D. J., Comella, I. A., Tasirin, J. S., and Klinck, H. (2022). Tarsier islands: exploring patterns of variation in tarsier duets from offshore islands of North Sulawesi. *Am. J. Primatol.* 76:e23410. doi: 10.1002/ajp.23410

Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour* 156, 1471–1494. doi: 10.1163/1568539X-00003575

Clink, D. J., Lau, A. R., Kanthaswamy, S., Johnson, L. M., and Bales, K. L. (2022). Moderate evidence for heritability in the duet contributions of a south American primate. *J. Evol. Biol.* 35, 51–63. doi: 10.1111/jeb.13962

Conley, A. J., Berger, T., Del Razo, R. A., Cotterman, R. F., Sahagún, E., Goetze, L. R., et al. (2022). The onset of puberty in colony-housed male and female titi monkeys (*Plecturocebus cupreus*): possible effects of oxytocin treatment during peri-adolescent development. *Horm. Behav.* 142:105157. doi: 10.1016/j.yhbeh.2022.105157

Corley, M., Perea-Rodriguez, J. P., Valeggia, C., and Fernandez-Duque, E. (2021). Associations between fecal cortisol and biparental care in a pair-living primate. *Am. J. Phys. Anthropol.* 176, 295–307. doi: 10.1002/ajpa.24368

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

Depeine, C. C., Rotundo, M., Juárez, C. P., and Fernandez-Duque, E. (2008). Hoot calling in owl monkeys (*Aotus azarai*) of Argentina: sex differences and function. *Am. J. Primatol.* 70:69.

Dolotovskaya, S., and Heymann, E. W. (2022). Coordinated singing in coppery Titi monkeys (*Plecturocebus cupreus*): resource or mate defense? *Front. Ecol. Evol.* 10. doi: 10.3389/fevo.2022.898509

Dolotovskaya, S., Roos, C., and Heymann, E. W. (2020a). Genetic monogamy and mate choice in a pair-living primate. *Sci. Rep.* 10, 1–13. doi: 10.1038/s41598-020-77132-9

Dolotovskaya, S., Walker, S., and Heymann, E. W. (2020b). What makes a pair bond in a Neotropical primate: female and male contributions. *R. Soc. Open Sci.* 7:191489. doi: 10.1098/rsos.191489

Douglas, S. B., and Mennill, D. J. (2010). A review of acoustic playback techniques for studying avian vocal duets. *J. Field Ornithol.* 81, 115–129. doi: 10.1111/j.1557-9263.2010.00268.x

Dunbar, R. I., and Shultz, S. (2021). Social complexity and the fractal structure of group size in primate social evolution. *Biol. Rev.* 96, 1889–1906. doi: 10.1111/brv.12730

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., et al. (2018). More than a feeling: a unified view of stress measurement for population science. *Front. Neuroendocrinol.* 49, 146–169. doi: 10.1016/j.yfrne.2018.03.001

Fernandez-Duque, E., Valeggia, C. R., and Mason, W. A. (2000). Effects of pair-bond and social context on male–female interactions in captive titi monkeys (*Callicebus moloch*, primates: *Cebidae*). *Ethology* 106, 1067–1082. doi: 10.1046/j.1439-0310.2000.00629.x

Fichtel, C., and Hilgartner, R. (2013). "Noises in the dark: vocal communication in Lepilemur ruficaudatus and other nocturnal pair-living primates" in *Leaping ahead:*

Advances in prosimian biology. eds. J. Master, M. Gamba and F. Génin (New York: Springer), 297–304. doi: 10.1007/978-1-4614-4511-1_33

Fischer, J., Noser, R., and Hammerschmidt, K. (2013). Bioacoustic field research: a primer to acoustic analyses and playback experiments with primates. *Am. J. Primatol.* 75, 643–663. doi: 10.1002/ajp.22153

Fishbein, A. R., Prior, N. H., Brown, J. A., Ball, G. F., and Dooling, R. J. (2021). Discrimination of natural acoustic variation in vocal signals. *Sci. Rep.* 11:916. doi: 10.1038/s41598-020-79641-z

Garcia de la Chica, A., Wood, D. B., Rotundo, M., and Fernandez-Duque, E. (2021). Responses of a pair-living, sexually monogamous primate to the simulated presence of solitary individuals: a field playback experiment. *Ethology* 127, 1002–1018. doi: 10.1111/eth.13222

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/S1464793101005826

Ghazanfar, A. A., and Hauser, M. D. (2001). The auditory behaviour of primates: a neuroethological perspective. *Curr. Opin. Neurobiol.* 11, 712–720. doi: 10.1016/s0959-4388(01)00274-4

Hennessy, M. B., Mendoza, S. P., Mason, W. A., and Moberg, G. P. (1995). Endocrine sensitivity to novelty in squirrel monkeys and titi monkeys: species differences in characteristic modes of responding to the environment. *Physiol. Behav.* 57, 331–338. doi: 10.1016/0031-9384(94)00250-9

Hoffman, K. A., Mendoza, S. P., Hennessy, M. B., and Mason, W. A. (1995). Responses of infant titi monkeys, *Callicebus moloch*, to removal of one or both parents: evidence for paternal attachment. *Dev. Psychobiol.* 28, 399–407. doi: 10.1002/dev.420280705

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., Scheibe, S., et al. (2016). "Package 'multcomp'." *Simultaneous inference in general parametric models. Project for statistical computing, Vienna, Austria*.

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2022). *Raven pro: Interactive sound analysis software (version 1.6.3) [computer software]*. Ithaca, NY: The Cornell Lab of Ornithology.

Kappeler, P. M., and van Schaik, C. P. (2002). Evolution of primate social systems. *Int. J. Primatol.* 23, 707–740. doi: 10.1023/A:1015520830318

Karaskiewicz, C. L., Witczak, L. R., Lau, A. R., Dufek, M. E., and Bales, K. L. (2021). Parenting costs time: changes in pair bond maintenance across pregnancy and infant rearing in a monogamous primate (*Plecturocebus cupreus*). *New Dir. Child Adolesc. Dev.* 2021, 21–42. doi: 10.1002/cad.20438

Kleiman, D. G. (1977). Monogamy in mammals. *Q. Rev. Biol.* 52, 39–69. doi: 10.1086/409721

Kroodsma, D. E. (1989). Suggested experimental designs for song playbacks. *Anim. Behav.* 37, 600–609. doi: 10.1016/0003-3472(89)90039-0

Kroodsma, D. E. (1990). Using appropriate experimental designs for intended hypotheses in 'song' playbacks, with examples for testing effects of song repertoire sizes. *Anim. Behav.* 40, 1138–1150. doi: 10.1016/S0003-3472(05)80180-0

Lau, A. R., Clink, D. J., and Bales, K. L. (2020). Individuality in the vocalizations of infant and adult coppery titi monkeys (*Plecturocebus cupreus*). *Am. J. Primatol.* 82:e23134. doi: 10.1002/ajp.23134

Lau, A. R., Grote, M. N., Dufek, M. E., Franzetti, T. J., Bales, K. L., and Isbell, L. A. (2021). Titi monkey neophobia and visual abilities allow for fast responses to novel stimuli. *Sci. Rep.* 11, 1–9. doi: 10.1038/s41598-021-82116-4

Lukas, D., and Clutton-Brock, T. H. (2013). The evolution of social monogamy in mammals. *Science* 341, 526–530. doi: 10.1126/science.1238677

Maninger, N., Mendoza, S. P., Williams, D. R., Mason, W. A., Cherry, S. R., Rowland, D. J., et al. (2017). Imaging, behavior and endocrine analysis of "jealousy" in a monogamous primate. *Front. Ecol. Evol.* 5:119. doi: 10.3389/fevo.2017.00119

Marshall, J. T. Jr., and Marshall, E. R. (1976). Gibbons and their territorial songs. *Science* 193, 235–237. doi: 10.1126/science.193.4249.235

Mendoza, S. P., and Mason, W. A. (1986a). Contrasting responses to intruders and to involuntary separation by monogamous and polygynous new world monkeys. *Physiol. Behav.* 38, 795–801. doi: 10.1016/0031-9384(86)90045-4

Mendoza, S. P., and Mason, W. A. (1986b). Parental division of labour and differentiation of attachments in a monogamous primate (*Callicebus moloch*). *Anim. Behav.* 34, 1336–1347. doi: 10.1016/S0003-3472(86)80205-6

Mendoza, S. P., and Moberg, G. P. (1985). Species differences in adrenocortical activity of New World primates: response to dexamethasone suppression. *Am. J. Primatol.* 8, 215–224. doi: 10.1002/ajp.1350080304

Mercier, F., Witczak, L. R., and Bales, K. L. (2020). Coppery titi monkey (*Plecturocebus cupreus*) pairs display coordinated behaviors in response to a simulated intruder. *Am. J. Primatol.* 82:e23141. doi: 10.1002/ajp.23141

Moynihan, M. (1966). Communication in the titi monkey, Callicebus. *J. Zool.* 150, 77–127. doi: 10.1111/j.1469-7998.1966.tb02999.x

Müller, A. E., and Anzenberger, G. (2002). Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. doi: 10.1159/000064788

Nietsch, A. (1999). Duet vocalizations among different populations of Sulawesi tarsiers. *Int. J. Primatol.* 20, 567–583. doi: 10.1023/A:1020342807709

Ord, T. J. (2021). Costs of territoriality: a review of hypotheses, meta-analysis, and field study. *Oecologia* 197, 615–631. doi: 10.1007/s00442-021-05068-6

Palombit, R. (1994). Dynamic pair bonds in hylobatids: implications regarding monogamous social systems. *Behaviour* 128, 65–101. doi: 10.1163/156853994X00055

Pinheiro, J. (2009). Nlme: linear and nonlinear mixed effects models. R package version 3. 1-96. Available at: http://cran. r-project. org/web/packages/nlme/.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., et al. (2017). Package 'nlme'. Linear and nonlinear mixed effects models, version, 3(1).

Place, V. A., and Nichols, K. C. (1991). Transdermal delivery of testosterone with TESTODERM™ to provide a Normal circadian pattern of testosterone. *Ann. N. Y. Acad. Sci.* 618, 441–449. doi: 10.1111/j.1749-6632.1991.tb27263.x

R Core Team (2022). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, J. G. (1979a). An analysis of the organization of vocal communication in the titi monkey Callicebus moloch. *Z. Tierpsychol.* 49, 381–405.

Robinson, J. G. (1979b). Vocal regulation of use of space by groups of titi monkeys Callicebus moloch. *Behav. Ecol. Sociobiol.* 5, 1–15. doi: 10.1007/BF00302691

Robinson, J. G. (1981). Vocal regulation of inter-and intragroup spacing during boundary encounters in the titi monkey, *Callicebus moloch*. *Primates* 22, 161–172. doi: 10.1007/BF02382607

Rothwell, E. S., Carp, S. B., Savidge, L. E., Mendoza, S. P., and Bales, K. L. (2020). Relationship tenure differentially influences pair-bond behavior in male and female socially monogamous titi monkeys (*Callicebus cupreus*). *Am. J. Primatol.* 82:e23181. doi: 10.1002/ajp.23181

Savidge, L. E., and Bales, K. L. (2020). An animal model for mammalian attachment: infant titi monkey (*Plecturocebus cupreus*) attachment behavior is associated with their social behavior as adults. *Front. Psychol.* 11:25. doi: 10.3389/fpsyg.2020.00025

Silk, J. B. (2007). The adaptive value of sociality in mammalian groups. *Philos. Trans. R. Soc. B* 362, 539–559. doi: 10.1098/rstb. 2006.1994

Singletary, B., and Tecot, S. (2020). Multimodal pair-bond maintenance: a review of signaling across modalities in pair-bonded nonhuman primates. *Am. J. Primatol.* 82:e23105. doi: 10.1002/ajp.23105

Smith, T. E., and French, J. A. (1997). Psychosocial stress and urinary cortisol excretion in marmoset monkeys. *Physiol. Behav.* 62, 225–232. doi: 10.1016/S0031-9384(97)00103-0

Tardif, S., Bales, K., Williams, L., Moeller, E. L., Abbott, D., Schultz-Darken, N., et al. (2006). Preparing New World monkeys for laboratory research. *ILAR J.* 47, 307–315. doi: 10.1093/ilar.47.4.307

Terborgh, J., and Janson, C. H. (1986). The socioecology of primate groups. *Annu. Rev. Ecol. Syst.* 17, 111–136. doi: 10.1146/annurev.es.17.110186.000551

Therneau, T. M. (2019). A package for survival analysis in R.

Valeggia, C. R., Mendoza, S. P., Fernandez-Duque, E., Mason, W. A., and Lasley, B. (1999). Reproductive biology of female titi monkeys (*Callicebus moloch*) in captivity. *Am. J. Primatol.* 47, 183–195. doi: 10.1002/(SICI)1098-2345(1999)47:3<183::AID-AJP1>3.0.CO;2-J

Van Goozen, S. H., Wiegant, V. M., Endert, E., Helmond, F. A., and Van de Poll, N. E. (1997). Psychoendocrinological assessment of the menstrual cycle: the relationship between hormones, sexuality, and mood. *Arch. Sex. Behav.* 26, 359–382. doi: 10.1023/A:1024587217927

Wickham, H., Chang, W., and Wickham, M. H. (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version, 2(1), 1–189.

Wilke, C. O., Wickham, H., and Wilke, M. C. O. (2019). Package 'cowplot'. Streamlined plot theme and plot annotations for 'ggplot2.

Witczak, L. R., Arias del Razo, R., Baxter, A., Conley, A. J., Cotterman, R., Dufek, M., et al. (2021). Relationships between cortisol and urinary androgens in female titi monkeys (*Plecturocebus cupreus*). *Gen. Comp. Endocrinol.* 314:113927. doi: 10.1016/j.ygcen.2021.113927

frontiers | Frontiers in Ecology and Evolution

TYPE Hypothesis and Theory
PUBLISHED 16 June 2023
DOI 10.3389/fevo.2023.1004384

OPEN ACCESS

EDITED BY
Patrice Adret,
Universidad Autónoma Gabriel René Moreno, Bolivia

REVIEWED BY
Leonardo Francisco Barón Birchenall,
Corporación Universitaria Minuto de Dios, Colombia
David Schruth,
University of Washington, United States

*CORRESPONDENCE
Gisela Kaplan
✉ gkaplan@une.edu.au

RECEIVED 27 July 2022
ACCEPTED 25 April 2023
PUBLISHED 16 June 2023

CITATION
Kaplan G (2023) Evolution of human language: duetting as part of prosociality and cognition. *Front. Ecol. Evol.* 11:1004384.
doi: 10.3389/fevo.2023.1004384

Evolution of human language: duetting as part of prosociality and cognition

Gisela Kaplan*

School of Science and Technology, University of New England, Armidale, NSW, Australia

The evolution of human language is a topic that has received undiminished attention. Numerous hypotheses for the origin of human language have been proposed, including gestural communication found specifically among apes. This study advances the hypothesis that human evolution, including human language development, is three-pronged: prosocial, cognitive, and collaborative. Duetting and turn-taking in primates are used as pivotal examples of how bonding leads to joint action and collaboration. It points out that such vocal behavior itself may be a crucial precursor of language evolution in the sense that it is explicitly focused on a conspecific. Some current hypotheses have acknowledged duetting as an important perceptual and behavioral example of synchronicity. Some forms of synchronized behavior, as found in duetting, synchronized dance, or even shared song, were perhaps crucial evolutionary steps preceding the evolution of human language. Duetting signifies more than that, however, because it is an observable and significant cognitive investment that signals attention toward a partner. This study also advances the hypothesis that affect and cognition would have needed to precede any form of duetting or signs of affiliation such as grooming. Hence, this study, asking what duetting in primates signifies in evolutionary terms, takes a multidisciplinary and multimodal approach to suggest important affective and cognitive steps in the evolution of human language and speech, the chief of which is prosociality. Prosociality, as an attitude and awareness of another, be this as a friend or partner for whom one can do favors or whom one can help, is a model for collaboration and cooperation, and also increased cognition.

KEYWORDS

prosociality, duetting, cooperation, synchronicity, cognition and emotions in primates, human language evolution, communication (vocal, gestural)

1. Introduction

Duetting exemplifies a significant step in the evolution of language for several reasons. It is usually a time-sensitive vocal activity performed by a pair of closely connected individuals. It further requires coordination of vocal production and a degree of vocal flexibility. In duetting, listening is a key element in the switch from self- to other-oriented and affiliative behavior that may signal cooperation on a broader scale (i.e., beyond duetting). Such behavior may be termed "prosocial". Prosociality has often been understood as a main facilitating driver of cooperation (Martin et al., 2021). Coordination in the sense of prosociality, unlike empathy, carries no direct cost to the actor but presupposes a positive attitude toward another and doing things together, even supporting others (Silk, 2007).

Accommodation of the behavior and even needs of others may then develop into a new awareness and affective sensibilities toward others for which a new cognitive framework may be needed. Since duetting is an exemplification of one of the most basic forms of joint vocal action of committed pairs, it will be discussed in light of prosocial tendencies.

The literature seems to agree that during the last 2 million years, hominins had become more and more socially complex animals in comparison to other primates (Dunbar, 2014). According to James Baldwin's insights (called the Baldwin effect) evolution by natural selection occurs in three stages: (1) the appearance of new environmental challenges, (2) the adoption of a new behavior through learning (natural selection favoring cognitive plasticity), and (3) new genetically based predispositions when natural selection favors individuals that exhibit a particular adaptive behavior (Podlipniak, 2017). Certainly, the first stage can be readily reconstructed, i.e., the changing physical environments in which early humans moved (Suzuki, 1970). Africa was drying, vegetation reducing, leaving a band of ill-equipped hominins surviving on the savannah exposed to formidable predators. Confronting such new environmental challenges, as Baldwin argued, would lead to the invention of new behavior and this might have forged how individuals acquired synchronizing behavior and collaboration and probably did so as the best or even only chance of survival (Klein, 1977; Caley et al., 2018).

Baldwin's third stage (new genetically based predispositions when natural selection favors individuals that exhibit a particular adaptive behavior), such as the shift to more prosocial, even verbal communication, may have been a step too far for chimpanzees and even bonobos. Chimpanzees, although many attempts were made, could not be taught, or made to speak (Gardner and Gardner, 1969; Gardner et al., 1989). Hence, acquiring the ability first to be able to articulate sounds in the sequence required physiological and cognitive changes (brain nuclei to process information), memory, and the ability to expand vocal expression. Vocal convergence, in which adjustments to one sound type result in similarities between individuals, occurs in a wider range of mammalian orders including primates, mole-rats, goats, and mice (Janik and Knörnschild, 2021). Duetting is part of these parameter adjustments. Learning for a purpose, most likely for cooperation, might well have been a crucial element that fostered the human species' survival and eventually might also have led to the development of human language.

I argue in this study that new environmental pressures forced the development and expression of innovative and new socio-psychological traits and that prosociality is a key characteristic as a driver for this change. Duetting is used as an example of one possible tipping point toward prosociality and eventually cooperation.

To develop these ideas, the study will first introduce the prosocial hypothesis along with hypotheses on human language evolution and then present duetting as a special case of vocal interactive behavior that leads to cooperation and cognitive expansion, and it finally shows why prosociality has an important place, or might well be the lynchpin, in the evolution of human language.

1.1. Background

The evolution of human language has been of undiminished interest and has been pursued by vastly different scholarly disciplines and, sometimes, these disciplines either do not read each other's conclusions and insights or their respective conclusions expose chasms. For instance, biology-based evolutionary theories and linguistic explanations concerning human language evolution have often been at loggerheads with each other (Hockett, 1959; Cadková, 2015). The alleged uniqueness of the human language proposed by 19th-century linguists was irreconcilable with evolutionary theory. Darwin (1859) certainly outraged Oxford University linguist Friedrich Max Müller who proclaimed that "language is the Rubicon which divides man from beast, and no animal will ever cross it" (cited by Fitch, 2013). Müller was not the only critic. Later, researchers adopted an *a priori* position arguing that primates were incapable of engaging in vocal learning (i.e., did not possess the ability to modify acoustic and syntactic sounds and were unable to imitate sounds and words) and hence primate communication was far inferior to that of humans, and to suggest otherwise was indefensible (Penn and Povinelli, 2007). As recently as 20 years ago, some linguists still expressed the belief that animals only produced sounds whose signal inventories "are limited and not subject to cultural modification" or, more precisely, animals were only able to produce innate sounds (Studdert-Kennedy, 2000). Studdert-Kennedy and Goldstein (2003) further argued that human language is defined by the dissociation of sound and meaning and has no precedence in animal vocalizations. Dissociation is seen as a critical discontinuity that separates human language from other primate systems of vocal communication (Studdert-Kennedy and Goldstein, 2003).

Not surprisingly, despite the controversies about primate linguistic abilities, comparative research into the origins of human language has focussed on the primate line, the closest extant relatives of humans (Fedurek and Slocombe, 2011; Wheeler and Fischer, 2012; Townsend and Manser, 2013; Levinson, 2016; Vonk, 2020). Some primates, particularly apes, actually show a great diversification of communicative acts, from gestures (Liebal and Call, 2012; Hobaiter et al., 2022) to body movements (Gasser and Arbib, 2019), from singular vocal acts to sustained vocal expressions (Liebal and Oña, 2018), and, finally, to joint vocal actions (Sekulic and Chivers, 1986; Baker-Médard et al., 2013) and even "song"; the latter largely limited to gibbons (Geissmann, 2000), Malagasy Indri, *Indri indri* (Maretti et al., 2010; De Gregorio et al., 2019), titi monkeys, of the following three genera: *Cheracebus*, *Callicebus*, and *Plecturocebus* (Adret et al., 2018; Aldrich et al., 2023), and Sulawesi tarsiers (MacKinnon and MacKinnon, 1980; Clink et al., 2022). These various and diverse examples of communicative behavior in primates have provided a rich canvas as starting points for human language origins, be they initially gestural or vocal (Deacon, 2003).

Indeed, theories of language evolution have proposed a vast range of different possibilities, be this via gestures, music, and rhythm (alternatives to be discussed later) but the puzzle remains how the switch from non-speaking great apes to speaking humans could have occurred. We know now that apes can form concepts and abstract ideas concerning the passage of time (Patterson, 1978). Through American Sign Language, ample evidence has been accumulated that apes use this means of communication to create new meanings, invent new signs, and combine words in ways that create a message (Miles, 1990, 1994). And Koko, a gorilla, showed that he was able to remember past events and plan or imagine the future (Patterson, 1978). Experience, memory, and learning produced new outcomes. Barton (1998) and later Barrett and Henzi (2005) explained that, as primates formed larger and socially more cohesive groups, their perceptual system needed to be enhanced to process details of dynamic social stimuli, such as facial expression, posture, gaze direction, and the like (Barton, 1998).

Significantly, research has shown that proto-language or gestures in great apes are mapped to specific areas of the brain used in human language such as Wernicke's and Broca's areas (Cantalupo and Hopkins, 2001). Interestingly though, the greatest expansion of the primate brain over evolutionary time apparently occurred in the visual cortex (in particular, area V1; Das and Gilbert, 1995), in the parvocellular region, which is associated with the analysis of fine detail and color in diurnal primates (Harting et al., 1973; Smaers and Vanier, 2019). Also largely located in layer V1 of the visual cortex are the recently discovered spindle cells, probably unique to great apes and the human brain (Banovac et al., 2021). There is some evidence that specialized spindle cells project to highly specific motor centers "controlling vocalization, facial expression, or autonomic function" (Nimchinsky et al., 1999). Perceptual abilities and the ability to discriminate vocal and facial expressions are certainly of benefit when subtleties in communication and an understanding of the emotions and intentions of others are increasingly important. Gesturing is a non-linguistic act but, as had been shown time and again, it can convey very specific meaning. Pointing, in particular, has often been identified as a key behavior for understanding the development of language and theory of mind (Camaioni et al., 2004). More of this later. Recent work has also identified a "primate mosaic brain evolution" (De Casien and Higham, 2019). The authors concluded that these were in the area of sensory and cognitive specializations that enabled effective communication even at a non-linguistic level (De Casien and Higham, 2019).

The apes' proven physiological inability to speak required morphological changes. Such changes included the dropping of the larynx before speech could occur and this led to humans' ability to speak (Lieberman, 1985). This theory had lost some traction in favor of suggesting different processes. Nishimura et al. (2022), for instance, have now shown that important physiological changes did occur but in an unexpected direction. The adaptations involved a process of *shedding* anatomical features of the vocal apparatus via structural simplifications: the laryngeal air sacs of great apes disappeared (Trenbeath, 2021) and as humans evolved, they also lost the standard primate laryngeal feature of thin upward projections of the vocal folds, and they considered these the crucial adaptations for speech (Nishimura et al., 2022).

As recent research has shown, however, some primate and avian vocal abilities are far more complex and varied than once thought (Kaplan, 2014), starting, in primates, with the discovery of referential signals in vervet monkeys, *Chlorocebus pygerythrus* (Seyfarth et al., 1980; Seyfarth and Cheney, 1986), continuing with the discovery of referential food grunts in chimpanzees, *Pan troglodytes* (Watson et al., 2015), food calls in common marmosets, *Callithrix jacchus* (Rogers et al., 2018), and the vocal modifications found in pygmy marmosets, *Cebuella pygmaea* (Snowdon, 2018). Great apes and even new world monkeys (such as black-fronted titi monkeys, *Callicebus nigrifrons*: Caesar and Zuberbuehler, 2012; and white-faced capuchin monkey, *Cebus imitator*: Coss et al., 2019) have been shown to use referential gesturing and vocalizations. Indeed, the detailed linguistically based studies of the 1980s and 1990s confirmed that apes were able to learn American Sign Language (Gardner et al., 1989; Miles, 1994). They understood words, commands, and objects, even showed some sense of grammar (Greenfield and Savage-Rumbaugh, 1990), and were able to count (Boysen and Bernston, 1989). This was confirmed for all four great ape species (bonobos, *Pan paniscus,* and chimpanzees, *P. troglodytes*: Savage-Rumbaugh, 1984; gorillas, *Gorilla gorilla*: Patterson, 1978; orangutans, *Pongo abelii*: Miles, 1990). While some avian species trump some of the primate skills (from chickens being able to count (Rugani et al., 2011) to the ability to understand speech (Pepperberg, 2007), the point here is that there are many precursors to human language evolution, be this conceptually and semantically, and thus cognitively already present in primates (Lameira, 2017).

By the 1990s, experts in the field spoke openly about the "minds" of great apes, rather than about "cognition" (Russon et al., 1996). In primates, this abstract ability to be able to deal with symbolic representations of language and thus display complex cognitive processes led to a host of detailed investigations both of behavior and of the structure and function of the primate brain (Maestripieri, 1999; Reader and Laland, 2002). Such investigations and comparative studies between primates and humans continue to this day and have clarified differences (Palomero-Gallagher and Zilles, 2019) and similarities (Miller et al., 2021) between the brain of apes and the human brain. The discoveries of mental time travel (conceiving of past, present, and future) conveyed in sign language added depth to the view that apes are cognitively very advanced, can readily cope with abstract concepts, and imbue gestures with meaning (Leavens, 2004; Liebal and Call, 2012; Fröhlich and Hobaiter, 2018; Hobaiter et al., 2022). Cognitive features of primate behavior, such as cooperation, have also been identified as essential qualities for human language evolution (Williams et al., 2022).

2. The prosocial hypothesis

Prosocial behavior has long been of central concern and research interest in human psychology, partly because adolescents who show weakly developed prosocial behavior tend to display several behavioral problems (Card et al., 2008; Carson, 2013). The prosocial hypothesis proposed here is that human evolution and human language development depend on a three-pronged model of key pillars: prosocial, cognitive, and collaborative actions. It is not a combination of those three elements but a sequential development, i.e., of prosocial behavior leading to sharing of cognitive insights and eventually collaborative actions. These advances tended to offer or help solve a range of environmental and inter- or intra-group challenges. The argument of the prosocial hypothesis is well in line with other hypotheses of complex, often multilevel social structures (Cronin, 2012; Sewall, 2015; Aureli and Schino, 2019; Kappeler et al., 2019; Morrison et al., 2020), communication, cognition (Sewall, 2015), and, importantly, cooperation (Jaeggi and Gurven, 2013) as drivers of evolution. They also fit well with the human self-domestication hypothesis (Hare et al., 2012). As Hare (2017) states: the human self-domestication hypothesis entails (a) selection for prosocial behavior linked to derived human cooperative-communicative abilities and (b) the domestication syndrome in human morphology, physiology, development, and cognition, as seen in other self-domesticated species (such as dogs).

However, there are some contradictory and unresolved problems between the studies of primatology and anthropology. The occurrence of prosociality in animal studies has spawned two main hypotheses, called the cooperative breeding hypothesis and the self-domesticated hypothesis. According to Amici et al. (2014), the cooperative breeding hypothesis, at least in primates, predicts *low levels of prosociality* when specific species are *non-cooperative* breeders, while the self-domestication hypothesis predicts high levels of prosocial behavior because self-domestication presumes high levels of tolerance of each other (Amici et al., 2014). Humans and callitrichid monkeys are the only primate species described as cooperative breeders, so they should show high levels of prosocial behavior and they do (Martin et al., 2021). All great apes should show *low* levels of prosocial behavior as Amici and colleagues found when they tested chimpanzees, bonobos, gorillas, orangutans, tufted capuchin monkeys (*Sapajus apella*), and Geoffroyi's spider monkeys (*Ateles geoffroyi*). Indeed, Amici and colleagues found little to no prosocial behavior in any of the great apes and New World monkeys they studied. This very much runs counter to other research results but, importantly, also to hominin evolution that has argued repeatedly that prosociality, indeed the human ability to support each other, is an essential precondition for the success of humans, perhaps the main reason why this species survived and thrived (Hare, 2017).

The results by Amici et al. (2014) showing little evidence of prosocial behavior especially in the four great ape species may be explicable by different circumstances and housing as well as gender. However, their results have been duplicated. Three years after the publication of their results, Verspeek et al. (2022) conducted experiments with bonobos and equally found no evidence of prosocial behavior, confirming the results and conclusion of Amici et al. (2014).

However, these results run counter to the prediction that self-domesticated primates should show *high* levels of prosocial behavior. The anthropological literature on human evolution from the Lower Paleolithic (ca 1.5 million to 200,000 years ago) to the Holocene Epoch (11,700 years ago to the present) periods strongly argues that later humans are self-domesticated (one hypothesis of prosociality) and, by the time of the Holocene, show strong prosocial behavior. The human self-domestication hypothesis (HSD) (Hare and Tomasello, 2005; Hare et al., 2012; Hare, 2017) seems very convincing and supports evolutionary trends also in other species, especially dogs (Hare, 2017). The assumption is, of course, that the nearest relatives to early humans, chimpanzees and bonobos, should share the same traits of prosociality and high levels of mutual tolerance or even spontaneous altruism, as has been described in humans, and certainly high levels of tolerance were found in children and chimpanzees (Warneken et al., 2007; Warneken, 2015). Such incompatible results give at least pause for thought.

Equally, evolutionary theories, such as Darwin's and Baldwin's, suggest that environmental pressures led to the invention of a new behavior by means of learning (natural selection favoring cognitive plasticity) and gradually an increase in cognitive abilities in humans. However, more social pressure does not always require more cognitive ability but can lead to more subdivision of tasks and a lowering of individual cognitive ability (Fedorova et al., 2017).

The first imperative would seem to be that individuals had to bond with conspecifics in some social way. The social brain hypothesis Dunbar proposed in the late 1990s (Dunbar, 1998) was at first designed to explain why primates had unusually large brains for body size compared to all other vertebrates: He attributed this to their complex social system but later he extended this hypothesis to human evolution (Dunbar, 2014). The social brain hypothesis that Dunbar developed largely seemed to explain the expansion of cognitive abilities particularly in the primate line and chiefly in chimpanzees. It did not necessarily explain the evolution of complex communication and prosociality in humans until the human self-domestication hypothesis was developed and tested, showing that apart from the physical, physiological, and other changes, self-domestication selected for high prosociality (Cieri et al., 2014; Hare, 2017).

3. Hypothesizing the evolution of human language

The prosocial hypothesis advanced in this study does not conflict with the social brain hypothesis (Dunbar, 1998) or the hypothesis of a gestural origin of human language (Corballis, 2002, 2010). Both, as well as several others, rightly emphasize the gradual sequencing of psycho-social developments, including nuances of communicative behavior. The term "communication" is chosen deliberately here. In agreement with Fitch (2020), even sophisticated, vocalized (referential signaling), or verbalized (human speech) communication does not address the cognitive richness of concepts that may or may not be expressed in words and may not leave measurable behavioral evidence. Also, the multifarious, at times instantly changeable, and flexible interactions between environment and organism need to remain firmly in view. We know that extant apes are capable of distinguishing gestures, facial expressions, and vocal information and, presumably, so was the hominid and hominin brain.

Even gene expressions can change relatively quickly. Wiles et al. (2005) gave as an example the genetic ability of mammals to synthesize vitamin C in the body. But in primates, by a process called genetic redistribution, this gene expression was eliminated so that, from then on, the only way to acquire vitamin C had to occur exogenously. A second example, a purely morphological change, was provided by Darwin's Galapagos finches. Darwin concluded that consistent environmental differences in different habitats in the Galapagos promoted *directional natural selection* on resident finches for optimal beak morphology. This process has produced more than a dozen distinct species of finches, all unique to the archipelago, further cementing Darwin's idea of natural selection (Grant, 2017).

Thus, in psycho-cognitive developments, both behavioral synchronicity (the ability to match the behavior of another, be this in movement, sound, and mood) and prosocial inclinations need to *precede* the development of intentional acts toward conspecifics (see Table 1) and create a niche for enhancing cognitive abilities in what has been termed "emotional intelligence" (Salovey and Mayer, 1990). Communication is a very important part of this but so is finding a *reason for* extended communication, namely the emergence of "other-directedness", of the importance of a partner or a group for one's survival.

Table 1 should be read from left to right as a cumulative and dynamic development toward prosocial and affiliative behavior. Note that the Australian shingleback lizard (*Tiliqua rugosa*) is a monogamous lizard but the pair separates outside the breeding season. Only those species are included here with pairs also staying together outside the breeding season and that particular condition alone limits the number of species included in animal bonds. Table 1 also indicates that the life history data of species, including their reproductive strategies, are important variables.

The point of this study revolves largely around two main social characteristics of any form of cooperation (one is biparental care and the other is prosociality) as two imperative milestones in the evolution of human bonding (Launay et al., 2016) and human language. However, biparental care in humans has a slim evolutionary base. When, for instance, examining reproductive strategies in fish, reptiles, and amphibians (Andrew-DeWoody et al., 2000), or even mammals, the number of species across classes of animals remaining paired for years is very small indeed. While Table 1 has identified species from marine life to a broad range of land animals, pair bonds, let alone monogamous life-long bonds, are overall very rare in any vertebrates, except for birds. In birds, at least 95–97% of more than 10,000 avian species pair bond with a mate and jointly raise their offspring (Cockburn, 2006).

In mammals, according to Clutton-Brock (1991), only about 5%, including some primates such as marmosets (Burkart and van Schaik, 2020; Martin et al., 2021), meerkats, wild dogs, and certain species of mice, form lifelong pair bonds or even short-term pair bonds and practice biparental care. But the 5% of mammals that practice biparental care still tend to live in troops, groups, prides, or packs, in which the breeding pair typically consists of the alpha male and the alpha female. Hence, the social configuration of pair bonding of two humans and the evolution of complex communication systems, including language, in humans, have few direct evolutionary predecessors, and, with some exceptions, the various elements required for creating a prosocial context are often not in the one species together.

Birds and humans thus have in common that they both raise their offspring as pairs (biparental care) or raise offspring cooperatively and even join forces in group defense. Cooperation and bonding in hominin evolution may not be an innovation *de novo* but evidence of such social relationships and task coordination still offers challenges to our understanding of their developments, be this in humans or birds (Issa et al., 2023).

Great apes generally provide many variations in mating and alliance systems, however, making meaningful comparisons with human society more difficult. Of course, the mating system of a species does not always mirror its social system (Dixson, 2009). By and large, orangutans are solitary (Kaplan and Rogers, 2000), western lowland gorilla (*Gorilla gorilla gorilla*) groups with several females and offspring are usually ruled by a single silverback (Forcina et al., 2019), and, with some variations, mountain gorillas, *Gorilla beringei beringei*, although classified as one male group may be up to 40% multimale groups (Robbins, 1999; Morrison et al., 2020). Chimpanzees live in multimale and multifemale social groups and may strongly compete with and aggressively fight other groups, and bonobos have a matriarchal system (Sommer et al., 2011). At some stage, the human social organization may have been the closest to that of gorillas, living with them in forests (White et al., 2009) or having moved to savannahs (the oldest established hypothesis on human bipedalism; Senut et al., 2018) or, as has also been suggested, living largely near water and exploiting its resources (Stewart, 1994; Finlayson, 2014). As Schacht and Kramer (2019) noted recently, consensus on a human-typical mating system remains elusive. "While a simple classification would be useful for cross-species comparisons, monogamous, polyandrous, and polygynous marriage systems exist across contemporary human societies" (Schacht and Kramer, 2019).

In discussing various mating systems, respective benefits for offspring are worth mentioning here. For instance, there is evidence that stable, socially monogamous pairs or stable small family groups in whatever species or class of animal create a safe and largely stress-free emotional and learning environment (Raposa et al., 2016) strengthening survival and long-term health. In many species with these characteristics, there is also an extra social layer—that of socializing juveniles. As I have explored elsewhere (Kaplan, 2020a, 2023), such environments encourage extensive social play behavior that is intimate, communicative, and creative (Bateson and Martin, 2013), and this is usually regarded as beneficial for the individual concerned. Whatever one might call the effects of play: they are now recognized as generating positive emotions (rats: Panksepp, 2005; Pellis and Pellis, 2007; Vanderschuren et al., 2016; ravens: Osvath and Sima, 2014; primates: Loizos, 2017). Positive emotions are themselves reinforcing to seek similar contact in future. Hence, regardless of how social interactions proceeded to evolve into human language—be this via gestures, music (song and dance), drumming, whistling, or extension of referential vocal signals—it required motivation first to even get to a position of seeking expansion of any form of communication.

Second, Table 1 is meant to emphasize the centrality of evolving prosocial behavior in the formation and maintenance of strong affiliative bonds. "Prosociality", as already mentioned, has been highly topical in the field of psychology for some time (Luengo-Kanacri et al., 2021), especially in human developmental studies (Spataro et al., 2020). The social circumstances for the absence or presence of prosociality (which may be variable and flexible) continue to be explored, especially the consequences when a well-defined profile of prosociality is absent or weakly developed (Donald et al., 2021). However, its role in animal communication and bond strength (a) in flexibly functioning pairs, families, and animal communities and (b) as a trigger for the intentional sharing of goods, such as food (Feistner and McGrew, 1989; Jaeggi and Gurven, 2013; Güroglu et al., 2014), in communication and

TABLE 1 Types of animal bond.

Prosocial behavior											
Taxonomy		Species	Pair-living breeding	Pair-living non-breeding	Maintain proximity	Joint territorial defense	Coordinated behavior	Biparental care	Affiliative behavior	Family defense (close-knit group)	Stress buffering
Invertebrates	Mollusca	Giant false limpet	*	*	*						
	Anthropoda	Snapping shrimp	*	*		*					
		Giant wood cockroach	*	*	*	*					
		Termites	*	*	*	*					
Vertebrates	Fish	Butterfly fish, Goby cleaner	*	*	*	*	*				
	Reptiles	Skinks	*	*	*	*	*	*			
	Amphibians	Poison dart frogs	*	*	*	*	*	*			
	Birds	Geese, Swans	*	*	*	*		*			
		Blue ducks	*	*	*	*		*			
		Cockatoos, Mackaws, Keas	*	*	*	*	*	*	*	*	
		Parrots	*	*	*	*	*	*	*	*	
		Song birds	*	*	*	*	*	*	*	*	
	Mammals	Prairie voles	*	*	*	*	*	*	*	*	*
		California mice	*	*	*	*	*	*	*	*	
		Wolves, Jackals	*	*	*	*	*	*	*	*	
		Marmosets/Tamarins	*	*	*	*		*	*	*	
		Titi monkey	*	*	*	*	*	*	*	*	*
		Great and lesser apes^	^	*	*	*	*	^	*	*	*
		Humans	*	*	*	*	*	*	*	*	*

Prosocial tendencies increasing.

^Great apes do not form pair bonds but are included here because they form long-term group bonds and males protect infants [what Wrangham (1979) called "permanent consortships"]. Gibbons (genera *Hylobates, Nomascus,* and *Hoolock*) and the siamang (genus *Symphalangus*), however, are the only hominoids to exhibit pair bonds and two-adult groups (Fuentes, 2000).

The double line is a demarcation line that indicates the emergence of prosocial behavior which is expressed as affiliative behavior.

ultimately in human language evolution is rarely considered. This is surprising, given the question: why are we the only primate that can speak? remains an open question.

I am suggesting a domino effect from joint action to prosocial tendencies, generating more detailed communication, leading in turn to an increase in differentiated acts of communication in both referential signals (vocal and gestural) as well as semantic content. Another point to be made here is that even the most occasional acts of prosocial behavior in great apes mean that prosocial behavior is, and most likely was, an option in primate culture.

Furthermore, prosocial behavior is causally linked to the evolution of human language because language is more than a linguistic manifestation. It is a tool for a *continuing motivation* to address a conspecific or partner. To achieve and maintain such motivation, both emotional and cognitive complexity needs to have developed and, if already present, increased further. In agreement with the dynamic systems paradigm (see Shanker and King, 2002; King, 2009), converging feelings and intentions among partners may continue to be enhanced in a dynamic of ongoing negotiation at inter- and intra-personal levels, leading perhaps to closer bonds. The latter is a claim of the involvement of emotions, recently discussed by Dukes et al. (2021).

Another point that at times seems to have been lost in debates between selfish and prosocial actions within pairs and groups of primates is to consider evolutionary principles: Whatever format of skills, communication, or affiliations is more sustainable, these traits are more likely to survive via natural selection. In some cases, they may even develop further, be this at the cellular level, in morphology, physiology, or even chemistry. Exhuberant morphological features are generally associated with food acquisition. Well-known examples are the elongated middle finger of one of the Madagascar's nocturnal lemurs, the Aye-Aye, *Daubentonia madagascariensis* (Sterling and McCreless, 2007), or the exaggerated beak length of the sword-billed hummingbird, *Ensifera ensifera* (Abrahamczyk et al., 2014), or, as Darwin described, the diversification of beak strength in finches in different environments. While the finch model of natural selection is well-known and can explain so many other variations in biology, it should be applied rigorously to behavior because the same evolutionary principles ought to apply.

4. Duetting

4.1. Characteristics

Duetting, a vocal manifestation of synchronicity, is one of the most studied vocal behaviors in mammals and birds and occurs in many forms, referred to as antiphonal singing, turn-taking, counter-calling, or counter-singing. Some of these exchanges are expressed between males. Whatever the dyadic composition, most interactions between two members of the same species are between male and female partners.

In the broadest sense, duetting and counter-singing are vocal behaviors that exist in many songbirds, in some primates, but also in Alston's singing mouse, *Scotinomys teguina* (Neff, 2019), Klipspringer antelopes, *Oreotragus oreotragus* (Tilson and Norton, 1981), the maned wolf (Ferreira et al., 2022), whales (sperm whale, *Physeter macrocephalus:* Schulz et al., 2008; long-finned pilot whale, *Globicephala melas:* Courts et al., 2020; reviewed in Vanderhoff and Bernal Hoverud, 2022), amphibians (chorus frogs such as spring peepers, *Pseudacris crucifer;* Forester and Harrison, 1987; south African clawed frog, *Xenopus laevis*: Tobias et al., 1998; Legler's stream frog, *Ptychohyla legleri*: Etzel et al., 2020), toadfish, *Tetraodontidae* (Vieira et al., 2021), and even in a range of invertebrates (Bailey, 2003; Henry et al., 2013).

Duetting in the narrowest sense is defined as a temporally coordinated interactive vocalization between two adults, usually of established pair bonds. Such vocal exchanges tend to have specific temporal patterns and may overlap even substantially while, in birds, few or no overlaps occur. Taking turns, as Banerjee and Vallentin (2022) noted, requires a fast sensory perception of the sender's vocal output but also the precise control of the responder's vocal onset. During these interactions, participants simultaneously plan upcoming vocalizations while listening to respond as early as possible without interrupting the initiator of the duet (Levinson and Torreira, 2015; Banerjee and Vallentin, 2022). Many avian duets fit into this characterization. Duets consist of calls or syllables in rapidly produced vocalizations and even these can be void of specific meaning (Arriaga and Jarvis, 2013; Dahlin and Benedict, 2014; Barón Birchenall, 2016). Any of the turn-taking vocalizations can be defined as an orderly exchange of communicative vocal signals that may or may not overlap.

However, among those primates that are mated pairs in stable monogamous bonds and are in stable territories, duetting is a rare social phenomenon and involves clear-cut examples of closely temporally matched vocalizations. We know only of a few diverse primate families—Tarsiidae, Indriidae, Lemuridae, Hylobatidae, Cercopithecidae, and Pitheciidae—to which these conditions apply (e.g., Tarsius spp.; indri *Indri indri*; Mentawai langur, *Presbytis potenziani*; and gibbons, Hylobatidae). After studying the duetting and vocal behavior of some of these taxa Haimoff (1986) concluded that the occurrence of duetting in these primate species and the similarities found in the acoustical features of their vocal behavior, represented a case of functional convergence. Such convergence was possibly a result of their evolution of a common social organization or similar ecological niche (Haimoff, 1986). To my knowledge, this conclusion has not been challenged to date.

Duetting can have several functions, some of which might even be present in one single species (Dahlin and Benedict, 2014). These are mate-guarding (Dowling and Webster, 2018; Dolotovskaya et al., 2020), to signify and or strengthen partnerships (Mèndez-Càrdenas and Zimmermann, 2009; Smith et al., 2010; Singletary and Tecot, 2020), and may serve as an indicator of the presence of a well-versed territorial defense team that may send a warning to potential intruders (Adret et al., 2018; Amorim et al., 2022). In sperm whales, Schulz et al. (2008) studied the frequent exchanges of short sequences of clicks, termed codas. They found that the sequencing of exchanges into duet-like chains with overlapping and matching functions reinforced social bonds between whales, which is attributed to the same or very similar function to duetting as in primates or birds.

The act of duetting also seems to have some measurable, "feel-good" consequences for the participating partners, be this in hormonal changes in oxytocin and vasopressin and increased brain-to-brain synchrony in frontal and pre-frontal brain areas

(Amodio and Frith, 2006; Reindl et al., 2018), confirmed in bats (Zhang and Yartsev, 2019; Rose et al., 2021), primates (Smith et al., 2010), and human studies (Atzil et al., 2012; Bales et al., 2021).

We thus have some cumulative evidence that duets are largely partner and pair dependent and contribute to the bond, be this in inhibition driven by auditory feedback (Coleman et al., 2021: plain-tailed wrens, *Pheugopedius euophrys*) or in very precise timing but different frequencies (Hoffmann et al., 2019: white-browed sparrow-weavers, *P. mahali*). The evidence also suggests that, over time, coordination of duetting improves in timing and auditory adjustments to the partner's specific auditory characteristics of their part of the duet. In my research of duetting in wild free-ranging magpie larks (*Grallina cyanoleuca*), the duets I recorded of a local pair in coastal New South Wales, Australia (Coordinates 30.5869° S, 153.0001° E), were not just timed precisely but the segments of each partner were near identical (Figure 1). In one of the rare longitudinal studies of duetting, Hall and Magrath (2007) showed that, in magpie larks at least, duets in newly established pairs were not precisely timed and their vocalizations would even overlap. By contrast, in well-established pairs, timing became very precise in all measures. Presumably, a potential territorial invader can audibly ascertain whether a pair is well-established and has perfected the art of territorial defense or the pair was newly formed and relatively inexperienced in which case its territorial claims could be challenged. In this avian species at least, duetting has a dual function as a form of mate-guarding and as a warning for potential intruders that they are dealing with well-experienced pairs (Vanderhoff and Bernal Hoverud, 2022). In most cases, the coordination of a song tends to have a leader and a follower. The partner who maintains the rhythm becomes the leader and the partner who maintains the synchrony of the joint behavior becomes the follower, arguing that maintaining synchrony requires greater adaptation (Hoffmann et al., 2019).

Interestingly, in one of the larger nocturnal sportive lemurs (*Lepilemur edwardsi*) that Smith et al. (2010) studied, pair partners synchronized behavioral activities, especially *after* duetting. In other words, duetting is not an isolated skill but one that, in mammals and birds at least, is a well-evolved expression of social rules and bonds. The latter may readily lead to ever-increasing invention of sound symbols, i.e., sounds with semantic meaning (Vonk, 2020).

4.2. The relevance of duetting to human language evolution

Much has been made of the gestural origin of human language and for good reason. In apes, some 80 gestural referential signals have been identified (Leavens and Hopkins, 1998). As was mentioned before, apes trained in American Sign Language were able to show human researchers that they were capable of thinking of the past and the future (theory of mind), and of being linguistically innovative by making new combinatory words and even sentences (Corballis, 2002, 2010; Hobaiter et al., 2022). These discoveries were significant in showing that concepts and theory of mind existed in apes before the evolution of human language and that these were applied intentionally and directed toward another individual or group.

One might argue (with some justification) that duetting is a very weak link to human language evolution especially when compared to the rich conceptual and symbolic range of ape gestures. With some exceptions (Clarke et al., 2006; Andrieu et al., 2020), duetting tends not to carry complex and personal messages as gestures can. But this is not the point. Lifting out any vocal behavior in extant species is providing a *static* snapshot of how and how far each species has taken its cognitive and affective abilities.

The question is why an expanded need for more communication arose in the first place, what its motivation was, and in what specific social context vocal communication eventually arose. Equally, the question remains as to why language as speech had to come about at all. A static snapshot may discover the extent of the cognitive achievements of a species, but it needs an evolutionary, dynamic perspective to address the question as to why and how vocal signals developed to the extent to which they did in humans.

It is generally agreed that biological changes can be due to mechanisms such as natural selection, random genetic drift (Santangelo et al., 2018; Miles et al., 2019), sexual selection (Kuijper et al., 2012), and other extraneous events or features (such as climate change: traditional food sources dwindling and changes in environmental topography; Veit, 2021). Such changes are responses and adaptations vital for enhanced chances of survival.

The changes that occurred in the hominin brain are structurally and functionally substantial. After investigating the differences in the brains of chimpanzees and humans, Ardesch et al. (2019) concluded, "…[our] findings suggest an evolutionary shift in the human brain toward investment of neural resources in multimodal connectivity facilitating neural integration, combined with an increase in language-related connectivity supporting functional specialization". The question is what possible internal or environmental factors could have made this happen? And how could language acquisition be achieved within the organisms' own biology and available social skills and resources?

In this regard, duetting is an important milestone, even if only shared by a few species among primates. This is not related to the less than frequent vocal displays of duets but for another reason: duetting can show the very point when adaptive behavior, that initially might have evolved for ecological reasons, can flip onto a cognitive and affective plane. First, unlike transitory mimicry of movement or sound, courtship dance rituals, or pre-copulatory synchronicity to *establish* a bond or common interests, this kind of synchrony investing in cooperative behavior means that such bonds have *already been established*. Such specific ongoing bonding practices may lead to further expressions of cooperative, prosocial, and even empathetic behavior (Hove and Risen, 2009). This is so because the partner has become a "significant other" and is given careful attention.

Clearly, the longitudinal study of magpie lark duetting, cited above (Hall and Magrath, 2007), showed that learning was involved when the duetting signals matched more closely after a year

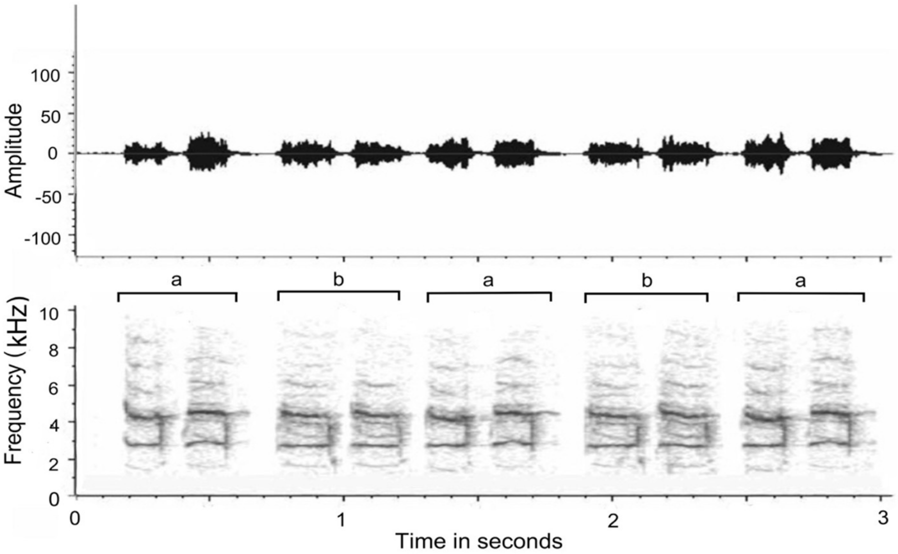


FIGURE 1
Waveform envelope **(Top)** and spectrogram **(Bottom)** of magpie lark duet (a) male; (b) female (author's recording). Note that the morphology of each call is matched almost exactly by the partner and the time intervals between initiated call and reply is reduced to split seconds. The miniscule time delay between a and b was not due to great distance or hesitancy by the replying bird but a function of the position of the remote-control microphone (closer to male) and wind direction (toward female).

than they did initially. Mastering precision requires close listening to and comprehending the other's rhythm, tempo, frequency, emphases, and even length of the duet. Nuances of duetting can vary substantially in terms of developmental plasticity (Adret, 2022), and calls can be sophisticated and distinct in expression (Clink et al., 2021) or may not seem sophisticated at all but are nevertheless significant as a collective behavior (Logue and Krupp, 2016).

This ability to create precise duetting is well-supported by identified brain mechanisms that allow such processes to occur. For instance, Okobi et al. (2019) pointed out that acoustic communication such as duetting often requires rapid modification of motor output in response to sensory cues. When they examined the vocal exchanges in Alston's singing mouse *Scotinomys teguina*, they found that males could modify their singing behavior on a sub-second time course that resembled "both traditional sensorimotor tasks and even conversational speech" in humans.

Two summary points about duetting can be made here in relation to the concept of synchronicity. First, duetting is just one manifestation of synchronicity, if a powerful one when the communication is intentional and practiced. Second, duetting is overwhelmingly found in stable and long-term relationships (Dahlin and Benedict, 2014). In pair duetting, "tuning in" to the bonded partner more than suggests that there is some flexibility both to innovate, learn, and adjust and to fit more directly with the vocal expressions of the bonded partner (Haraway and Maples, 1998; Oller and Griebel, 2008, 2021). Given these sustained observations, it becomes more plausible to suggest that some types of vocal behavior can lead to complex sociality and cognition (Roberts and Roberts, 2020).

5. Beyond synchronicity and toward cooperation

Synchronizing, as discussed above, denotes the precise timing and coordination of movements to coincide with those of another (Bernieri and Rosenthal, 1991). Coordination is socially not far removed from synchronizing behavior and thus plays a fundamental role in social interaction (Yu and Tomonaga, 2015), and such coordination can be a crucial step toward voluntary, intentional cooperation (Valdesolo et al., 2010; Michael et al., 2020). Unlike courtship dances or pre-copulatory synchronicity to *establish* a bond or common interests, this kind of synchrony investing in cooperative behavior presumes that such bonds *have already been established*. Past research has shown that synchronicity can also be tested behaviorally because it is interactional and observable (Hoehl et al., 2020).

5.1. Cognition and emotions

Results of many studies confirm that the brain of great apes and hominins, while expanding substantially from earlier primates (Smaers et al., 2017), did not do so uniformly, identifying some high-expanding areas within the forebrain (Sneve et al., 2019). According to Sneve et al. (2019), especially the brain of *Homo habilis* marked transverse expansion of the cerebrum and the frontal and parieto-occipital parts, and increases in the mass of the frontal and parietal lobes and the two major cerebral areas governing spoken language (Tobias, 1987). One notes also, that while brain mass increased, estimated body mass did not change appreciably (Table 2).

TABLE 2 Mean endocranial capacity and body mass for select hominins.

Species	Mean endocranial capacity (cm^2)	Estimated body mass (kg)
Pan troglodytes (chimp.)	395.0	45.0
Homo habilis	640.2	48.0
Homo erectus	937.2	53.0
Homo sapiens	1,350.0	57.0

Data excerpted from Tobias (1987).

Such an increase in neocortical neurons comes with a high metabolic cost. Sneve et al. (2019) believed that the "capacity of high-expanding cortex to connect flexibly with various specialized brain 'networks' suggests an involvement in 'supramodal' cognition". Whatever is implied in this statement, it is clear that some of these expanding cortical areas are associated with language function in humans. For instance, both in humans and extant great apes, strong asymmetries are present at the population level in the frontal cortex, including a left hemisphere dominant asymmetry of the planum temporale, and in the brain region of Wernicke's area (Figure 2), which supports a critical component of speech production (Gannon et al., 1998; Hopkins et al., 1998; Spocter et al., 2010). Also, the sulci within the inferior frontal cortex, which contains Broca's area, displays left hemisphere dominant asymmetry in both humans and great apes (Sherwood et al., 2008; Hill et al., 2010). Both areas are specific to language and speech. Such patterns of select cortical expansions happened also during human evolution (Hill et al., 2010). In other words, the primate brain was already rather well-equipped to handle cooperation and coordinate activities in ways that required cognitive flexibility.

Those substantial expansions of some brain areas have come at a cost, however. The brain has been described as the most "expensive" part of the body (called "the expensive brain hypothesis", see Isler and van Schaik, 2009), demanding substantially higher energy input than the rest of the body. The "expense" is one of the nutritional requirement because neurons use up to 10 times more energy than body cells (Yu et al., 2014). We know from humans and mammals that energy consumption in the brain accounts for over 20% of total oxygen metabolism (Watts et al., 2018) and neurons consume 75–80% of energy produced in the brain (Hyder et al., 2013). The expensive brain hypothesis argues that the increased length and difficulties to raise an offspring usually lower the number of offspring that can be raised, which can lead to a creeping extinction, a process whereby replenishment of offspring falls below the death rate.

Such metabolic and cytoarchitectural changes in the brain would likely have occurred only if (a) there were substantial evolutionary pressures for new adaptations, (b) the "cost" could be offset by some external compensatory benefits and action, i.e., co-opt others to help protect and raise offspring, and (c) incurred benefits including higher survival rates of self and offspring. To have some negotiated position with a partner, family, or group to feed and care for an individual for a long period also raises the stakes as to the quality of social bonds and responsiveness to a partner—any close social bond or commitment thus involves the communication of some kind, creating a fertile social framework for the expression of emotions and the expansion of cognitive abilities.

Older theories of animal behavior tended to imply, influenced by the views of the French philosopher René Descartes (1596–1650), that animals were mere automata without minds, morality, language, or general intelligence (Thomas, 2020). In this instinct-dominated model, any behavioral expression by an animal was not based on choice but was elicited by a present stimulus that determined the frequency and form of the response. The behavioral form is the same from episode to episode of its elicitation and across animals of the same kind (Epstein, 1982; Miller, 2013). Against the affect-based theories, Richard Lazarus had argued from the 1960s onwards (Lazarus, 1982) that cognitive processes *precede* emotional ones, establishing a clear link between cognition and emotions well before neuroscience could confirm the brain processes involved. He argued that cognitive processes generate, influence, and shape the emotional response in every species that react with emotion.

According to the Lazarus doctrine, cognition is not a postscript to emotions, but for any species, no matter how limited its cognitive abilities may be, any event or encounter in the environment undergoes some evaluative process first. This suggested that most organisms, as far as tested, should come with an array of cognitive skills. This has since been confirmed experimentally. For instance, tests of young chicks have shown some abilities to form abstract concepts using geometrical cues (Vallortigara et al., 1990; Tommasi and Vallortigara, 2004). Indeed, young chickens were found to come equipped with a "package" of conceptual skills in geometry, physics, and mathematics (Vallortigara et al., 2010). Among non-vertebrates, similar cognitive skills were identified. For instance, bees can acquire the ability to deal with conceptual relationships such as "above" and "below", "same," "different," "larger than," and "better than," among others (Avarguès-Weber et al., 2011; Avarguès-Weber and Giurfa, 2013), and were recently shown to solve numerical cognition tasks (MaBouDi et al., 2021), but they may do so using quite different neural processes than birds or other vertebrates (Kaplan, 2015), and octopuses have multiple cognitive abilities that have now been identified (Amodio, 2019; Mather, 2022).

Hence, following several decades of research, it is now generally agreed that all of these elements described above—a basis in natural physics, mathematics, geometry, and natural psychology (for a review of these four pillars of animal cognition, see Vallortigara et al., 2010) is present in primates, many other mammals, birds, and even some insects so far tested.

The evaluative processes that animals may undertake, as Lazarus (1982) had argued, however, did not imply anything about deliberate reflection, rationality, or awareness but suggested that responses are based on learning and recall of previous and similar situations (accessible memory). Social learning undoubtedly plays a key role, both in an ecological and a psychological sense (Whiten and de Waal, 2018). Part of that learning process is taking note of someone else and, if a partner, that someone else may even be openly acknowledged by signs of affection (preening, for instance), in responding to requests, or in simple forms, by just walking in step, mirror imaging movements.

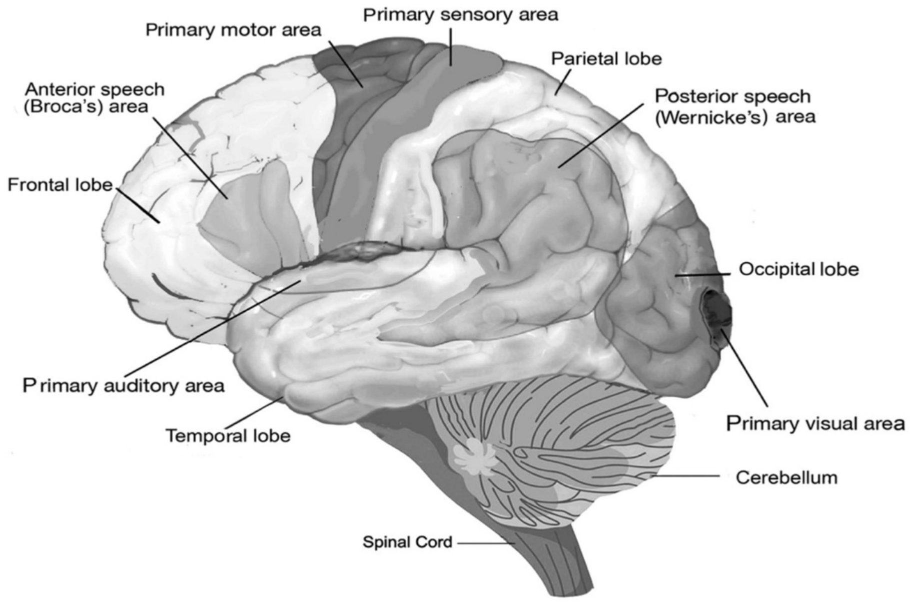


FIGURE 2
Functional areas of the human brain. The diagram shows the areas for speech and the location of basic perceptual areas (audition, vision, and primary sensory area) and motor area, as well as the anterior Broca's area and posterior Wernicke's areas, both of which are indispensable for speech and for which homologs have been found in non-human primates (Hopkins, 2022).

5.2. Multimodal perception, expression, and cognition

Both at functional levels and one that involves regulation of emotions in some way, duetting utilizes one single modality (audition), largely because individuals may be visually separated from one another (Smith et al., 2010). Duets may function as ways to reassure the two partners of their current location, be an example of mate-guarding or warn a potential intruder against invasion (Grafe and Bitz, 2004; Marshall-Ball et al., 2006). However, turn-taking in communication can happen in wider contexts and in visual contact with each other. Animals, be they diurnal or nocturnal, operate in a multi-sensory world (Partan and Marler, 1999; Hiramatsu et al., 2009). In addition to auditory information, individuals may simultaneously be exposed to and respond to visual and olfactory cues that may either confirm and strengthen the information received or contradict or annul information received in another modality (there is food but there is also a predator—a negative sensory input). Such stimuli combined may produce very different outcomes in behavior (Zhou et al., 2010). New World monkeys, such as the common marmoset (*Callithrix jacchus*), have a well-developed olfactory system and display a range of olfaction-based social behavior (Epple, 1993; Lazaro-Perea et al., 1999). As yet, however, there are too few studies that address the effects on the behavior of multi-modal signaling or incidental information on the response choices.

In our laboratory, we tested the idea of whether predator-naive marmosets (*Callithrix j. jacchus*) would show aversion to and withdrawal from fecal odors of predators and curiosity (approach) to food-based odors and found that marmosets perceive and respond to specific olfactory information and that olfaction may be more important for a broad range of functions not previously considered (Kemp and Kaplan, 2012). Although the importance of olfaction gradually declined in the primate line, it is worth remembering that olfaction has played an important role in perception apart from vision and audition (red-bellied tamarins: Caine and Weldon, 1989; cotton-top tamarins: Buchanan-Smith et al., 1993; wild mousse lemurs: Kappel et al., 2011). In the few research projects in which multimodal perception and responses have been investigated in detail, performance and success (whatever the measure might have been) tend to be enhanced by multimodal signaling. Rek and Magrath (2020), for instance, showed that visual display *enhances* vocal duet production in Australian magpie-larks, *Grallina cyanoleuca*.

Facial expressions, as visual stimuli, belong to another form of non-verbal communication that is shared by many primates, all apes, and humans, because we share the same facial musculature with the apes (Burrows, 2008). These expressions have been studied extensively, starting with Richard Andrew's first very detailed account (Andrew, 1963) and followed by an unbroken plethora of research publications until now, be this of great apes, some other primates, or humans (apes: van Lawick-Goodall, 1968; Parr and Waller, 2006; Kret et al., 2020; macaques: Hinde and Rowell, 1962; Partan, 2002; marmosets: Epple, 1967; Stevenson and Poole, 1976). We were interested to see how well marmosets could "read" the facial expressions of their cage mates and devised video footage, played back on large screens behind a food dish, and then tested whether specific facial responses to food and predator-related stimuli might act as social signals to conspecifics (Kemp and Kaplan, 2013). We recorded two contrasting facial expressions (fear and pleasure) as separate sets of video clips and then presented food together with these images of cage mates. Results showed that the expression of a fearful face on the screen significantly reduced time spent near the food bowl compared to the duration of staying near the food bowl when a face showing pleasure was screened.

These multifarious non-verbal forms of communication in addition to gestural signals (Fröhlich and Hobaiter, 2018) remind one that all these aspects of primate and human social life act in unison, in one body and often simultaneously, providing a rich palate of possible emotions, messages, and intentions to be interpreted by the recipient (Kret et al., 2020).

The central cognitive task lies in the ability of the partner, offspring, or wider group to read these signals correctly and in conjunction with one another (Fröhlich and van Schaik, 2018). Waller's objection to viewing these communicative acts together is that they may have different underlying cognitive processes (Waller et al., 2013). Processing simultaneous signals can be far more challenging than one might suspect. The combinatory signals allow for strong messages in the negative and positive sense (Crivelli and Fridlund, 2018) by providing tools for deception (Gyger and Marler, 1988), contradictions, ambiguities, and misunderstandings—a possibility that does not improve with the evolution of speech (Herman et al., 2022). The understanding of non-verbal messages is supported by the brain's mirroring system that is shaped by individual experience. Tight links, therefore, exist between action and perception, both within an individual and between several individuals (Roelfsema et al., 1997; Dinstein et al., 2007; Schippers et al., 2010).

Michael Corballis has been particularly persuasive over the years in his argument that gestural communication was the forerunner of human language evolution (Corballis, 2002, 2010). Many have agreed with him, and they have been supported by further evidence, as already mentioned, showing homologous areas of the human brain for speech production (Broca's area) and for language comprehension (Wernicke's area) are found in great apes and macaque brains (Cantalupo and Hopkins, 2001; Gil-da-Costa et al., 2006). Infants make pointing gestures spontaneously from an early age (Liszkowski et al., 2004), a key to understanding the development of language and theory of mind (Butterworth, 2003; Camaioni et al., 2004). Others have argued that the act of pointing is a complex cultural and cognitive behavior (Kita, 2003).

Undoubtedly, such evidence of referential gesturing adds to the duetting paradigm of coordinated action involving a conspecific. However, it is not enough to explain the substantial expansion of the hominin brain and the actual development of human speech because gesturing itself is already a clear sign of motivation to expand communication. The question is rather, what events, ecological and social circumstances, prompted and motivated the expansion of communication and cooperation in partners and groups.

6. Prosociality and cognition

To be in synchrony with another individual on a specific task may be the beginning of some ongoing collaboration (Duguid and Melis, 2020) and thus create openings for entering into some level of the ongoing bond. When Heyes (2009) summarized her research interest in imitation and mimicry in human development, she might as well have spoken about prosociality in primates, birds, and humans. Heyes said: "Imitation is an important and intriguing neurocognitive process: a process that bridges the gap between one mind and another; that powers cognitive and social development in infancy and childhood; that promotes empathy, cooperation and well-being in our relationships with others" (Heyes, 2009). How these variables might interact is presented in Figure 3, showing that cognition is both shaped by learning or knowledge already gained and memorized and by perception-action systems (Savage et al., 2020). Evolving sociality, synchronicity of movement, body or facial expression, or synchronized vocalizations, such as duetting, increases the chances for further communicative acts, including the development of a gestural repertoire that is shaped and enhanced by cognitive abilities.

Note that, in Figure 3, prosociality is not featured as central to this diagram but it occurs two times: as part of the affective system and as part of the cognitive system. The cognitive system relies on perception and then relays its emotional response via a network of prefrontal nuclei for learning and memory to action a response. Figure 3 also shows that motivation is influenced and reinforced by emotions which in turn are regulated by a set of reward hormones: the dopaminergic reward system, the endogenous opioid system, and oxytocin (Savage et al., 2020). These systems also regulate moods and behavior in humans. However, they can only become active and functional if the individual has developed an ability to identify and respond to social cues, such as gaze and head or body orientation, clearly beneficial for the survival of any social animal, even in fish (Leadner et al., 2021).

Prosocial tendencies represent the next cognitive leap (post simple synchrony) in that a conspecific, or a group of conspecifics, come to constitute valuable "others" and are recognized as having their personalities, needs, moods, and demands. In addition, prosocial tendencies seem to require some basic form of bonding with another individual or a group beyond a mother–infant bond (that, in birds, can be achieved by imprinting, McCabe, 2019). Prosociality is sometimes referred to as "self-other resonance" to emphasize the interactive nature of this trait (Christov-Moore and Iacoboni, 2016) and is as much a social, emotional, and a cognitive process.

In human developmental psychology, prosociality has been a key topic for research into children's and adolescent behavior (Ferraro, 2019), but it is relatively rarely considered in animals. One of the reasons why it is so central in human developmental psychology is defined by the behavioral damage done to adolescent individuals in whom "prosocial" attitudes are missing or are poorly developed (Meehan et al., 2019).

However, de Waal and Suchak (2010) discussed prosociality in non-human primates at some length and emphasized the difference between empathy and prosociality, as is also used in this study. In brief, empathy is the capacity of the observer to

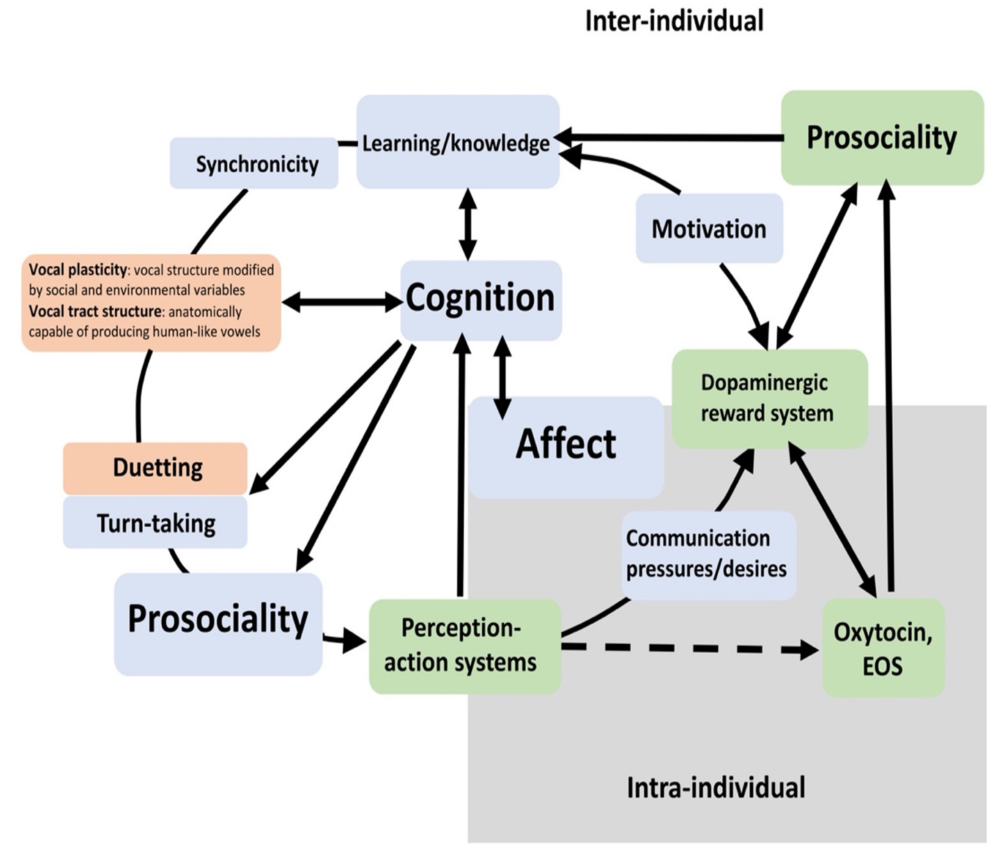


FIGURE 3
Prosociality in context—Adapted from Savage et al. (2020) (in green) and argument included from Snowdon (2018) (in orange). EOS refers to Endogenous Opioid System; see also the article by Hurlemann and Marsh (2019) which provides an overview of the neurobiology of prosocial behavior and the modulatory role of oxytocin in human prosociality.

feel with and indirectly experience the emotional state or even pain of the observed, while prosocial responses can be entirely unselfconscious, unaware, and consist of spontaneous helpful acts that demand no reciprocity. Prosociality and empathy have in common that either may be readily expressed within the social network of the observer or, in rare cases, may also be extended to strangers (Norscia and Palagi, 2011; Decety et al., 2016).

7. Human language evolution

In 2015, a specialist in evolutionary anthropology wrote an article that argued for human uniqueness on the grounds of three inherently human characteristics: an evolved advanced cognition, hyper-prosociality, and psychology for social learning (Marean, 2015). The narrative about human evolution and the development of human language as a set of linear prehistoric events seems to border on story-telling and certainly suggests an over-simplification, based on fossil finds that are possibly chronologically tens or even hundreds of thousands of years apart (de León et al., 2021). Nevertheless, fossil finds so far indicate that there was a modern human lineage in Africa (Templeton, 2002; Carotenuto et al., 2016; Husson et al., 2022), at least one archaic African lineage (Hammer et al., 2011), and two archaic Eurasian lineages, Neanderthals and Denisovans (Mithen, 2006; Petr et al., 2020; de León et al., 2021). Certainly, the hypotheses of human evolution are getting more complex. The more fossil skulls are found and the more improved techniques of dating them in the 21st century, the less clear are the genetic and anatomical elements, involving consideration of admixtures and radiations which have made some evolutionary trajectories more confusing and unresolved (Lieberman, 2001). It is clear, however, that there was a substantial increase in brain volume from chimps and bonobos to *Homo habilis* and to *Homo erectus*, as shown in Table 2.

In between the estimated departure dates of hominins from Africa, there are long periods without any fossil evidence of any kind, in which various human groups would presumably have moved about, probably in small isolated bands. Genetically, socially and cognitively, much could have transpired. All hominin lineages eventually went extinct, leaving one single remaining homo member of the large family tree and perhaps its survival was contingent on precisely the qualities that were described in this study. From very different perspectives, the present paper and Marean's article have arrived at the same conclusion of the centrality of cognition, prosociality, and the ability to work closely together, be these primates or humans.

Assuming the above is correct, the gap in explaining human language evolution is still large and might remain an open question. One can agree with Marean that the surviving branch was an "anomaly" in so far, as it was the only branch surviving despite tough climatic conditions and the species' very poor physical attributes. Humans, compared to other primates, had no fur to protect themselves from insects, from cold or heat, had poor climbing ability, only average speed in running, no claws, and little physical strength against any predators. But they did get one advantage: a large brain equipped for problem-solving and close cooperation, both enough to survive.

Human language is an arbitrary construct, and all bands of humans developed their own. One of the oldest living cultures in the world, Australia's Aboriginal culture, consisted of more than 250 nations and could boast as many languages, most of them bearing no similarity to each other (Blevins, 2001; Dixon and Dixon, 2011), except for the additional many dialects. There is no reason to think that all human communities developed language at the same time or had similar vocabulary sizes or even names for the same concepts or objects (Blevins, 2001; Dixon and Dixon, 2011).

An argument, rarely raised but possibly of substantial importance is to consider life histories in hominin species. Based on available evidence, John L. Locke and Barry Bogin did exactly that: they calculated the mean age of eruption of the first permanent molar and built the length of childhood around such available physical data. According to Locke and Bogin (2006), stages of childhood gradually lengthened from *Homo habilis* (3.8 years), early *Homo erectus* (4.5 years), and late *Homo erectus* (5.0 years) to *Homo sapiens* (6.2 years). Juvenile and adolescent stages also lengthened from 12 years in *Homo habilis* to 17 years in *Homo sapiens* (Locke and Bogin, 2006).

Lengthening childhood and juvenile stages over time suggests an increased biparental or family group commitment to protect and food-support their offspring for an ever-increasing period. We know from primates under group or biparental care, as well as from biparental care in avian species with protracted "childhoods", that the offspring seem to get three main benefits from this delay in maturation: 1, protection (low-stress levels); 2, long learning time; and 3, more play time with other juveniles fostering prosocial development. These social conditions, as I have shown elsewhere (Kaplan, 2020b), tend to correlate with growing large brains. In chimpanzees, offspring are typically weaned at ∼4 years of age, and thereafter the immatures of the western chimpanzees (*Pan troglodytes verus*), a subspecies of the genus *Pan troglodytes,* continue to associate with their mothers for up to 10 years beyond weaning (Samuni et al., 2020). From studies of both wild and captive gibbons, it is thought that gibbons reach sexual maturity at about 6–8 years of age, and the siamang (*Hylobates syndactylus*) at about 8–9 years (Geissmann, 1991). Similarily, in birds, some cockatoos reach sexual maturity when they are 6–8 years of age. To them and other avian species with similar life histories, the benefits tend to be identical to those in long-nurtured primates and hominin societies, such as longevity, cognitive complexity, and strong social bonds (Kaplan, 2019).

Finally, as Arbib (2013) rightly pointed out: "language" is not speech. Arbib (2013) and others before and since have seen song and dance as a bridge between music and language. The latter can exist as speech or in signs and can exploit voice, hands, and face (be this via voice utterances, whistles, drumming, clapping, and gesturing) using hearing and/or vision so that there is always a duality of patterning. To this day, there are sign languages, many whistled languages (Meyer, 2008), and also drum languages (Seifart et al., 2018; Ros, 2021). And there is dance combining rhythm, sound, and even song and movement. Laland et al. (2016) reminded us that dance has representational properties that "rely on the dancers' ability to imitate particular people, animals or events, as well as the audience's ability to recognize these correspondences." The beginnings of language might well have occurred via imitation and mimicry of animals and were expressed in music and dance. Both are ubiquitous among humans (Lewis, 2009; Knight and Lewis, 2017). Mimicry of sound (entrainment to a musical beat) or of body movement (dance) is suggestive of the capabilities of motor and vocal imitation (Fitch, 2016; Laland et al., 2016; Fink et al., 2021). Mimicry of sounds, songs, and dance may first have evolved from imitated movements (say of animals they have seen and might have hunted) to communicate socially relevant information about them accurately. Indeed, such information could have been conveyed in many ways, be this via gestures, pointing, sound imitation, or even dancing. These articulations may well be processed by a similar neural network as those responsible for vocal learning in songbirds (Schuppe et al., 2022). Darwin thought that different aspects of language were acquired *sequentially* and possibly over vast stretches of evolutionary time. Vocal actions needed partners, such as in duetting (Clink and Lau, 2020; Clink et al., 2020) or turn-taking (Takahashi et al., 2016), joint-calling as in choruses (Mitani and Gros-Louis, 1998; De Gregorio et al., 2021, 2022), and referential signaling addressed to a conspecific or a family group and groups (Seyfarth et al., 1980; Snowdon, 2020; Vonk, 2020). And in such partnerships in dyadic or group vocalizations and movements, coalitions and partnerships were forged that could solve problems and innovate.

Tobias et al. (2016) argued that communal signaling (which includes duetting and choruses) is perhaps the most complex and least understood form of communication in social animals. They used Bayesian phylogenetic models to test whether acoustic communal signals are explained by a range of life history and environmental variables across 10,328 bird species worldwide and estimated that duets and choruses occur in some 1830 (18%), and in these, evolutionary transitions between communal signaling and solo signaling were "not explained by latitude, migration, climate, or habitat, and only weakly correlated with cooperative breeding. Instead, they are most strongly associated with year-round territoriality, typically in conjunction with stable social bonds" (Logue and Hall, 2014; Tobias et al., 2016).

I suspect that in some cases, if not all, prosociality was a vital step toward communicating with others on a broader basis, be this out of necessity or to share information that was about matters not immediately visible. Beyond the speculative, the neurobiological and anatomical evidence and the behavior of extant vertebrates, especially primates, have provided mounting evidence of the importance of the development of prosociality which makes its centrality in human language evolution very plausible.

8. Concluding remarks

In his treatise *The Expression of the Emotions in Man and Animals* (Darwin, 1872) and in chapters 2 and 3 of *The Descent of Man* (Darwin, 1871), Darwin talked about attention and imitation and he argued that if an individual can attend to something then it is possible for that individual either to imitate what it has seen or to be taught to do something (Kaplan and Rogers, 2004). Duetting and synchronized movements are both hallmarks of communication and group affiliations known in the primate line and particularly evident in many songbird species.

Furthermore, there is an ancestral social behavior network within the basal forebrain and midbrain that is common to all vertebrates from teleosts to birds and mammals and a mesolimbic reward system that forms a larger social decision-making network (Goodson, 2005; O'Connell and Hofmann, 2011). At the very least, one can say that a path to express and develop the ability for adaptive social behavior toward conspecifics has been in existence in ancient and well-preserved networks of the brain. Many research projects have also shown that interpersonal synchrony increases affiliation and increases cooperative behavior (Hove and Risen, 2009; Reddish et al., 2013). Note, however, that the social and vocal aspects of behavior can be mutually reinforcing. In a study of vocal behavior in bonobos, the researchers concluded that social bonds drive vocal exchanges (Levréro et al., 2019).

To have identified some potential sources for precursors of the evolution of human language should not be seen at the exclusion of many other evolutionary elements that might well have played into such a momentous innovation as speech. One might well speculate that any form of "language" in humans was evolutionarily a late development, suggested by complex activation of brain areas when such communicative acts occur. Kaan and Swaab (2002) found neuroimaging support for arguing that syntactical processing of multimodal information does not just recruit one specific brain area. Instead, a network of areas including Broca's area and anterior, middle, and superior areas of the temporal lobes are involved. Okobi et al. (2019) identified the neural control needed for duetting. Although this applied to Alston's singing mice, the model has been proposed as an emerging vocalization model also for duets in primates (Neff, 2019). Indeed, in primates, duetting happens to be one of the most convincing examples of vocal flexibility. How else would bonded couples achieve their voiced synchrony if it were not for the ability to adjust any specific features in vocal production, be they syntactical, rhythmic, or in frequency.

Anatomically, the road from pre-speech to speech in the hominid line was not blocked by the inability for vocal learning in primates or for lack of ability to form concepts, think of things past, and even plan a future. Primates and specifically great apes and some New World monkeys have shown remarkable cognitive abilities in solving problems and vocal learning.

Whether the gestural thesis of the origin of human language might explain the evolution of speech is not the point of argument here. The language might as well have developed via music and dance as said above. Moreover, "language" did not always result in speech as has also been pointed out above. These evolving systems of complex communication all reflect forms of self-expression as well as stable, communally agreed, unambiguous vocal labels for objects or concepts. While they well describe how rich in communicative abilities they may be, none of them show why any of them would have evolved in the first place.

The argument here has focussed on the possible motivators for the evolutionary precursors of such manifestations of communication.

First, in evolution, change tends to happen when an organism is stressed to fulfill its basic needs and/or when a small change in behavior or physiology gives one species a significant advantage over another. The hominid line had a poor record in meeting the challenges. All hominid ancestors eventually went extinct except *Homo sapiens,* suggesting that substantial innovations were needed to make this last hominin species viable. Studies on stress responses in modern humans interestingly found that stress triggers social approach behavior, which operates as a potent stress-buffering strategy, thereby providing evidence for the context and triggers in prosocial behavior, also referred to as the tend-and-befriend hypothesis (Von Dawans et al., 2012). How much speech has to do with it is yet another question.

Second, one constraint in the formulations of theories on human language evolution has been the need to remain focussed on one major variable, such as gestural origins, vocal synchronisations, and concept of musicality or dance. Hence, theories have tended to be single-focused on one singular candidate as a precursor of human language evolution. However, focus on any of these visual or vocal social expressions (and their expansions) is a focus on vocal expressions that are all, to varying degrees, *outcom*es in the communicative refinement of expressions of vocal and movement behavior.

Instead, this paper has posed the question of what impetus could have led to any of these impressive self-expressions and communicative complexities. It has been the contention of this study to ask why such outcomes occurred at all and which evolutionary steps had to precede these developments. As Hoehl et al. (2020) argued: Synchronizing benefits arise from an increased predictability of incoming signals and include many positive outcomes ranging from basic information processing at the individual level to the bonding of dyads and larger groups. Cooperative behavior, starting with specific vocal expressions such as duetting, fostered social cohesion (Launay et al., 2016). To achieve some synchronicity in duetting, as has been shown in many studies mentioned here, requires vocal flexibility. A recent study of lar gibbons, *Hylobates lar* (Raimondi et al., 2023), revealed not only substantial sophistication in the gibbon's rhythmic vocal expressions but showed that isochrony, at the core of human musicality, is present in lar gibbon duetting. Raimondi et al. (2023) found that gibbons are more isochronous when duetting than singing solo, achieving a higher-than-chance degree of synchrony in their duets because of this ability to rhythmically adjust their part of the duet and coordinate it (Raimondi et al., 2023).

In conclusion, the evolution of prosocial behavior may well be the vital precondition for, and the motivational link to, any expansion of cognition and communication and ultimately causally related to the evolution of human language. Furthermore, evidence that has been provided in the duetting literature of primates and dolphins is the degree of flexibility in vocal exchanges. The remarkable vocal communication among dolphins has no bearing on human language evolution but is a case of convergent evolution. Their social behavior also showed consistency in some other social factors, comparable with primates (King et al., 2022). Indeed, in dolphins and some avian species, the same or very similar basic biological and social factors can be observed: high cognitive ability, strong social bonds, and a high degree of vocal flexibility and individuality as the vocal labeling of dolphins (King et al., 2018). Clink et al. (2022) discovered flexibility in vocal exchanges of Gursky's spectral tarsier, *Tarsius spectrum gurskyae*. They rightly argued that vocal flexibility (and individuality) is a precursor to human language, and it evolved early in the primate lineage and long before the emergence of modern humans (Clink et al., 2022).

It seems from the physical evidence on record that joint actions led to more cooperation, more communication, further brain growth, better problem-solving, and a more secure place for humans in the natural environment, despite the many physical inadequacies of the modern human species. The motivation to pursue shared goals and indulge in creative models of ever-expanding communication, eventually language, also has to do with the extensive reward system the brain provided. This probably came about because positive rewards accompanied acts and attitudes of prosociality and this, in turn, helped increase affiliative bonds.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This project was funded by the University of New England, Armidale, NSW, Australia.

Acknowledgments

I thank Dr. Patrice Adret for excellent editorial support, am very grateful to my university for funding this project, and am thankful for the very helpful comments on this paper by Prof. Lesley Rogers.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abrahamczyk, S., Souto-Vilarós, D., and Renner, S. S. (2014). Escape from extreme specialization: passionflowers, bats and the sword-billed hummingbird. *Proc. R. Soc. B: Biol. Sci.* 281, 20140888. doi: 10.1098/rspb.2014.0888

Adret, P. (2022). Developmental plasticity in primate coordinated song: parallels and divergences with duetting songbirds. *Front. Ecol. Evol.* 10, 862196. doi: 10.3389/fevo.2022.862196

Adret, P., Dingess, K. A., Caselli, C. B., Vermeer, J., Martínez, J., and Luna Amancio, J. C. (2018). Duetting patterns of titi monkeys (Primates, Pitheciidae: Callicebinae) and relationships with phylogeny. *Animals* 8, 178–211. doi: 10.3390/ani8100178

Aldrich, B. C., Feddema, K., Fourage, A., Nekaris, K. A.I., and Shanee, S. (2023). "Primate portrayals: narratives and perceptions of primates in entertainment," in *Primates in Anthropogenic Landscapes*,eds T. McKinney, S. Waters, and M. A. Rodrigues (Cham: Springer), 307–326.

Amici, F.,Visalberghi, E., and Call, J. (2014). Lack of prosociality in great apes, capuchin monkeys and spider monkeys: convergent evidence from two different food distribution tasks. *Proc. R. Soc. B:Biol. Sci.* 281, 1699. doi: 10.1098/rspb.2014.1699

Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884

Amodio, P. (2019). Octopus intelligence: the importance of being agnostic. *Anim. Senti.* 4, 20. doi: 10.51291/2377-7478.1507

Amorim, P. S., Diniz, P., Rossi, M. F., and Guaraldo, A. C. (2022). Out of sight, out of mind: dear enemy effect in the rufous hornero, *Furnarius rufus*. *Anim. Behav.* 187, 167–176. doi: 10.1016/j.anbehav.2022.03.010

Andrew, R. J. (1963). The origin and evolution of the calls and facial expressions of the primates. *Behaviour* 20, 1–107. doi: 10.1163/156853963X00220

Andrew-DeWoody, J., Fletcher, D. E., David Wilkins, S., Nelson, W. S., and Avise, J. C. (2000). Genetic monogamy and biparental care in an externally fertilizing fish, the large mouth bass (*Micropterus salmoides*). *Proc. R. Soc. Series B Biol. Sci.* 267, 2431–2437. doi: 10.1098/rspb.2000.1302

Andrieu, J., Penny, S. G., Bouchet, H., Malaivijitnond, S., Reichard, U. H., et al. (2020). White-handed gibbons discriminate context-specific song compositions. *PeerJ* 8, e9477. doi: 10.7717/peerj.9477

Arbib, M. A. (2013). *Language, Music, and the Brain. A Mysterious Relationship*. Cambridge, MA; London: MIT Press.

Ardesch, D. J., Scholtens, L. H., Li, L., Preuss, T. M., Rilling, J. K., and van den Heuvel, M. P. (2019). Evolutionary expansion of connectivity between multimodal association areas in the human brain compared with chimpanzees. *PNAS* 116, 7101–7106. doi: 10.1073/pnas.1818512116

Arriaga, G., and Jarvis, E. D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? *Brain Lang.* 124, 96–116. doi: 10.1016/j.bandl.2012.10.002

Atzil, S., Hendler, T., Zagoory-Sharon, O., Winetraub, Y., and Feldman, R. (2012). Synchrony and specificity in the maternal and the paternal brain: relations to oxytocin and vasopressin. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 798–811. doi: 10.1016/j.jaac.2012.06.008

Aureli, F., and Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. *Behav. Ecol. Sociobiol.* 73, 6. doi: 10.1007/s00265-018-2604-5

Avarguès-Weber, A., Dyer, A. G., and Giurfa, M. (2011). Conceptualization of above and below relationships by an insect. *Proc. R. Soc. B* 278, 898–905. doi: 10.1098/rspb.2010.1891

Avarguès-Weber, A., and Giurfa, M. (2013). Conceptual learning by miniature brains. *Proc. R. Soc. B* 280, 1907. doi: 10.1098/rspb.2013.1907

Bailey, W. J. (2003). Insect duets: underlying mechanisms and their evolution. *Physiol. Entomol.* 28, 157–174. doi: 10.1046/j.1365-3032.2003.00337.x

Baker-Médard, M. S., Baker, M. C., and Logue, D. M. (2013). *Chorus song of the Indri (Indri indri: Primates, Lemuridae): group differences and analysis of within-group vocal interactions*. e-scholarship Publishing, University of California. Available online at: https://escholarship.org/uc/item/0gg070fd

Bales, K. L., Ardekani, C. S., Baxter, A., Karaskiewicz, C. L., Kuske, J. X., et al. (2021). What is a pair bond? *Horm. Behav.* 136, 105062. doi: 10.1016/j.yhbeh.2021.105062

Banerjee, A,., and Vallentin, D. (2022). Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Curr. Opin. Neurobiol.* 73, 102529. doi: 10.1016/j.conb.2022.102529

Banovac, I., Sedmak, D., Judaš, M., and Petanjek, Z. (2021). Von Economo neurons–primate-specific or commonplace in the mammalian brain?. *Front. Neur. Circ.* 89, 714611. doi: 10.3389/fncir.2021.714611

Barón Birchenall, L. (2016). Animal communication and human language: an overview. *Intern. J Comp. Psych.* 29, 1–26. doi: 10.46867/ijcp.2016.29.00.07

Barrett, L., and Henzi, P. (2005). The social nature of primate cognition. *Proc. R. Soc. B* 272, 1865–1875. doi: 10.1098/rspb.2005.3200

Barton, R. A. (1998). Visual specialization and brain evolution. *Proc. R. Soc. B* 265, 1933–1937. doi: 10.1098/rspb.1998.0523

Bateson, P. P. G., and Martin, P. (2013). *Play, Playfulness, Creativity and Innovation*. Cambridge: Cambridge University Press.

Bernieri, F., and Rosenthal, R. (1991). "Interpersonal coordination: behavior matching and interactional synchrony," in *Fundamentals of Nonverbal Behavior*, eds R. S. Feldman, and B. Rime (Cambridge: Cambridge University Press,) 401–431.

Blevins, J. (2001). *Nhanda: An Aboriginal Language of Western Australia*. University of Hawai'i Press.

Boysen, S. T., and Bernston, G. G. (1989). Numerical competence in a chimpanzee (Pan troglodytes). *J. Comp. Psychol.* 10323, 23. doi: 10.1037/0735-7036.103.1.23

Buchanan-Smith, H. M., Anderson, D. A., and Ryan, C. W. (1993). Responses of cotton-top tamarins (*Saguinus oedipus*) to faecal scents of predators and non-predators. *Anim. Welf.* 2, 17–32. doi: 10.1017/S0962728600015438

Burkart, J. M., and van Schaik, C. P. (2020). Marmoset prosociality is intentional. *Anim.Cogn.* 23, 581–594. doi: 10.1007/s10071-020-01363-6

Burrows, A. M. (2008). The facial expression musculature in primates and its evolutionary significance. *Bioessays* 30, 212–225. doi: 10.1002/bies.20719

Butterworth, G. (2003). "Pointing is the royal road to language for babies," in *Pointing: Where Language, Culture, and Cognition Meet,* ed S. Kita (Mahwah, NJ: Erlbaum), 9–33.

Cadková, L. (2015). Do they speak language. *Biosemiot* 8, 9–27. doi: 10.1007/s12304-014-9225-9

Caesar, C., and Zuberbuehler, K. (2012). Referential alarm calling behaviour in New World primates. *Curr. Zool.* 58, 680–697. doi: 10.1093/czoolo/58.5.680

Caine, N. G., and Weldon, P. J. (1989). Responses by red-bellied tamarins (*Saguinus labiatus*) to fecal scents of predatory and non-predatory neotropical mammals. *Biotropica* 21, 186–189. doi: 10.2307/2388709

Caley, T., Extier, T., Collins, J. A., Schefuß, E., Dupont, L., Malaizé, B., et al. (2018). A two-million-year-long hydroclimatic context for hominin evolution in southeastern Africa. *Nature* 560, 76–79. doi: 10.1038/s41586-018-0309-6

Camaioni, L., Perucchini, P., Bellagamba, F., and Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy* 5, 291–308, doi: 10.1207/s15327078in0503_3

Cantalupo, C., and Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature* 414, 505. doi: 10.1038/35107134

Card, N. A., Stucky, B. D., Sawalani, G. M., and Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: a meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Dev.* 79, 1185–1229. doi: 10.1111/j.1467-8624.2008.01184.x

Carotenuto, F., Tsikaridze, N., Rook, L., Lordkipanidze, D., Longo, L., Condemi, S., et al. (2016). Venturing out safely: the biogeography of Homo erectus dispersal out of Africa. *J. Hum. Evol.* 95, 1–12. doi: 10.1016/j.jhevol.2016.02.005

Carson, D. C. (2013). Perceptions of prosocial and delinquent peer behavior and the effect on delinquent attitudes: a longitudinal study. *J. Crim. Just.* 41, 151–161. doi: 10.1016/j.jcrimjus.2013.01.005

Christov-Moore, L., and Iacoboni, M. (2016). Self-other resonance, its control and prosocial inclinations: brain-behavior relationships. *Hum. Brain Mapp.* 37, 1544–1558. doi: 10.1002/hbm.23119

Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., and Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Curr. Anthropol.* 55, 419–443. doi: 10.1086/677209

Clarke, E., Reichard, U. H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE* 1, e73. doi: 10.1371/journal.pone.0000073

Clink, D. J., Comella, I. A., Tasirin, J. S., and Klinck, H. (2022). Tarsier islands: exploring patterns of variation in tarsier duets from offshore islands of North Sulawesi. *Am. J. Primat.* e23410. doi: 10.1002/ajp.23410

Clink, D. J., and Lau, A. R. (2020). Adherence to Menzerath's Law is the exception (not the rule) in three duetting primate species. *R. Soc. Open Sci.* 7, 201557. doi: 10.1098/rsos.201557

Clink, D. J., Tasirin, J. S., and Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Curr. Zool.* 66, 173–186. doi: 10.1093/cz/zoz035

Clink, D. J., Zafar, M., Ahmad, A. H., and Lau, A. R. (2021). Limited evidence for individual signatures or site-level patterns of variation in male northern gray gibbons (*Hylobates funereus*) Duet Codas. *Intern. J. Primatol.* 42, 896–914. doi: 10.1007/s10764-021-00250-2

Clutton-Brock, T. (1991). *The Evolution of Parental Care*. Princeton, NJ: Princeton University Press.

Cockburn, A. (2006). Prevalence of different modes of parental care in birds. *Proc. R. Soc. B* 273, 1375–1383. doi: 10.1098/rspb.2005.3458

Coleman, M. J., Day, N. F., Rivera-Parra, P., and Fortune, E. S. (2021). Neurophysiological coordination of duet singing. *PNAS* 118, e2018188118. doi: 10.1073/pnas.2018188118

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Corballis, M. C. (2010). Language as gesture. *Hum. Mov. Sci.* 28, 556–565. doi: 10.1016/j.humov.2009.07.003

Coss, R. G., Cavanaugh, C., and Brennan, W. (2019). Development of snake-directed antipredator behavior by wild white-faced capuchin monkeys: III. The signaling properties of alarm-call tonality. *Am. J. Primatol.* 81, e22950. doi: 10.1002/ajp.22950

Courts, R., Erbe, C., Wellard, R., Boisseau, O., Jenner, K. C., and Jenner, M. N. (2020). Australian long-finned pilot whales (Globicephala melas) emit stereotypical, variable, biphonic, multi-component, and sequenced vocalisations, similar to those recorded in the northern hemisphere. *Sci. Rep.* 10, 1–14. doi: 10.1038/s41598-020-74111-y

Crivelli, C., and Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends Cogn. Sci.* 22, 388–399. doi: 10.1016/j.tics.2018.02.006

Cronin, K. A. (2012). Prosocial behaviour in animals: the influence of social relationships, communication and rewards. *Anim. Behav.* 84, 1085–1093. doi: 10.1016/j.anbehav.2012.08.009

Dahlin, C. R., and Benedict, L. (2014). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 120, 1–10. doi: 10.1111/eth.12182

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals.* London: John Murray.

Das, A., and Gilbert, C. D. (1995). Receptive field expansion in adult visual cortex is linked to dynamic changes in strength of cortical connections. *J. Neurophysiol.* 74, 779–792. doi: 10.1152/jn.1995.74.2.779

De Casien, A. R., and Higham, J. P. (2019). Primate mosaic brain evolution reflects selection on sensory and cognitive specialization. *Nat. Ecol. Evol.* 3, 1483–1493. doi: 10.1038/s41559-019-0969-0

De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., et al. (2021). Born to sing! Song development in a singing primate. *Curr. Zool.* 67, 585–596. doi: 10.1093/cz/zoab018

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evo.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R. M., et al. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Curr. Zool.* 65, 89–97. doi: 10.1093/cz/zoy058

de León, M. S. P., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., Warren, J. L. A., et al. (2021). The primitive brain of early Homo. *Science* 372, 165–171. doi: 10.1126/science.aaz0032

de Waal, F. B. M., and Suchak, M. (2010). Prosocial primates: Selfish and unselfish motivations. *Philos. Trans. R. Soc. B* 365, 2711–2722. doi: 10.1098/rstb.2010.0119

Deacon, T. W. (2003). "Multilevel selection in a complex adaptive system: the problem of language origins," in *Evolution and Learning: The Baldwin Effect Reconsidered*, eds B. H. Weber, and D. J. Depew (Cambridge, MA: MIT Press).

Decety, J., Bartal, I. B.-A., Uzefovsky, F., and Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour: highly conserved neurobehavioural mechanisms across species. *Philos. Trans. R. Soc. B* 371, 20150077. doi: 10.1098/rstb.2015.0077

Dinstein, I., Hasson, U., Rubin, N., and Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 198, 1415–1427. doi: 10.1152/jn.00238.2007

Dixon, R. M., and Dixon, R. M. (2011). *The Languages of Australia*. Melbourne, VIC: Cambridge University Press.

Dixson, A. F. (2009). *Sexual Selection and the Origins of Human Mating Systems*. Oxford: Oxford University Press.

Dolotovskaya, S., Walker, S., and Heymann, E. W. (2020). What makes a pair bond in a Neotropical primate: female and male contributions. *R. Soc. Open Sci.* 7, 191489. doi: 10.1098/rsos.191489

Donald, J. N., Bradshaw, E. L., Conigrave, J. H., Parker, P. D., Byatt, L. L., and Noetel, M. (2021). Paths to the light and dark sides of human nature: a meta-analytic review of the prosocial benefits of autonomy and the antisocial costs of control. *Psychol. Bull.* 147, 921. doi: 10.1037/bul0000338

Dowling, J., and Webster, M. S. (2018). Acoustic and physical mate guarding have different effects on intruder behaviour in a duetting songbird. *Anim. Behav.*135, 69–75. doi: 10.1016/j.anbehav.2017.11.011

Duguid, S., and Melis, A. P. (2020). How animals collaborate: underlying proximate mechanisms. *Wiley Interdisc. Rev. Cogn. Sci.* 11, e1529. doi: 10.1002/wcs.1529

Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., et al. (2021). The rise of affectivism. *Nat. Hum. Behav.* 5, 816–820. doi: 10.1038/s41562-021-01130-8

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evol. Anthrop.* 6, 178–190. doi: 10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M. (2014). *Human Evolution.* London: Penguin Books Ltd.

Epple, G. (1967). Vergleichende Untersuchungen über Sexual-und Sozialverhalten der Krallenaffen (Hapalidae). *Folia Primatol.* 7, 37–65. doi: 10.1159/000155095

Epple, G. (1993). "Making sense out of scents: species differences in scent glands, scent-marking behaviour, and scent-marking composition in the Callitrichidae," in *Marmosets and Tamarins: Systematics, Behaviour and Ecology*, ed A. B. Rylands (Oxford: Oxford University Press), 123–151.

Epstein, A. N. (1982). "Instinct and motivation as explanations for complex behavior," in *The Physiological Mechanisms of Motivation*, ed D. W. Pfaff (New York, NY: Springer), 25–58.

Etzel, R., Cornish, M., Kifer, M. S., Nuñez, L., Valladao, G., and Folt, B. (2020). Subterranean advertisement and duet calling behavior in Ptychohyla legleri (Legler's Stream Frog). *Alytes* 37, 57–61. doi: 10.6084/m9.figshare.11797191.v1

Fedorova, N., Evans, C. L., and Byrne, R. W. (2017). Living in stable social groups is associated with reduced brain size in woodpeckers (Picidae). *Biol. Lett.* 13, 20170008. doi: 10.1098/rsbl.2017.0008

Fedurek, P., and Slocombe, K. E. (2011). Primate vocal communication: a useful tool for understanding human speech and language evolution? *Human Biol.* 83, 153–173. doi: 10.3378/027.083.0202

Feistner, A. T., and McGrew, W. C. (1989). Food-sharing in primates: a critical review. *Persp. in Primate Biol.* 3, 21–36.

Ferraro, F. R. (2019). Special issue on psychology of prosocial behavior. *Curr. Psych.* 38, 909. doi: 10.1007/s12144-019-00336-4

Ferreira, L. S., Sábato, V., Pinheiro, T. A., Neto, E., Rocha, L. H., Baumgarten, J., et al. (2022). Long-distance counter calling in maned wolves: friends or foes? *Animal* 12, 1081. doi: 10.3390/ani12091081

Fink, B., Bläsing, B., Ravignani, A., and Shackelford, T. K. (2021). Evolution and functions of human dance. *Evol. Human Behav.* 42, 351–360. doi: 10.1016/j.evolhumbehav.2021.01.003

Finlayson, C. (2014). *The Improbable Primate: How Water Shaped Human Evolution*. Oxford: OUP.

Fitch, W. T. (2013). "Musical protolanguage: Darwin's theory of language evolution revisited," in *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, eds R. C. Berwick and N. Chomsky (Cambridge, MA: MIT Press), 489–504.

Fitch, W. T. (2016). Dance, music, meter and groove: a forgotten partnership. *Front. Hum. Neurosci.* 10, 64. doi: 10.3389/fnhum.2016.00064

Fitch, W. T. (2020). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Philos. Trans. R. Soc. B* 375, 20190046. doi: 10.1098/rstb.2019.0046

Forcina, G., Vallet, D., Le Gouar, P. J., Bernardo-Madrid, R., Illera, G., et al. (2019). From groups to communities in western lowland gorillas. *Proc. R. Soc. B* 286, 20182019. doi: 10.1098/rspb.2018.2019

Forester, D. C., and Harrison, W. K. (1987). The significance of antiphonal vocalisation by the spring peeper, *Pseudacris crucifer* (Amphibia, Anura). *Behaviour* 103, 1–15. doi: 10.1163/156853987X00233

Fröhlich, M., and Hobaiter, C. (2018). The development of gestural communication in great apes. *Behav. Ecol. Sociobiol.* 72, 1–14. doi: 10.1007/s00265-018-2619-y

Fröhlich, M., and van Schaik, C. P. (2018). The function of primate multimodal communi-cation. *Anim. Cogn.* 21, 619–629. doi: 10.1007/s10071-018-1197-8

Fuentes, A. (2000). Hylobatid Communities: Changing views on pair bonding and social organization in hominoids. *Yrbk Phys. Anthrop.* 43, 33–60. doi: 10.1002/1096-8644(2000)43:31+<33::AID-AJPA3>3.0.CO;2-D

Gannon, P. J., Holloway, R. L., Broadfield, D. C., and Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science* 279, 220–222. doi: 10.1126/science.279.5348.220

Gardner, A. R.A., Gardner, B. T., and van Cantfort, T. E. (1989). *Teaching Sign Language to Chimpanzees.* Albany, NY: State University of New York Press).

Gardner, R. A., and Gardner, B. T. (1969). Teaching sign language to a chimpanzee: a standardized system of gestures provides a means of two-way communication with a chimpanzee. *Science* 165, 664–672. doi: 10.1126/science.165.3894.664

Gasser, B., and Arbib, M. (2019). A dyadic brain model of ape gestural learning, production and representation. *Anim. Cogn.* 22, 519–534. doi: 10.1007/s10071-018-1228-5

Geissmann, T. (1991). Reassessment of age of sexual maturity in gibbons (hylobates spp.). *Am. J. Primatol.* 23, 11–22. doi: 10.1002/ajp.1350230103

Geissmann, T. (2000). "Gibbon songs and human music from an evolutionary perspective" in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press). Ch.7.

Gil-da-Costa, R., Martin, A., Lopes, M. A., Munoz, M, Fritz, J. B., and Braun, A. R. (2006). Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque. *Nat. Neurosci.* 9, 1064–1070. doi: 10.1038/nn1741

Goodson, J. L. (2005). The vertebrate social behaviour network: Evolutionary themes and variations. *Horm. Behav.* 48, 11–22. doi: 10.1016/j.yhbeh.2005.02.003

Grafe, T. U., and Bitz, J. H. (2004). Functions of duetting in the tropical boubou, *Laniarius aethiopicus*: territorial defence and mutual mate guarding. *Anim. Behav.* 68, 193–201. doi: 10.1016/j.anbehav.2003.11.003

Grant, P. R. (2017). *Ecology and Evolution of Darwin's Finches (Princeton Science Library Edition)*. Princeton, NJ: Princeton University Press.

Greenfield, P. M., and Savage-Rumbaugh, E. S. (1990). "Grammatical combination in Pan paniscus: processes of learning and invention in the evolution and development of language," in *"Language" and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives*, ed S. T.Parker (Cambridge: Cambridge UP), 540–578.

Güroglu, B., van den Bos, W., and Crone, E. A. (2014). Sharing and giving across adolescence: an experimental study examining the development of prosocial behaviour. *Front. Psych.* 5, 291. doi: 10.3389/fpsyg.2014.00291

Gyger, M., and Marler, P. (1988). Food calling in the domestic fowl, Gallus gallus: the role of external referents and deception. *Anim. Behav.* 36, 358–365. doi: 10.1016/S0003-3472(88)80006-X

Haimoff, E. H. (1986). Convergence in the duetting of monogamous old world primates. *J. Human Evol.* 15, 51–59. doi: 10.1016/S0047-2484(86)80065-3

Hall, M. L., and Magrath, R. D. (2007). Temporal coordination signals coalition quality. *Curr. Biol.* 17, R406–R407. doi: 10.1016/j.cub.2007.04.022

Hammer, M. F., Woerner, AE, Mendez, F. L., Watkins, J. C., and Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *PNAS* 108, 15123–15128 doi: 10.1073/pnas.1109300108

Haraway, M. M., and Maples, E. G. (1998). Flexibility in the species-typical songs of gibbons. *Primates* 39, 1–12. doi: 10.1007/BF02557739

Hare, B. (2017). Survival of the Friendliest: homo sapiens evolved via selection for prosociality. *Annu. Rev. Psych.* 68, 155–186. doi: 10.1146/annurev-psych-010416-044201

Hare, B., and Tomasello, M. (2005). Human-like social skills in dogs? *Trends Cogn. Sci.* 9, 439–444. doi: 10.1016/j.tics.2005.07.003

Hare, B., Wobber, V., and Wrangham, R. (2012). The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Anim. Behav.* 83, 573–585. doi: 10.1016/j.anbehav.2011.12.007

Harting, J. K., Glendenning, K. K., Diamond, I. T., and Hall, W. C. (1973). Evolution of the primate visual system: anterograde degeneration studies of the tecto-pulvinar system. *Am. J. Phys. Anthrop.* 38, 383–392. doi: 10.1002/ajpa.1330380237

Henry, C. S., Brooks, S. J., Duelli, P., Johnson, J. B., Wells, M. M., and Mochizuki, A. (2013). Obligatory duetting behaviour in the Chrysoperla carnea-group of cryptic species (Neuroptera: Chrysopidae): its role in shaping evolutionary history. *Biol. Rev.* 88, 787–808. doi: 10.1111/brv.12027

Herman, H., Purba, R., Sijabat, P. A., Saputra, N., and Van Thao, N. (2022). Investigating the realization of speech function in a speech through systemic functional linguistics perspective. *Script J. Ling. Engl. Teach.* 7, 31–41. doi: 10.24903/sj.v7i1.917

Heyes, C. (2009). Evolution, development and intentional control of imitation. *Philos. Trans. R. Soc. B* 364, 2293–2298. doi: 10.1098/rstb.2009.0049

Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., and Van Essen, D. (2010). Similar patterns of cortical expansion during human development and evolution. *PNAS* 107, 13135–13140. doi: 10.1073/pnas.1001229107

Hinde, R. A., and Rowell, T. E. (1962). Communication by postures and facial expressions in the rhesus monkey (*Macaca mulatta*). *Proc. Zool. Soc. London* 138, 1–21. doi: 10.1111/j.1469-7998.1962.tb05684.x

Hiramatsu, C., Melin, A. D., Aureli, F., Schaffner, C. M., Vorobyev, M., and Kawamura, S. (2009). Interplay of olfaction and vision in fruit foraging of spider monkeys. *Anim. Behav.* 77, 1421–1426. doi: 10.1016/j.anbehav.2009.02.012

Hobaiter, C., Graham, K. E., and Byrne, R. W. (2022). Are ape gestures like words? Outstanding issues in detecting similarities and differences between human language and ape gesture. *Philos. Transact. R. Soc. B* 377, 20210301. doi: 10.1098/rstb.2021.0301

Hockett, C. F. (1959). "Animal ≪languages≫ and human language," in *The Evolution of Man's Capacity for Culture*, ed J. N. Spuhler (Detroit: Wayne State UP), 32–39.

Hoehl, S., Fairhurst, M., and Schirmer, A. (2020). Interactional synchrony: signals, mechanisms, and benefits. *Soc.Cogn. Affect. Neurosci.* 16, 5–18. doi: 10.1093/scan/nsaa024

Hoffmann, S., Trost, L., Voigt, C., Leitner, S., Lemazina, A., Sagunsky, H. (2019). Duets recorded in the wild reveal that inter-individually coordinated motor control enables cooperative behavior. *Nat. Commun.* 10, 2577. doi: 10.1038/s41467-019-10593-3

Hopkins, W. D. (2022). Neuroanatomical asymmetries in nonhuman primates in the homologs to Broca's and Wernicke's areas: a mini-review. *Emerg. Top. Life Sci.* 6, 271–284. doi: 10.1042/ETLS20210279

Hopkins, W. D., Marino, L., Rilling, J. K., and MacGregor, L. A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). *NeuroRep* 9, 2913–2918. doi: 10.1097/00001756-199808240-00043

Hove, M. J., and Risen, J. L. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi: 10.1521/soco.2009.27.6.949

Hurlemann, R., and Marsh, N. (2019). Unraveling the role of oxytocin in the motivational structure of conflict. *Behav. Brain Sci.* 42, e126. doi: 10.1017/S0140525X19000785

Husson, L., Salles, T., Lebatard, A. E., Zerathe, S., Braucher, R., Noerwidi, S., et al. (2022). Javanese Homo erectus on the move in SE Asia circa 1.8 Ma. *Sci. Rep.* 12, 1–12. doi: 10.1038/s41598-022-23206-9

Hyder, F., Rothman, D. L., and Bennett, M. R. (2013). Cortical energy demands of signaling and nonsignaling components in brain are conserved across mammalian species and activity levels. *PNAS* 110, 3549–3554. doi: 10.1073/pnas.1214912110

Isler, K., and van Schaik, C. P. (2009). The expensive brain: a framework for explaining evolutionary changes in brain size. *J Hum. Evol.* 57, 392–400. doi: 10.1016/j.jhevol.2009.04.009

Issa, M. M., Sikora, B., Rusiecki, S., and Osiejuk, T. S. (2023). The Yellow-breasted Barbet (*Trachyphonus margaritatus*) introduces vocal duets and choruses with a specific multimodal signal, during territorial advertisement. *J. Ornithol.* 164, 183–192. doi: 10.1007/s10336-022-02016-w

Jaeggi, A. V., and Gurven, M. (2013). Natural cooperators: food sharing in humans and other primates. *Evol. Anthrop. Issues News Rev.* 22, 186–195. doi: 10.1002/evan.21364

Janik, V. M., and Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Phil. Transact. Roy. Soc. B.* 376, 20200244. doi: 10.1098/rstb.2020.0244

Kaan, E., and Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends Cogn. Sci.* 6, 350–356. doi: 10.1016/S1364-6613(02)01947-2

Kaplan, G. (2014). Animal communication. *WIREs Cogn Sci.* 5, 661–677. doi: 10.1002/wcs.1321

Kaplan, G. (2015). *Bird Minds. Cognition and Behaviour of Australian Native Birds.* Melbourne, VIC: CSIRO Publishing.

Kaplan, G. (2019). *Bird Bonds. Sex, Mate-choice and Cognition in Australian Native Birds.* Sydney, NSW: Pan Macmillan.

Kaplan, G. (2020a). Long-term attachments and complex cognition in birds and humans are linked to pre-reproductive prosociality and cooperation. Constructing a hypothesis. *Ann. Cogn. Sci.* 4,127–142. doi: 10.36959/447/347

Kaplan, G. (2020b). Play behaviour, not tool using, relates to brain mass in a sample of birds. *Sci. Rep.* 0, 20437. doi: 10.1038/s41598-020-76572-7

Kaplan, G. (2023). The evolution of social play in songbirds, parrots and cockatoos - emotional or highly complex cognitive behaviour or both? *Neurosci. Biobehav. Rev. [In Preparation].*

Kaplan, G., and Rogers, L. J. (2000). *The Orang-Utans. Their Evolution, Behavior, and Future.* Cambridge, M: Perseus Publishing.

Kaplan, G., and Rogers, L. J. (2004). "Charles Darwin and animal behaviour," in *Encyclopaedia of Animal Behaviour, 3 vols),* eds M. Bekoff, and J. Goodall (Westport CT: Greenwood Publishing), Introductory essay to vol. 2, 471–479.

Kappel, P., Hohenbrink, S., and Radespiel, U. (2011). Experimental evidence for olfactory predator recognition in wild mouse lemurs. *Am. J. Primatol.* 73, 928–938. doi: 10.1002/ajp.20963

Kappeler, P. M., Clutton-Brock, T., Shultz, S., and Lukas, D. (2019). Social complexity: patterns, processes, and evolution. *Behav. Ecol. Sociobiol.* 73, 5. doi: 10.1007/s00265-018-2613-4

Kemp, C., and Kaplan, G. (2012). Olfactory cues modify and enhance responses to visual cues in the common marmoset (*Callithrix jacchus*). *J. Primatol.* 1, 102. doi: 10.4172/jpmt.1000102

Kemp, C., and Kaplan, G. (2013). Facial expressions in common marmosets (*Callithrix jacchus*) and their use by conspecifics. *Anim. Cogn.* 16, 773–788. doi: 10.1007/s10071-013-0611-5

King, B. J. (2009). *The Dynamic Dance: Nonvocal Communication in African Great Apes.* Boston, MA: Harvard University Press.

King, S. L., Connor, R. C., and Montgomery, S. H. (2022). Social and vocal complexity in bottlenose dolphins. *Trends Neurosci.* 45, 881–883. doi: 10.1016/j.tins.2022.09.006

King, S. L., Friedman, W. R., Allen, S. J., Gerber, L., Jensen, F. H., et al. (2018). Bottlenose dolphins retain individual vocal labels in multi-level alliances. *Curr. Biol.* 28, 1993–1999. doi: 10.1016/j.cub.2018.05.013

Kita, S. (2003). *Pointing: Where Language, Culture, and Cognition Meet.* Hillsdale, NJ: Erlbaum.

Klein, R. G. (1977). The ecology of early man in Southern Africa: the relationship between man and environment is traced through 3 million years of Southern African prehistory. *Science* 197, 115–126. doi: 10.1126/science.197.4299.115

Knight, C., and Lewis, J. (2017). Wild voices: mimicry, reversal, metaphor, and the emergence of language. *Curr. Anthrop.* 58, 435–453. doi: 10.1086/692905

Kret, M. E., Prochazkova, E., Sterck, E. H., and Clay, Z. (2020). Emotional expressions in human and non-human great apes. *Neurosci. Biobehav. Rev.* 115, 378–395. doi: 10.1016/j.neubiorev.2020.01.027

Kuijper, B., Pen, I., and Weissing, F. J. (2012). A guide to sexual selection theory. *Ann. Rev. Ecol. Evol. Systemat.* 43, 2012. doi: 10.1146/annurev-ecolsys-110411-160245

Laland, K., Wilkins, C., and Clayton, N. (2016). The evolution of dance. *Curr. Biol.* 26, R5–R9. doi: 10.1016/j.cub.2015.11.031

Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrate for speech evolution. *Neurosci. Biobehav. Rev.* 83, 429–439. doi: 10.1016/j.neubiorev.2017.09.021

Launay, J., Tarr, B., and Dunbar, R. I. M. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology.* 122, 779–789. doi: 10.1111/eth.12528

Lazaro-Perea, C., Snowdon, C. T., and de Fatima Arruda, M. (1999). Scent-marking behavior in wild groups of common marmosets (*Callithrix jacchus*). *Behav. Ecol. Sociobiol.* 46, 313–324. doi: 10.1007/s002650050625

Lazarus, R. (1982). Thoughts on the relations between emotions and cognition. *Am. Psych.* 37, 1019–1024. doi: 10.1037/0003-066X.37.9.1019

Leadner, K., Sekely, L., Klein, R. M., and Gabay, S. (2021). Evolution of social attentional cues: evidence from the archerfish. *Cognition* 207, 104511. doi: 10.1016/j.cognition.2020.104511

Leavens, D. A. (2004). Manual deixis in apes and humans. *Interact. Stud.* 5, 387–408. doi: 10.1075/is.5.3.05lea

Leavens, D. A., and Hopkins, W. D. (1998). Intentional communication by chimpanzees: a cross-sectional study of the use of referential gestures. *Dev. Psychol.* 34, 813–822. doi: 10.1037/0012-1649.34.5.813

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implicationsfor processing models of language. *Front. Psychol.* 6, 731. doi: 10.3389/fpsyg.2015.00731

Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J. P., and Lemasson, A. (2019). Social bonding drives vocal exchanges in bonobos. *Sci. Rep.* 9, 711. doi: 10.1038/s41598-018-36024-9

Lewis, J. (2009). "As well as words: Congo Pygmy hunting, mimicry, and play," in *The Cradle of Language Vol.12: African Perspectives, Series: Studies in the Evolution of Language,* eds R. Botha and C. Knight (Oxford: Oxford University Press), 236–256.

Liebal, K., and Call, J. (2012). The origins of non-human primates' manual gestures. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 118–128. doi: 10.1098/rstb.2011.0044

Liebal, K., and Oña, L. (2018). Different approaches to meaning in primate gestural and vocal communication. *Front. Psychol.* 9, 478. doi: 10.3389/fpsyg.2018.00478

Lieberman, D. E. (2001). Another face in our family tree. *Nature* 410, 419–420. doi: 10.1038/35068648

Lieberman, P. (1985). On the evolution of human syntactic ability. Its pre-adaptive Bases—Motor control and speech. *J. Hum. Evol.* 14, 657–668. doi: 10.1016/S0047-2484(85)80074-9

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., and Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Dev. Sci.* 7, 297–307. doi: 10.1111/j.1467-7687.2004.00349.x

Locke, J. L., and Bogin, B. (2006). Language and life history: a new perspective on the development and evolution of human language. *Behav. Brain Sci.* 20, 259–325. doi: 10.1017/S0140525X0600906X

Logue, D. M., and Hall, M. L. (2014). Migration and the evolution of duetting in songbirds. *Proc. R. Soc. B* 281, 20140103. doi: 10.1098/rspb.2014.0103

Logue, D. M., and Krupp, D. B. (2016). Duetting as a collective behavior. *Front. Ecol. Evol.* 4, 7. doi: 10.3389/fevo.2016.00007

Loizos, K. (2017). *A Multiscale Computational Modeling Platform for Design and Analysis of Electrical Neural Stimulation (Doctoral Dissertation)*. The University of Utah.

Luengo-Kanacri, B. P., Eisenberg, N., Tramontano, C., Zuffiano, A., Caprara, M. G., Regner, E., et al. (2021). Measuring prosocial behaviors: psychometric properties and cross-national validation of the prosociality scale in five countries. *Front. Psych.* 12, 693174. doi: 10.3389/fpsyg.2021.693174

MaBouDi, H., Barron, A. B., Li, S., Honkanen, M., Loukola, O. J., Peng, F., et al. (2021). Non-numerical strategies used by bees to solve numerical cognition tasks. *Proc. R. Soc. B* 288, 20202711. doi: 10.1098/rspb.2020.2711

MacKinnon, J., and MacKinnon, K. (1980). The behavior of wild spectral tarsiers. *Int. J. Primat.* 1, 361–379. doi: 10.1007/BF02692280

Maestripieri, D. (1999). "Primate social organization, gestural repertoire size, and communication dynamics: a comparative study of macaques," in *The Evolution of Language: Assessing the Evidence from Nonhuman Primates,* ed B. J. King (Santa Fe, NM: School of American Research).

Marean, C. W. (2015). An evolutionary anthropological perspective on modern human origins. *Annu. Rev. Anthrop*. 44, 533–556. doi: 10.1146/annurev-anthro-102313-025954

Maretti, G., Sorrentino, V., Finomana, A., Gamba, M., and Giacoma, C. (2010). Not just a pretty song: an overview of the vocal repertoire of Indri indri. *J. Anthrop. Sci.* 88, 151–165.

Marshall-Ball, L., Mann, N., and Slater, P. J.B. (2006). Multiple functions to duet singing: hidden conflicts and apparent cooperation. *Anim. Behav*. 71, 823–831. doi: 10.1016/j.anbehav.2005.05.021

Martin, J. S., Koski, S. E., Bugnyar, T., Jaeggi, A. V., and Massen, J. J. (2021). Prosociality, social tolerance and partner choice facilitate mutually beneficial cooperation in common marmosets, *Callithrix jacchus*. *Anim. Behav*. 173, 115–136. doi: 10.1016/j.anbehav.2020.12.016

Mather, J. (2022). The case for octopus consciousness: temporality. *Neuro Sci.* 3, 245–261. doi: 10.3390/neurosci3020018

McCabe, B. J. (2019). Visual imprinting in birds: behavior, models, and neural mechanisms. *Front. Physiol*. 10, 658. doi: 10.3389/fphys.2019.00658

Meehan, A. J., Maughan, B., and Barker, E. D. (2019). Health and functional outcomes for shared and unique variances of interpersonal callousness and low prosocial behavior. *J. Psychopathol. Behav. Assess.* 41, 353–365. doi: 10.1007/s10862-019-09756-9

Mèndez-Càrdenas, M. G., and Zimmermann, E. (2009). Duetting—a mechanism to strengthen pair bonds in a dispersed pair-living primate (Lepilemur edwardsi)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Meyer, J. (2008). Typology and acoustic strategies of whistled languages: phonetic comparison and perceptual cues of whistled vowels. *J. Int. Phonet. Assoc*. 38, 69–94. doi: 10.1017/S0025100308003277

Michael, J., McEllin, L., and Felber, A. (2020). Prosocial effects of coordination–What, how and why? *Acta Psychol.* 207, 103083. doi: 10.1016/j.actpsy.2020.103083

Miles, H. L. (1990). "The cognitive foundations for reference in a signing orangutan" in *"Language" and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives.* eds S. T. Parker, and K. R. Gibson (Cambridge: Cambridge University Press), 511–539.

Miles, H. L. (1994). "Chantek: the language ability of an enculturated orangutan (Pongo pygmaeus)," in *Proc. Intern. Orang Utan Conf.* (Davis, CA: UCLA).

Miles, L. S., Rivkin, L. R., Johnson, M. T., Munshi-South, J,., and Verrelli, B. C. (2019). Gene flow and genetic drift in urban environments. *Mol. Ecol.* 28, 4138–4151. doi: 10.1111/mec.15221

Miller, E. N., Hof, P. R., Sherwood, C. C., and Hopkins, W. D. (2021). The paracingulate sulcus is a unique feature of the medial frontal cortex shared by great apes and humans. *Brain Behav. Evol.* 96, 26–36. doi: 10.1159/000517293

Miller, M. R. (2013). Descartes on animals revisited. *J. Philos. Res.* 38, 89–114. doi: 10.5840/jpr2013386

Mitani, J. C., and Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour.* 135, 1041–1064.

Mithen, S. (2006). *The Singing Neanderthals* (London: Phoenix Paperback/Orion Books).

Morrison, R. E., Eckardt, W., Stoinski, T. S., and Brent, L. J. N. (2020). Comparing measures of social complexity: larger mountain gorilla groups do not have a greater diversity of relationships. *Proc. R. Soc. B* 287, 20201026. doi: 10.1098/rspb.2020.1026

Neff, E. P. (2019). Neural control of duets between Alston's singing mice, an emerging vocalization model. *Lab. Anim.* 48, 137–137. doi: 10.1038/s41684-019-0293-y

Nimchinsky, E. A., Gilissen, E., Allman, J. M., Perl, D. P., Erwin, J. M., and Hof, P. R. (1999). A neuronal morphologic type unique to humans and great apes. *PNAS* 96, 5268–5273. doi: 10.1073/pnas.96.9.5268

Nishimura, T., Tokuda, I. T., Miyachi, S., Dunn, J. C., Herbst, C. T., and Imai, H. (2022). Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science* 377, 760–763. doi: 10.1126/science.abm1574

Norscia, I., and Palagi, E. (2011). Yawn contagion and empathy in Homo sapiens. *PLoS ONE* 6, e28472. doi: 10.1371/journal.pone.0028472

O'Connell, L. A., and Hofmann, H. A. (2011). The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. *J. Comp. Neurol.* 519, 3599–3639. doi: 10.1002/cne.22735

Okobi, D. E. Jr., Banerjee, A., Matheson, A. M. M., Phelps, S. M., and Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988. doi: 10.1126/science.aau9480

Oller, D. K., and Griebel, U. (2021). Functionally flexible signaling and the origin of language. *Front. Psych.* 11, 626138. doi: 10.3389/fpsyg.2020.626138

Oller, D. K., and Griebel, U. (eds.). (2008). *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability*. Cambridge, MA: MIT Press.

Osvath, M., and Sima, M. (2014). Sub-adult ravens synchronize their play: a case of emotional contagion. *Anim.Behav. Cogn.* 1, 197–205. doi: 10.12966/abc.05.09.2014

Palomero-Gallagher, N., and Zilles, K. (2019). Differences in cytoarchitecture of Broca's region between human, ape and macaque brains. *Cortex* 118, 132–153. doi: 10.1016/j.cortex.2018.09.008

Panksepp, J. (2005). Beyond a joke: from animal laughter to human joy? *Science* 308, 62–63. doi: 10.1126/science.1112066

Parr, L. A., and Waller, B. M. (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Soc. Cogn. Affect. Neurosci.* 1, 221–228. doi: 10.1093/scan/nsl031

Partan, S. R. (2002). Single and multichannel signal composition: facial expressions and vocalizations of rhesus macaques (*Macaca mulatta*). *Behaviour* 139, 993–1027. doi: 10.1163/15685390260337877

Partan, S. R., and Marler, P. (1999). Communication goes multimodal. *Science* 283, 1272–1273. doi: 10.1126/science.283.5406.1272

Patterson, F. G. (1978). "Linguistic capabilities of a lowland gorilla," in *Sign Language and Language Acquisition in Man and Ape: New Dimensions in Comparative Pedolinguistics*, ed F. C. C. Peng (Boulder: Westview Press), 161–201.

Pellis, S. M., and Pellis, V. C. (2007). Rough-and-tumble play and the development of the social brain. *Curr. Dir. Psychol. Sci.* 16, 95–98. doi: 10.1111/j.1467-8721.2007.00483.x

Penn, D. C., and Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philos. Trans. R. Soc. B* 362, 731–744. doi: 10.1098/rstb.2006.2023

Pepperberg, I. M. (2007). Grey parrots do not always 'parrot': the roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Lang. Sci.* 29, 1–13. doi: 10.1016/j.langsci.2005.12.002

Petr, M., Hajdinjak, M., Fu, Q., Essel, E., Rougier, H., Crevecoeur, I., et al. (2020). The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science* 369, 1653–1656. doi: 10.1126/science.abb6460

Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. *Front. Neurosci.* 11, 542. doi: 10.3389/fnins.2017.00542

Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Ravignani, A., et al. (2023). Isochrony and rhythmic interaction in ape duetting. *Proc. R. Soc. B* 290, 20222244. doi: 10.1098/rspb.2022.2244

Raposa, E. B., Laws, H. B., and Ansell, E. B. (2016). Prosocial behavior mitigates the negative effects of stress in everyday life. *Clin. Psychol. Sci.* 4, 691–698. doi: 10.1177/2167702615611073

Reader, S. M., and Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *PNAS* 99, 4436–4441. doi: 10.1073/pnas.062041299

Reddish, P., Fischer, R., and Bulbulia, J. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PLoS ONE* 8, e71182. doi: 10.1371/journal.pone.0071182

Reindl, V., Gerloff, C., Scharke, W., and Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *NeuroIm* 178, 493–502. doi: 10.1016/j.neuroimage.2018.05.060

Rek, P., and Magrath, R. D. (2020). Visual displays enhance vocal duet production and the perception of coordination despite spatial separation of partners. *Anim. Behav.* 168, 231–241. doi: 10.1016/j.anbehav.2020.08.002

Robbins, M. M. (1999). Male mating patterns in wild multimale mountain gorilla groups. *Anim. Behav.* 57, 1013–1020. doi: 10.1006/anbe.1998.1063

Roberts, A. I., and Roberts, S. G.B. (2020). Communicative roots of complex sociality and cognition. *Biol. Rev.* 1, 51–73. doi: 10.1111/brv.12553

Roelfsema, P. R., Engel, A. K., König, P., and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nat. Rev. Neurosci.* 385, 157–161. doi: 10.1038/385157a0

Rogers, L. J., Stewart, L., and Kaplan, G. (2018). Food calls in common marmosets, *Callithrix jacchus*, and evidence that one is functionally referential. *Animal* 8, 99. doi: 10.3390/ani8070099

Ros, S. (2021). Rhythm-speech correlations in a corpus of senegalese drum language. *Front. Commun.* 6, 643683. doi: 10.3389/fcomm.2021.643683

Rose, M. C., Styr, B., Schmid, T. A., Elie, J. E., and Yartsev, M. M. (2021). Cortical representation of group social communication in bats. *Science* 374, eaba9584. doi: 10.1126/science.aba9584

Rugani, R., Regolin, L., and Vallortigara, G. (2011). Summation of large numerousness by newborn chicks. *Front. Psychol.* 2, 179. doi: 10.3389/fpsyg.2011.00179

Russon, A. E., Bard, K. A., and Parker, S. T. (eds.). (1996). *Reaching into Thought: The Minds of Great Apes*. Cambridge. Cambridge University Press.

Salovey, P., and Mayer, J. D. (1990). Emotional intelligence. *Imag. Cogn. Person*. 9, 185–211. doi: 10.2190/DUGG-P24E-52WK-6CDG

Samuni, L., Tkaczynski, P., Deschner, T., Löhrrich, T., Wittig, R. M., and Crockford, C. (2020). Maternal effects on offspring growth indicate post-weaning juvenile dependence in chimpanzees (Pan troglodytes verus). *Front. Zool.* 17, 1–12. doi: 10.1186/s12983-019-0343-8

Santangelo, J. S., Johnson, M. T., and Ness, R. W. (2018). Modern spandrels: the roles of genetic drift, gene flow and natural selection in the evolution of parallel clines. *Proc. R. Soc. B Biol. Sci.* 285, 20180230. doi: 10.1098/rspb.2018.0230

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L, and Fitch, W. T. (2020). Music as a coevolved system for social bonding. *Behav. Brain Sci.* 44, e59. doi: 10.31234/osf.io/qp3st

Savage-Rumbaugh, E. S. (1984). "Pan paniscus and Pan troglodytes: contrast in preverbal communicative competence," in *The Pygmy Chimpanzee: Evolutionary Biology and Behavior*, ed R. L. Susman (New York, NY: Plenum Press), 395–413.

Schacht, R., and Kramer, K. L. (2019). Are we monogamous? A review of the evolution of pair-bonding in humans and its contemporary variation cross-culturally. *Front. Ecol. Evol.* 230. doi: 10.3389/fevo.2019.00230

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *PNAS*. 107, 9388–9393. doi: 10.1073/pnas.1001791107

Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Anim. Behav.* 76, 1977–1988. doi: 10.1016/j.anbehav.2008.07.032

Schuppe, E. R., Cantin, L., Chakraborty, M., Biegler, M. T., Jarvis, E. R., Chen, C.-C., et al. (2022). Forebrain nuclei linked to woodpecker territorial drum displays mirror those that enable vocal learning in songbirds. *PLoS Biol.* 20, e3001751. doi: 10.1371/journal.pbio.3001751

Seifart, F., Meyer, J., Grawunder, S., and Dentel, L. (2018). Reducing language to rhythm: amazonian Bora drummed language exploits speech rhythm for long-distance communication. *R. Soc. Open Sci.* 5, 170354. doi: 10.1098/rsos.170354

Sekulic, R., and Chivers, D. J. (1986). The significance of call duration in howler monkeys. *Int. J. Primat.* 7, 183–190. doi: 10.1007/BF02692317

Senut, B., Pickford, M., Gommery, D., and Sègalen, L. (2018). Palaeoenvironments and the origin of hominid bipedalism. *Hist. Biol.* 30, 284–296. doi: 10.1080/08912963.2017.1286337

Sewall, K. B. (2015). Social complexity as a driver of communication and cognition. *Integr. Comp. Biol.* 55, 384–395. doi: 10.1093/icb/icv064

Seyfarth, R. M., and Cheney, D. L. (1986). Vocal development in vervet monkeys. *Anim. Behav.* 34, 1640–1658. doi: 10.1016/S0003-3472(86)80252-4

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: classification and semantic communication. *Science* 210, 801–803. doi: 10.1126/science.7433999

Shanker, S. G., and King, B. J. (2002). The emergence of a new paradigm in ape language research: beyond interactionism. *Behav. Brain Sci.* 25, 646–651. doi: 10.1017/S0140525X02510116

Sherwood, C. C., Subiaul, F., and Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *J. Anat.* 212, 426–454. doi: 10.1111/j.1469-7580.2008.00868.x

Silk, J. B. (2007). The strategic dynamics of cooperation in primate groups. *Adv. Study of Behav.* 37, 1–41.

Singletary, B., and Tecot, S. (2020). Multimodal pair-bond maintenance: a review of signalling across modalities in pair-bonded nonhuman primates. *Am. J. Primatol.* 82, e23105. doi: 10.1002/ajp.23105

Smaers, J. B., Gómez-Robles, A., Parks, A. N., and Sherwood, C. C. (2017). Exceptional evolutionary expansion of prefrontal cortex in great apes and humans. *Curr. Biol.* 27, 714–720. doi: 10.1016/j.cub.2017.01.020

Smaers, J. B., and Vanier, D. R. (2019). Brain size expansion in primates and humans is explained by a selective modular expansion of the cortico-cerebellar system. *Cortex* 118, 292–305. doi: 10.1016/j.cortex.2019.04.023

Smith, A. S., Ågmo, A., Birnie, A. K., and French, J. A. (2010). Manipulation of the oxytocin system alters social behavior and attraction in pair-bonding primates, Callithrix penicillata. *Horm. Behav.* 57, 255–262. doi: 10.1016/j.yhbeh.2009.12.004

Sneve, M. H., Grydeland, H., Rosa, M. G., Paus, T., Chaplin, T., and Fjell, A. M. (2019). High-expanding regions in primate cortical brain evolution support supramodal cognitive flexibility. *Cerebr. Cortex* 29, 3891–3901. doi: 10.1093/cercor/bhy268

Snowdon, C. T. (2018). Cognitive components of vocal communication: a case study. *Animal* 8, 126. doi: 10.3390/ani8070126

Snowdon, C. T. (2020). Vervet monkey alarm calls: Setting the historical context. *Anim. Behav. Cogn.* 7, 87–94. doi: 10.26451/abc.07.02.02.2020

Sommer, V., Bauer, J., Fowler, A., and Ortmann, S. (2011). "Patriarchal chimpanzees, matriarchal bonobos: Potential ecological causes of a Pan dichotomy," in *Primates of Gashaka*, eds V. Sommer, and C. Ross (New York, NY: Springer), 469–501.

Spataro, P., Calabrò, M., and Longobardi, E. (2020). Prosocial behaviour mediates the relation between empathy and aggression in primary school children. *Eur. J. Dev. Psychol.* 17, 727–745. doi: 10.1080/17405629.2020.1731467

Spocter, M. A., Hopkins, W. D., Garrison, A. R., Bauernfeind, A. L., Stimpson, C. D., Hof, P. R., et al. (2010). Wernicke's area homologue in chimpanzees (Pan troglodytes) and its relation to the appearance of modern human language. *Proc. R. Soc. B Biol. Sci.* 277, 2165–2174. doi: 10.1098/rspb.2010.0011

Sterling, E. J., and McCreless, E. E. (2007). "Adaptations in the aye-aye: a review," in *Lemur: Ecology and Adaptation*, eds L. Gould (New York, NY: Sauther Springer Science+Business Media LLC), 159–184.

Stevenson, M. F., and Poole, T. B. (1976). An ethogram of the common marmoset (*Calithrix jacchus* jacchus): general behavioural repertoire. *Anim. Behav.* 24, 428–451. doi: 10.1016/S0003-3472(76)80053-X

Stewart, K. M. (1994). Early hominid utilisation of fish resources and implications for seasonality and behaviour. *J. Hum. Evol.* 27, 229–245. doi: 10.1006/jhev.1994.1044

Studdert-Kennedy, M. (2000). "Evolutionary implications of the particulate principle: imitation and the dissociation of phonetic form from semantic function," in *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, eds C. Knight, M. Studdert-Kennedy, and J. Hurford (Cambridge: CUP), 161–176. doi: 10.1017/CBO9780511606441.011

Studdert-Kennedy, M., and Goldstein, L. (2003). Launching language: the gestural origin of discrete infinity. *Stud. Evol. Lang.* 3, 235–254. doi: 10.1093/acprof:oso/9780199244843.003.0013

Suzuki, A. (1970). An ecological study of Chimpanzees in a Savanna woodland. *Primates* 10, 103–148. doi: 10.1007/BF01730979

Takahashi, D. Y., Fenley, A. R., and Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Trans. R. Soc. B* 371, 20150370. doi: 10.1098/rstb.2015.0370

Templeton, A. (2002). Out of Africa again and again. *Nature* 416, 45–51. doi: 10.1038/416045a

Thomas, E. (2020). Descartes on the animal within, and the animals without. *Can. J. Philos.* 50, 999–1014. doi: 10.1017/can.2020.44

Tilson, R. L., and Norton, P. M. (1981). Alarm duetting and pursuit deterrence in an African antelope. *Am. Nat.* 118, 455–462. doi: 10.1086/283840

Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., and Nakagawa, S. (2016). Territoriality, social bonds, and the evolution of communal signalling in birds. *Front. Ecol. Evol.* 4, 74. doi: 10.3389/fevo.2016.00074

Tobias, M. L., Viswanathan, S. S., and Kelley, D. B. (1998). Rapping, a female receptive call, initiates male–female duets in the South African clawed frog. *PNAS* 95, 1870–1875. doi: 10.1073/pnas.95.4.1870

Tobias, P. V. (1987). The brain of *Homo habilis*: a new level of organization in cerebral evolution. *J. Hum. Evol.* 16, 741–761. doi: 10.1016/0047-2484(87)90022-4

Tommasi, L., and Vallortigara, G. (2004). Hemispheric processing of landmark and geometric information in male and female domestic chicks (*Gallus gallus*). *Behav. Brain Res.* 155, 85–96. doi: 10.1016/j.bbr.2004.04.004

Townsend, S. W., and Manser, M. B. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology* 119, 1–11. doi: 10.1111/eth.12015

Trenbeath, C. (2021). *Yakety Sacs: Laryngeal Air Sac Usage in Great Apes, Master of Science in Integrative Biology*. Kennesaw, GA: Kennesaw State University.

Valdesolo, P., Ouyang, J., and De Steno, D. (2010). The rhythm of joint action: synchrony promotes cooperative ability. *J. Exp. Soc. Psychol.* 46, 693–695. doi: 10.1016/j.jesp.2010.03.004

Vallortigara, G., Chiandetti, C., Rugani, R., Sovrano, V. A., and Regolin, L. (2010). Animal cognition. *Wiley Interdisc. Reviews: Cogn. Sci.* 1, 882–893. doi: 10.1002/wcs.75

Vallortigara, G., Zanforlin, M., and Pasti, G. (1990). Geometric modules in animal's spatial representation: a test with chicks. *J. Comp. Psych.* 104, 248–254. doi: 10.1037/0735-7036.104.3.248

van Lawick-Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe Stream Reserve. *Anim. Behav. Monogr.* 1, 61-IN12. doi: 10.1016/S0066-1856(68)80003-2

Vanderhoff, E. N., and Bernal Hoverud, N. (2022). Perspectives on antiphonal calling, duetting and counter-singing in non-primate mammals: an overview with notes on the coordinated vocalizations of bamboo rats (Dactylomys spp., Rodentia: Echimyidae). *Front. Ecol. Evol.* 10, 906546. doi: 10.3389/fevo.2022.906546

Vanderschuren, L. J., Achterberg, E. M., and Trezza, V. (2016). The neurobiology of social play and its rewarding value in rats. *Neurosci. Biobehav. Rev.* 70, 86–105. doi: 10.1016/j.neubiorev.2016.07.025

Veit, W. (2021). Scaffolding natural selection. *Biol. Theory* 17, 163–180. doi: 10.1007/s13752-021-00387-6

Verspeek, J.,van Leeuwen, E. J. C.,Laméris, D. W.,Staes, N., and Stevens, J. M. G. (2022). Adult bonobos show no prosociality in both prosocial choice task and group service paradigm. *PeerJ.* 10, e12849. doi: 10.7717/peerj.12849

Vieira, M., Amorim, M. C., and Fonseca, P. J. (2021). Vocal rhythms in nesting Lusitanian toadfish, *Halobatrachus didactylus*. *Ecol. Inform.* 63, 101281. doi: 10.1016/j.ecoinf.2021.101281

Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., and Heinrichs, M. (2012). The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. *Psych. Sci.* 23, 651–660. doi: 10.1177/0956797611431576

Vonk, J. (2020). Forty years on from the question of referential signals in nonhuman communication. *Anim. Behav. Cogn.* 7, 82–86. doi: 10.26451/abc.07.02.01.2020

Waller, B. M., Liebal, K., Burrows, A. M., and Slocombe, K. E. (2013). How can a multimodal approach to primate communication help us understand the evolution of communication? *Evol. Psychol.* 11, 147470491301100305. doi: 10.1177/147470491301100305

Warneken, F. (2015). Precocious prosociality: Why do young children help? *Child Dev. Perspect.* 9, 1–6. doi: 10.1111/cdep.12101

Warneken, F., Hare, B., Melis, A. P., Hanus, D., and Tomasello, M. (2007). Spontaneous altruism by chimpanzees and young children. *PLoS Biol.* 5, e184. doi: 10.1371/journal.pbio.0050184

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., et al. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Curr. Biol.* 25, 495–499. doi: 10.1016/j.cub.2014.12.032

Watts, M. E., Pocock, R., and Claudianus, C. (2018). Brain energy and oxygen metabolism: emerging role in normal function and disease. *Front. Mol. Neurosci.* 11, 216. doi: 10.3389/fnmol.2018.00216

Wheeler, B. C., and Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evol. Anthrop. Issues News Rev.* 21, 195–205. doi: 10.1002/evan.21319

White, T. D., Asfaw, B., Beyene, Y., Hailie-Selassie, Y., Lovejoy, C. O., Suwa, G., et al. (2009). *Ardipithecus ramidus* and the paleobiology of early hominids. *Science* 326, 75–86. doi: 10.1126/science.1175802

Whiten, A., and de Waal, E. (2018). The pervasive role of social learning in primate lifetime development. *Behav. Ecol. Sociobiol.* 72, 1–16. doi: 10.1007/s00265-018-2489-3

Wiles, J., Watson, J., Tonkes, B., and Deacon, T. W. (2005). Transient phenomena in learning and evolution: genetic assimilation and genetic redistribution. *Artif. Life* 11, 177–188. doi: 10.1162/10645460532 79026

Williams, L., Shultz, S., and Jensen, K. (2022). The primate workplace: cooperative decision-making in human and non-human primates. *Front. Ecol. Evol. Sect. Behav. Evo. Ecol.* 10, 887187. doi: 10.3389/fevo.2022.887187

Wrangham, R. (1979). On the evolution of ape social systems. *Soc. Sci. Info* 18, 336–368. doi: 10.1177/053901847901800301

Yu, L., and Tomonaga, M. (2015). Interactional synchrony in chimpanzees: examination through finger tapping experiment. *Sci. Rep.* 5, 1–9. doi: 10.1038/srep10218

Yu, Y., Karbowski, J., Sachdev, R. N. S., and Feng, J. (2014). Effect of temperature and glia in brain size enlargement and origin of allometric body-brain size scaling in vertebrates. *BMC Evo. Biol.* 14, 178–192. doi: 10.1186/s12862-014-0178-z

Zhang, W., and Yartsev, M. M. (2019). Correlated neural activity across the brains of socially interacting bats. *Cell* 178, 413–428. doi: 10.1016/j.cell.2019.05.023

Zhou, W., Jiang, Y., He, S., and Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Curr. Biol.* 20, 1356–1358. doi: 10.1016/j.cub.2010.05.059

Frontiers in Ecology and Evolution

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fevo.2023.956922

OPEN ACCESS

EDITED BY
Patrice Adret,
Universidad Autónoma Gabriel René Moreno, Bolivia

REVIEWED BY
Sofya Dolotovskaya,
German Primate Center, Germany
Chiara De Gregorio,
University of Turin, Italy

*CORRESPONDENCE
Michal Hradec
✉ hradecm@af.czu.cz

RECEIVED 30 May 2022
ACCEPTED 31 May 2023
PUBLISHED 28 July 2023

CITATION
Hradec M, Illmann G, Prikrylová M, Bolechová P and Vostrá-Vydrová H (2023) Difference in the songs of paired and unpaired southern yellow-cheeked gibbon males (*Nomascus gabriellae*): social status or age? *Front. Ecol. Evol.* 11:956922.
doi: 10.3389/fevo.2023.956922

Difference in the songs of paired and unpaired southern yellow-cheeked gibbon males (*Nomascus gabriellae*): social status or age?

Michal Hradec[1]*, Gudrun Illmann[1,2], Martina Prikrylová[1], Petra Bolechová[1] and Hana Vostrá-Vydrová[1]

[1]Department of Ethology and Companion Animal Science, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czechia, [2]Department of Ethology, Institute of Animal Science, Praha Uhříněves, Prague, Czechia

All gibbons (Primates: Hylobatidae) are well known for emitting loud vocalizations specific for species and sex. The songs of paired and unpaired male southern yellow-cheeked gibbons (*Nomascus gabriellae*) are characterized by the presence of staccato notes and multi-modulation phrases with two or more extremely rapid frequency modulations in the second note of each phrase. In addition, paired males also produce a coda vocalization, which is similar to the multi-modulation phrase of male calls but lacks the staccato notes and always occurs directly following the female great call as part of a pair-specific duet. The aim of this study was first to assess whether the songs of paired and unpaired males can be acoustically distinguished from one another and second, whether the coda vocalization differs from the multi-modulation phrase of the male call in paired males. To assess these issues, we analyzed 616 songs obtained from a long-term study of vocal development in 14 captive adult males (>7 years old), half of which were unpaired and significantly younger than paired subjects. For each song, we quantified nine acoustic features, for which we applied a suite of linear mixed effects models with social status as a fixed variable and age as a regression coefficient. This allowed us to compare (1) the structure of male calls (staccato notes and multi-modulation phrase) between paired and unpaired subjects, (2) the muti-modulation phrase of unpaired subjects to the coda vocalization of paired subjects, and (3) the multi-modulation phrase of paired males to the coda vocalization. We found that the male call of younger-unpaired subjects had a longer duration, broader frequency range, higher maximum frequency, and fewer staccato notes than their counterparts in paired subjects. The coda vocalization of older-paired males exhibited a larger number of frequency modulations than the multi-modulation phrase of all males. While the male call of younger-unpaired males differs from both the male call and the coda vocalization of older-paired males, further studies are necessary to disentangle the effects of age and pairing status.

KEYWORDS

gibbon, *Nomascus*, vocalization, coda, male call

1. Introduction

Vocal communication, unlike visual signaling, can convey information about the sender over long distances. Such information includes identity (Owren and Rendall, 2003; Price et al., 2009), social rank (Fischer et al., 2004), sex, and body size (Pfefferle and Fischer, 2006; Ey et al., 2007). Along with physical growth, changes in vocalization with increasing age were confirmed in birds (Nemeth et al., 2012) as well as in aquatic (Sanvito et al., 2008; Umeed et al., 2018) and terrestrial mammals (Reby and McComb, 2003; Charlton et al., 2009; Briefer et al., 2010).

Although various forms of vocal signaling have been documented in several mammalian taxa, such as primates, rodents, bats, and cetaceans (for recent reviews see Banerjee et al., 2019; Vernes and Wilkinson, 2020; Janik and Knornschild, 2021; De Gregorio et al., 2022; Vanderhoff and Hoverud, 2022), the critical acoustic features involved during inter-individual vocal exchanges, duetting included, are little known. Some advances have been made in singing primates, notably in Sulawesi tarsiers (genus Tarsius, Clink et al., 2020), indris (*Indri indri,* Torti et al., 2013), coppery titi monkeys (*Plecturocebus cupreus*, Lau et al., 2020), and gibbons (Geissmann, 2002). Additional studies have focused on song dynamics (e.g., rhythm, pitch, and degree of overlap) (Gamba et al., 2016), song flexibility (Clarke et al., 2006; Terleph et al., 2018a; Hradec et al., 2021a), song individuality (Sun et al., 2011; Clink et al., 2021), song ontogeny (Hauser, 1989; Pistorio et al., 2006; Hradec et al., 2017, 2021b; De Gregorio et al., 2021), song recognition (Raemaekers and Raemaekers, 1985; Mitani, 1987; Caselli et al., 2015), song taxonomy, and genetic relatedness (Konrad and Geissmann, 2006; Torti et al., 2017). However, the question of whether songs by singing primates transmit information about pairing status (i.e., paired vs. unpaired individuals) has received surprisingly little attention.

Gibbons (family Hylobatidae) are a uniform group of territorial and mostly pair-living apes that are well known for emitting loud, stable patterns of vocalizations specific for species and sex (Geissmann, 2002; but see Reichard et al., 2016; De Gregorio et al., 2022). The vocalizations are often referred to as "songs", uttered in succession and forming a recognizable temporal sequence of pure, melodic sounds (Thorpe, 1961; Haimoff, 1984a; Geissmann, 2002; Supplementary Table 1).

Gibbon songs are relatively stereotyped and are thought to be under strong genetic constraints (Brockelman and Schilling, 1984; Geissmann, 1984). For most gibbon species, paired individuals (i.e., adult male and adult female) often combine their respective songs into coordinated duets, except for the silvery gibbon (*Hylobates moloch*) and the Kloss's gibbon (*Hylobates klossii*) in which paired individuals produce only solo songs (Tenaza, 1976; Geissmann and Nijman, 2006). Duet songs are an alternation of sex-specific vocalizations contributed to by the female and her mate (Haimoff, 1984a). A typical male-female gibbon duet begins with a few introductory notes from both mates, followed by the female great call. During the build-up phase of the female great call, the paired male ceases his song and, after the completion of the great call, adds a coda vocalization. The coda vocalization is produced only by paired males and always occurs directly following the female great call (Geissmann, 2002). Subsequently, the paired male repeats several male calls during the so-called interlude sequence (Figure 1) until the female begins her next great call (Geissmann, 2002; Konrad and Geissmann, 2006).

The vocal structure of the coda vocalization and the male call has been studied in several gibbon species (Geissmann, 2002; Konrad and Geissmann, 2006; Terleph et al., 2018b). The stable structure of the coda vocalization has been found to be particularly useful in determining individuality and taxonomy in male gibbons, both in the absence of the male call (Sun et al., 2011; Lau et al., 2018; Clink et al., 2021) and in conjunction with the male call (Wanelik et al., 2013). However, recent studies in white-handed gibbons (*Hylobates lar*) have shown that the stable structure of both the male call and the female great call declines with age (Barelli et al., 2013; Terleph et al., 2016).

Wild unpaired Kloss's gibbons (*Hylobates klossii*) and white-handed gibbons (*Hylobates lar*) aged 8–9 years disperse from their natal group and become "floaters", i.e., individuals in search of a mate and territory (Tilson, 1981; Brockelman and Gittins, 1984; Cowlishaw, 1992). These unpaired adult males emit loud calls, exclusively as a solo song, without the coda vocalizations (Brockelman and Srikosamatara, 1984; Mitani, 1988; Cowlishaw, 1992). This raises the question of whether male calls differ according to pairing status, which could then be used by conspecific (female) receivers to assess the emitter's social status. The only available study (Mitani, 1988) did not find any evidence of such differences between the songs of paired and unpaired male agile gibbons (*Hylobates agilis*). In addition, playback experiments revealed that females (and other individuals) failed to distinguish paired males from unpaired males, based solely on vocal cues (Mitani, 1988). In non-human primates, paired males are usually older than unpaired males, at least in captivity, and it is also known that the acoustic parameters of vocalization change with advancing age (Simakobu monkeys: Erb et al., 2013; and titi monkeys: Clink et al., 2019; e.g., Japanese macaques: Inoue, 1988) or age range (e.g., baboons: Fischer et al., 2002, 2004). In fact, an age-related decline in acoustic parameters (i.e., duration and frequency) has been shown in paired males and females of white-handed gibbons (Barelli et al., 2013; Terleph et al., 2016). However, the influence of age on acoustic structure in unpaired males is completely unknown.

Given such limited information on male calls, it might be worth exploring, in other gibbon species, whether male songs differ according to pairing status or because of aging. To do this, we analyzed available data from a long-term study of vocalizations in captive southern yellow-cheeked gibbons (*Nomascus gabriellae*). In both paired and unpaired individuals of this species, the male call is characterized by the presence of irregular series of soft staccato notes that sometimes precede a multi-modulation phrase (Figures 1A, B). A previous report in the same species established that, when reaching 7.1 years of age, unpaired males produce the complete male call as a solo song (Hradec et al., 2021b), while paired males include it in duets with females (Haimoff, 1984a; Geissmann, 2002; Konrad and Geissmann, 2006).

In addition, paired males also produce a coda vocalization (Figure 1A) which is similar to the multi-modulation phrase in solo and duet songs of a male call but has a more stereotyped structure, and it is always produced immediately after the female great call without staccato notes (Geissmann, 2002). The multi-modulation

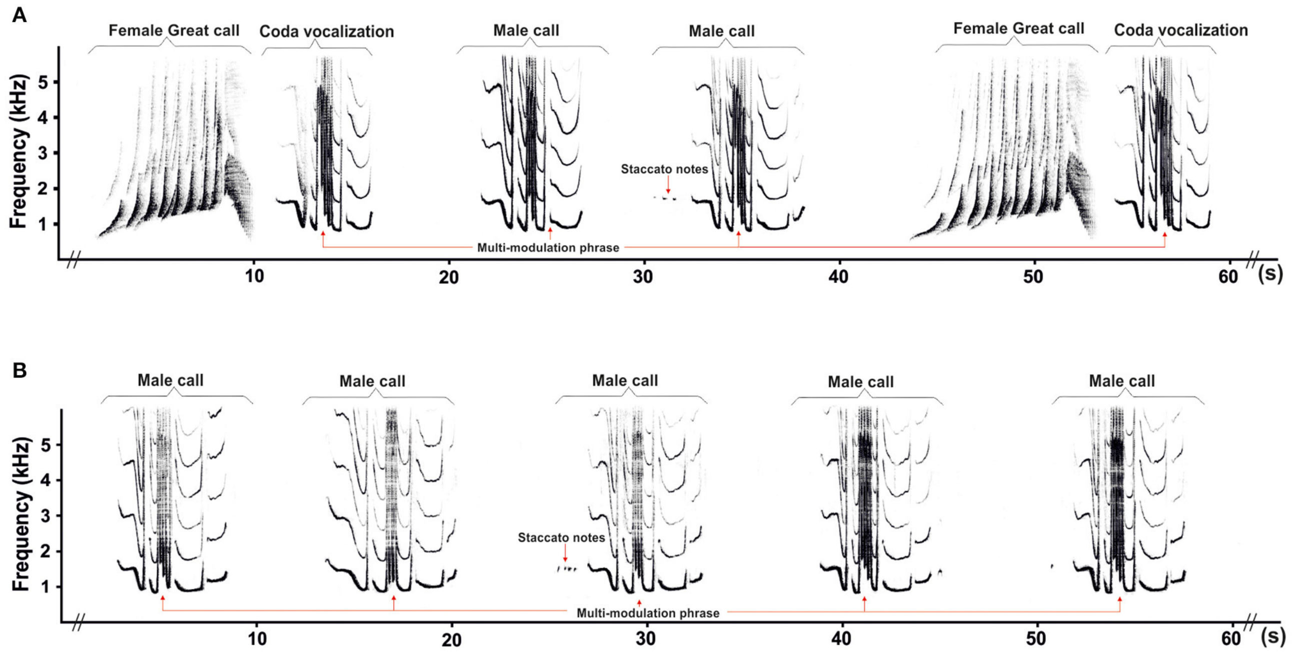


FIGURE 1
Representative spectrograms showing a duet song **(A)** and an interlude sequence of a solo song **(B)** in southern yellow-cheeked gibbons.

phrase is the predominant acoustic structure of the male call and is divided into several notes. The second note of the multi-modulation phrase and the coda vocalization feature two or more extremely rapid frequency modulations (a steep up-and-down sweeping) that set southern yellow-cheeked gibbons apart from other species of the genus *Nomascus* (Konrad and Geissmann, 2006).

It has been suggested that the extremely rapid changes in the frequency modulations in the second note of the multi-modulation phrase and coda vocalization (also called "roulade"; Demars and Goustard, 1978 or "roll"; Konrad and Geissmann, 2006) are reminiscent of songbird trills. The high repetition rate of those trills is costly to produce and might be considered an accurate signal of male quality (Gil and Gahr, 2002; Ballentine, 2009; Cramer, 2013) but we do not know whether this applies to females. Among non-human primates, a trade-off between call rate and frequency bandwidth has been first reported in the trills of the female great call of Bornean gibbons (*Hylobates muelleri*) but whether these performance constraints inform recipients of female quality remains to be investigated (Clink et al., 2018). At the same time, it has been shown that acoustic signals of longer durations, higher frequencies, and wider bandwidths are reliable indicators of the unpaired status of male rufous-and-white wrens, (*Thryothorus rufalbus*), that, in turn, could prove attractive to females (Hennin et al., 2009).

To date, it is still unclear whether structural differences exist in the multi-modulation phrase between paired and unpaired southern yellow-cheeked gibbon males. For instance, male calls displaying a wider frequency range and longer duration might be an indicator of unpaired status or else might reflect an age difference. Furthermore, it is essentially unknown whether the coda vocalization—as an integral part of the duet—structurally differs from the multi-modulation phrase of both paired (in duet song) and unpaired (in solo song) individuals. Such differences have been suggested based on the visual inspection of the calls. For instance, the coda vocalization of the Hainan black gibbon, *Hylobates concolor hainanus*—since renamed Hainan gibbon (*Nomascus hainanus*)—has a longer duration and displays a higher number of frequency modulations in the second note than that found in the multi-modulation phrase of the male call (Haimoff, 1984a,b). This indicates that, in paired individuals, the coda vocalization is structurally more complex. However, an in-depth analysis of the acoustical structure comparing the coda vocalization to male calls is still lacking.

The present study is based on a unique but unequally distributed dataset in which unpaired adult males were significantly younger (age: 7.7–10.10 years) than paired males (age: 11.4–34 years), which may not necessarily reflect the situation in the wild (Mitani, 1988). Such artificial distribution in the group composition of zoo-housed animals relies on the breeding management recommendations, as stated by the European Association of Zoos and Aquaria (EAZA) *Ex-situ* Program (EEP). The possibility of obtaining vocal samples from both paired and unpaired males that would overlap in age distribution is thus limited by the number of animals kept and the EEP regulations. This study is the first of its kind to compare the acoustical structure of unpaired and paired southern yellow-cheeked gibbon males while considering the age factor.

Our aim is to assess (1) whether social status and age synergically influence the song structure in these two groups of captive male southern yellow-cheeked gibbons namely by comparing a set of acoustic features in the male call and coda vocalizations of older-paired vs. younger-unpaired males and (2) whether the coda vocalization, indeed, differs from the male calls among older-paired males of the same species, as was found in the Hainan gibbon (Haimoff, 1984a,b). First, we predicted that, in comparison to older-paired males, younger-unpaired males would produce songs characterized by a higher pitch and longer duration, predominantly in the second note of the multi-modulation phrase. Second, we also predicted that focusing only on paired males, the coda vocalization would differ from the multi-modulation phrase in having a higher maximum frequency, a longer total duration, and a larger number of frequency modulations in the second note.

2. Materials and methods

2.1. Subjects under study

This study was conducted in Czech, Slovak, and Slovenian zoological parks and involved 14 southern yellow-cheeked gibbons (all adult males) distributed in nine groups (Supplementary Table 2, Figure 1). Four of the 14 subjects were born in the wild (Vietnam) and transported as juveniles to the European zoological parks in the late 1980s. The remaining 10 individuals were born in captivity. Seven of the 14 subjects were unpaired males, and the remaining seven were paired males that were housed with their mate (group 4) or as a family group (an adult pair and offspring). The Jihlava and Olomouc zoological parks each had two groups of gibbons that remained in visual and auditory contact with each other. Group 2 at the Jihlava Zoo comprised one adult female and her two male offspring of different ages; the adult male died in 2009. Group 7 at the Košice Zoo comprised two unpaired males (brothers); both were raised in a family group up to 5.5 years of age and 7.5 years of age, respectively (Birot, 2022). Although some individuals were related, in this study, the sample size was too small ($N = 7$ unpaired males) to include relatedness in the statistical model. Each group had permanent access to an indoor and outdoor enclosure or was confined to an island (group 9). The gibbons were fed four times per day, and their diet consisted of fruits, vegetables, seeds, leaves, cereals, and eggs. Water was available *ad libitum*. Further information about indoor and outdoor enclosure, as well as the overall composition of family groups, is available in Supplementary Tables 3, 4.

2.2. Data collection and acoustic analysis

This research was part of a long-term study focusing on the vocal behavior of captive *Nomascus* gibbons (Hradec et al., 2016, 2017, 2021a,b). Acoustic data were collected on the first and sixth visitations that took place between 2014 and 2021. All of the adult males were observed for 1 to 3 days during each observation. We classified all of the males as either older-paired or younger-unpaired (Supplementary Table 2, Figure 1). In our study, unpaired males are younger (7.7–10.10 years), and paired males are older (11.4–34 years), which indicates that the age of paired and unpaired males did not overlap. Consequently, the social effect cannot be differentiated from the age effect. Older-paired males and their female partner would regularly emit a duet song, typically in the morning (5:00 to 10:00 a.m.), which lasted approximately 10–25 min. Younger-unpaired males emitted only solo songs (i.e., male call) in the morning (5:20–11:00 a.m.), and which lasted 8–30 min.

Younger-unpaired males who lived in family groups emitted solo singing after the parental duets. Where two brothers shared the same enclosure (groups 2 and 6; Supplementary Table 2), these males sang solo songs and produced them independently from each other. All 14 subjects produced a fully matured male vocal pattern, as revealed by the production of staccato notes followed by a multi-modulation phrase exhibiting two or more frequency modulations on the second note (Hradec et al., 2021b). Our dataset derives from 34 recording sessions targeting 7 adult pairs and 17 recording sessions targeting 7 younger-unpaired males, making for a total of 51 recording sessions. From this pool of recordings, we extracted 616 male songs (older-paired: $N = 374$ songs comprising 184 multi-modulation phrases and 190 coda vocalizations; younger-unpaired: $N = 242$ songs). In older-paired subjects, we analyzed 190 codas and 184 male calls, 101 of which were preceded by staccato notes. In younger-unpaired subjects, we analyzed 242 male calls, 29 of which were preceded by staccato notes (Table 1).

We recorded the male vocalizations on an M-Audio Micro Track II recorder or a Marantz PMD661 with a Rode NTG-2 semi-directional microphone (in mono at 16-bit resolution, 44.1 kHz sampling rate) at a distance of 2 to 10 m in an outdoor enclosure or island. All of the recordings were saved as waveform audio files. Acoustic analysis was carried out using Avisoft SASLab Pro version 5.2 software (Avisoft Bioacoustics, Berlin, Germany). Spectrograms were generated under the following settings: FFT length = 1,024; frequency resolution = 12 Hz; temporal resolution = 21.3 ms; overlap = 75%; and window type = Hamming.

To examine the structural differences between male calls (staccato notes and multi-modulated phrases) and coda vocalization as a function of pairing status, we selected the nine acoustic parameters presented in Figure 2. We quantified these features in Avisoft SASLab Pro following the prior annotations of each note and call via visual inspection of spectrograms. Previous studies in the genus *Nomascus* (Schilling, 1984; Konrad and Geissmann, 2006; Thinh et al., 2011) have established that both the male call and the coda are made up of three notes, i.e., Note 1, Note 2, and Last notes (Figure 2). The Last notes consist of one or several notes. If more notes were produced, we did not divide them further; they were treated as one entity. Importantly, the second note of the multi-modulation phrase includes rapid frequency modulations consisting of a steep up-and-down sweep (Konrad and Geissmann, 2006; Hradec et al., 2021b), which is more flexible than the overall pattern of a multi-modulation phrase. Therefore, our analysis of frequency parameters focused on this second note and included only those notes that had two or more modulation frequencies. From this pool of notes, we measured the maximum frequency (kHz), minimum frequency (kHz), and frequency range (kHz), i.e., the difference between the maximum and minimum frequency of the second note. As for temporal parameters, we measured the total duration of the multi-modulation phrase (all three notes included) and the duration of the second note. Finally, the presence or absence of staccato notes in male calls was also quantified due to their considerable individual variation during

TABLE 1 Information on the number of songs (i.e., staccato notes, multi-modulation phrases, and coda vocalizations) in older-paired and younger-unpaired males.

Male	Social status	Recording sessions	Coda vocalization	Multi-modulation phrase	Staccato series	Total number of staccato notes
Male 1	Older-paired	9	60	43	16	95
Male 5	Older-paired	5	2	21	1	1
Male 6	Older-paired	11	71	114	83	259
Male 7	Older-paired	1	6	3	0	0
Male 9	Older-paired	5	44	1	1	3
Male 13	Older-paired	2	4	1	0	0
Male 14	Older-paired	1	3	1	0	0
Total		34	190	184	101	358
Male 2	Younger_unpaired	2	0	17	2	3
Male 3	Younger_unpaired	6	0	52	0	0
Male 4	Younger_unpaired	3	0	43	1	1
Male 8	Younger_unpaired	1	0	35	8	29
Male 10	Younger_unpaired	1	0	7	7	25
Male 11	Younger_unpaired	2	0	26	0	0
Male 12	Younger_unpaired	2	0	62	11	15
Total		17	0	242	29	73
Grand Total		51	190	426	130	431

The total number of songs ($N = 616$) corresponds to the sum of the multi-modulation phrases recorded from unpaired ($N = 242$) and paired ($N = 184$) males and the number of coda vocalizations ($N = 190$) recorded from paired males. These dates are marked in red.

singing. When present, we counted the number of staccato notes at the beginning of a song.

2.3. Statistical analysis

All data were analyzed with the aid of SAS software version 9.4 (SAS Institute Inc., Cary, NC). Results with a p-value of less than 0.05 ($P \leq 0.05$) were considered statistically significant. We applied six statistical Linear Mixed Models fit by REML (restricted maximum likelihood) to test for differences in the vocal structure of the following: (1) male calls in younger-unpaired and older-paired males. Seven features were tested (Figure 2): total duration of multi-modulation phrase (s); duration of the second note (s); maximum frequency (kHz); minimum frequency (kHz); frequency range (kHz); the number of frequency modulations in the second note; and presence of staccato notes; (2) the multi-modulation phrases in younger-unpaired males and coda vocalizations. Six features were tested: total duration of multi-modulation phrase (s); duration of the second note (s); maximum frequency (kHz); minimum frequency (kHz); frequency range (kHz); and the number of frequency modulations in the second note; and (3) the multi-modulation phrases in older-paired males and coda vocalizations. Six features were tested: total duration of multi-modulation phrase (s); duration of the second note (s); maximum frequency (kHz); minimum frequency (kHz); frequency range (kHz); and the number of frequency modulations in the second note. The structure of the multi-modulation phrase in the male calls of both younger-unpaired and older-paired males was similar to that of the coda vocalizations in older-paired males. Because most of our subjects were sampled multiple times, statistical differences in the multi-modulation phrase and coda vocalizations between younger-unpaired and older-paired males were analyzed by a mixed linear model procedure (PROC MIXED) with repeated measurements. The data were analyzed in two steps: In the first step, the effect of social status (older-paired vs. younger-unpaired) was analyzed by the following model:

$$y_{ij} = \ social\ status_i \ + a_j + e_{ij}$$

In the second step, the data set was divided into two subsets based on the determined differences in social status. Additionally, we first carried out a visual data inspection, which indicated a possible different age effect for the acoustical parameters in older-paired and younger-unpaired males. Subsequently, the effect of age on the acoustic parameters was analyzed separately for paired and younger-unpaired males using the following model:

$$y_{ij} = bAge + a_j + e_{ij}$$

In each of these models, y_{ij} is a dependent variable (maximum and minimum frequency, frequency range, number of frequency

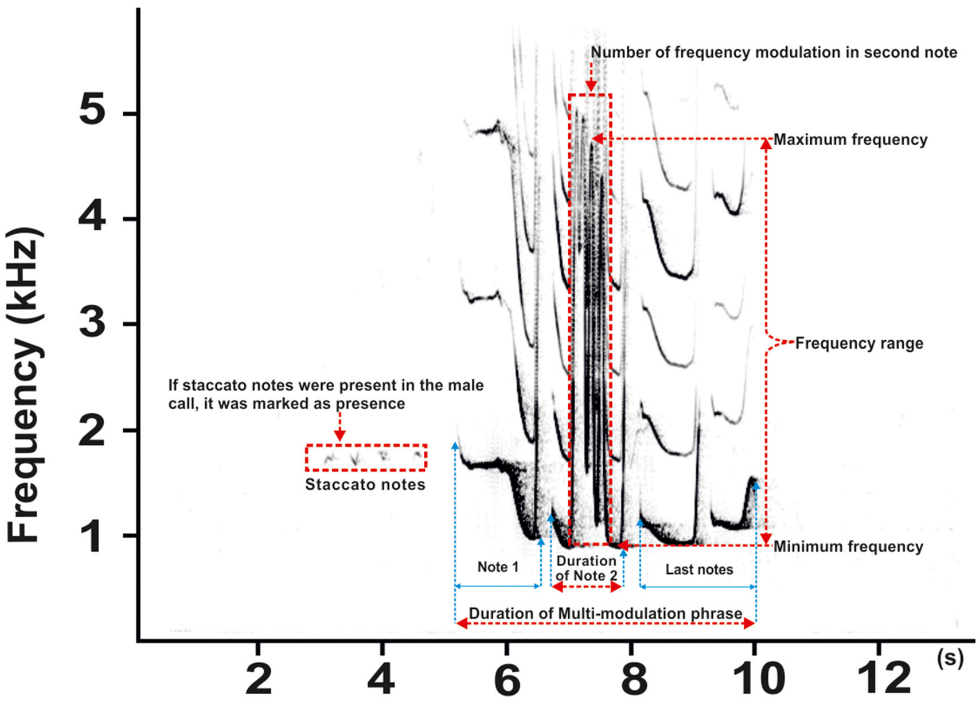


FIGURE 2
Representative spectrogram male call of southern yellow-cheeked gibbons (adapted from Hradec et al., 2017) displaying staccato notes and a multi-modulation phrase (divided into three notes). The most rapid change in frequency modulation is evident in the second note of the steep up-and-down sweeping sound (red dashed rectangle). Red dashed lines indicate the measured parameters of the multi-modulation phrase.

modulations, duration of the multi-modulation phrase, and duration of the second note), the social status is a fixed effect of the i_{th} social status (i = younger-unpaired, older-paired), *bAge* is a regression coefficient for the age of individuals, *a* is a random effect of j_{th} animals, and *e* is a residual error.

We determined statistical differences between younger-unpaired and older-paired males for the multi-modulation phrase of dependent variable "staccato" with a generalized linear mixed model (GLMM) using PROC GLIMMIX with repeated statements for the purpose of repeated measurements. Here, the dependent variable staccato notes (yi) has a value of 1 (presence) with a probability of πi or 0 (absence) and with a probability of 1 − πi for observation i. The logistic model uses a link function $g(p_i)$ linking the expected value to the following non-linear function:

$$y_i = g\left(\pi\right) = \log\left(\frac{\pi_i}{1-\pi_i}\right),$$

where π_i is the probability of "staccato" being recorded

$$\log\left(\frac{\pi_{ijkl}}{1-\pi_{ijkl}}\right) = social\ status_i + bAge + a_j$$

where social status is a fixed effect of the i_{th} (i = older-paired, younger-unpaired), *bAge* presents a regression coefficient for the age of individuals, and *a* is a random effect of j_{th} animals. The goodness of fit of each model (homoscedasticity, normality of errors, and independence) was checked by visually inspecting residuals using the SAS statement "PLOTS = PEARSON PANEL" and testing for normality according to the Kenward–Roger test.

3. Results

3.1. The structure of the male call differs between younger-unpaired and older-paired males

3.1.1. Social status as a fixed effect

Four of the seven acoustic parameters were significantly longer and higher for younger-unpaired males than they were for older-paired males (Table 2). Younger-unpaired males had a longer total duration of the multi-modulation phrase ($P < 0.0001$), a longer duration of the second note ($P < 0.0001$), a higher maximum frequency ($P < 0.0001$), and a wider frequency range ($P < 0.0001$). On the contrary, older-paired males had a higher minimum frequency ($P < 0.0001$). Younger-unpaired males exhibited a lower occurrence of staccato notes (11.98 %; $N = 73$; $P < 0.0001$) than their older-paired counterparts (50.54 %; $N = 358$; $P < 0.0001$). In younger-unpaired males, there were shorter series of staccato notes (range: 0–8 notes) than in older-paired males (range: 0–18 notes). Among the latter, two

TABLE 2 Results based on GLMM showing a comparison between male calls (staccato notes and multi-modulation phrases) by younger-unpaired males and those of older-paired males.

Dependent variables	Social status					Age effect of the male call in younger-unpaired males			Age effect of the male call in older-paired males		
	Male call in younger-unpaired males *(LS mean±SE)*	Male call in older-paired males *(LS mean±SE)*	Estimate ±SE (males call in older-paired males)	F-value	P-value	Estimate ±SE	F-value	P-value	Estimate ±SE	F-value	P-value
Total duration of multi-modulation phrase (s)	6.494 ± 0.063	5.550 ± 0.072	−0.944 ± 0.096	96.70	**<0.0001**	−0.318 ± 0.077	16.83	**<0.0001**	0.036 ± 0.007	25.53	**<0.0001**
Duration of second note (s)	2.287 ± 0.029	1.958 ± 0.034	−0.328 ± 0.045	51.92	**<0.0001**	−0.026 ± 0.029	0.77	0.3799	0.026 ± 0.005	27.74	**<0.0001**
Maximum frequency (kHz)	5,176.980 ± 17.256	5,016.030 ± 19.790	−160.950 ± 26.257	37.52	**<0.0001**	−25.856 ± 19.334	1.79	0.1824	14.840 ± 2.409	37.93	**<0.0001**
Minimum frequency (kHz)	779.750 ± 3.276	813.150 ± 3.757	33.400 ± 4.985	44.88	**<0.0001**	9.086 ± 2.983	9.28	**0.0026**	5.182 ± 0.459	127.32	**<0.0001**
Frequency range (kHz)	4,476.860 ± 69.666	4,118.160 ± 97.584	−358.700 ± 129.930	99.66	**<0.0001**	−34.943 ± 19.698	3.15	0.0774	9.658 ± 2.241	16.05	**<0.0001**
Number of frequency modulations in the second note	3.599 ± 0.044	3.652 ± 0.050	0.053 ± 0.067	0.62	0.4453	0.0796 ± 0.052	2.30	0.1308	0.004 ± 0.006	0.35	0.5531
Presence of staccato notes	0.119 ± 0.020	0.505 ± 0.037	0.022 ± 0.147 (older–paired males) −1.994 ± 0.197 (younger–unpaired males)	66.70	**<0.0001**						

Significant values are highlighted in bold.

subjects (Male 1 and Male 6) mainly contributed to this difference (Table 1).

3.1.2. Age as a regression coefficient

In younger-unpaired males, two of the six acoustic parameters were significantly influenced by age: the total duration of the multi-modulation phrase ($P < 0.0001$) and the minimum frequency ($P = 0.0026$). The total duration of the multi-modulation phrase decreased with age, while the minimum frequency increased with age (Table 2, Figure 3). In older-paired males, five of the six acoustic parameters increased with age: the total duration of the multi-modulation phrase, the duration of the second note, the maximum frequency, the minimum frequency, and the frequency range (all $p < 0.0001$).

3.2. The coda vocalization structure in older-paired males compared to the multi-modulation phrase structure of the male call in younger-unpaired males

3.2.1. Social status as a fixed effect

All six acoustic parameters differed significantly when comparing the multi-modulation phrase of younger-unpaired males with the coda vocalizations of older-paired males (Table 3). These consisted of a longer total duration of the multi-modulation phrase ($P < 0.0001$), a longer duration of the second note ($P = 0.0025$), a higher maximum frequency ($P = 0.0075$), and a wider frequency range ($P < 0.0001$). On the contrary, the coda vocalization of paired males indicated a higher minimum frequency ($P = 0.0006$) and a higher number of frequency modulations in the second note ($P = 0.0002$), than that of the multi-modulation phrase emitted by unpaired males. Both acoustic parameters increased with age (Table 3).

3.2.2. Age as a regression coefficient

For the multi-modulation phrase, two of the six acoustic parameters were significantly influenced by age: the total duration decreased with age ($P < 0.0001$), whereas the minimum frequency increased with age ($P = 0.0026$; Table 3, Figure 4). For the coda vocalization, all six acoustic parameters were significantly influenced by age: the total duration of the multi-modulation phrase ($P = 0.0014$), the duration of the second note ($P < 0.0001$), the maximum frequency ($P < 0.0001$), the minimum frequency ($P = 0.0309$), the frequency range ($P < 0.0001$), and the number of frequency modulations in the second note ($P = 0.0422$). The total duration of the multi-modulation phrase, the duration of the second note, the maximum and minimum frequency, and the frequency range increased with age, while the number of frequency modulations in the second note decreased with age (Table 3, Figure 4).

3.3. Comparing the structure of coda vocalization to that of the multi-modulation phrase in older-paired males

In older-paired males, only one of the six acoustic parameters was significantly higher for the coda vocalization than for the multi-modulation phrase of the male call, namely the number of frequency modulations in the second note ($P < 0.0001$, Table 4). On the contrary, the multi-modulation phrase of the male call of older-paired males had a longer total duration ($P < 0.0001$) and a higher minimum frequency ($P < 0.0084$) than the coda vocalization.

3.3.1. Age as a regression coefficient

Five of the six acoustic parameters significantly increased with age: the total duration of the multi-modulation phrase, the duration of the second note, the maximum frequency, the minimum frequency, and the frequency range (all $P < 0.0001$; Table 4).

4. Discussion

Overall, our analysis of the songs uttered by captive adult males of southern yellow-cheeked gibbons provides three new results. (1) We confirmed our first prediction that younger-unpaired males produce songs that differ in various acoustical features from older-paired males. However, it is unclear which features were influenced by social status or age. (2) We partly confirmed our second prediction that in older-paired males, the coda vocalization structurally differs from the multi-modulation phrase of male calls. (3) Compared with younger-unpaired males, longer series of staccato notes were found to precede the multi-modulation phrase in older-paired males.

4.1. Synergic effects of social status and age on male call and coda vocalization

Our analysis focused on the second note of the multi-modulation phrase of southern yellow-cheeked gibbons, which is the predominant acoustic structure that sets them apart from other species of the genus *Nomascus* (Konrad and Geissmann, 2006). Compared to the male call and coda vocalization of older-paired males, we found that the multi-modulation phrase of the male calls in younger-unpaired males was characterized by a longer total duration, with the second note having a longer duration, a higher maximum frequency, a lower minimum frequency, and a broader frequency range. In contrast, the coda vocalization in older-paired males was characterized by a higher minimum frequency and a larger number of frequency modulations in the second note compared to that same note in the multi-modulation phrase of male calls.

While the male call of unpaired, younger males can be clearly distinguished from the male call and the coda vocalization of paired older males, the effect of social status (paired vs.

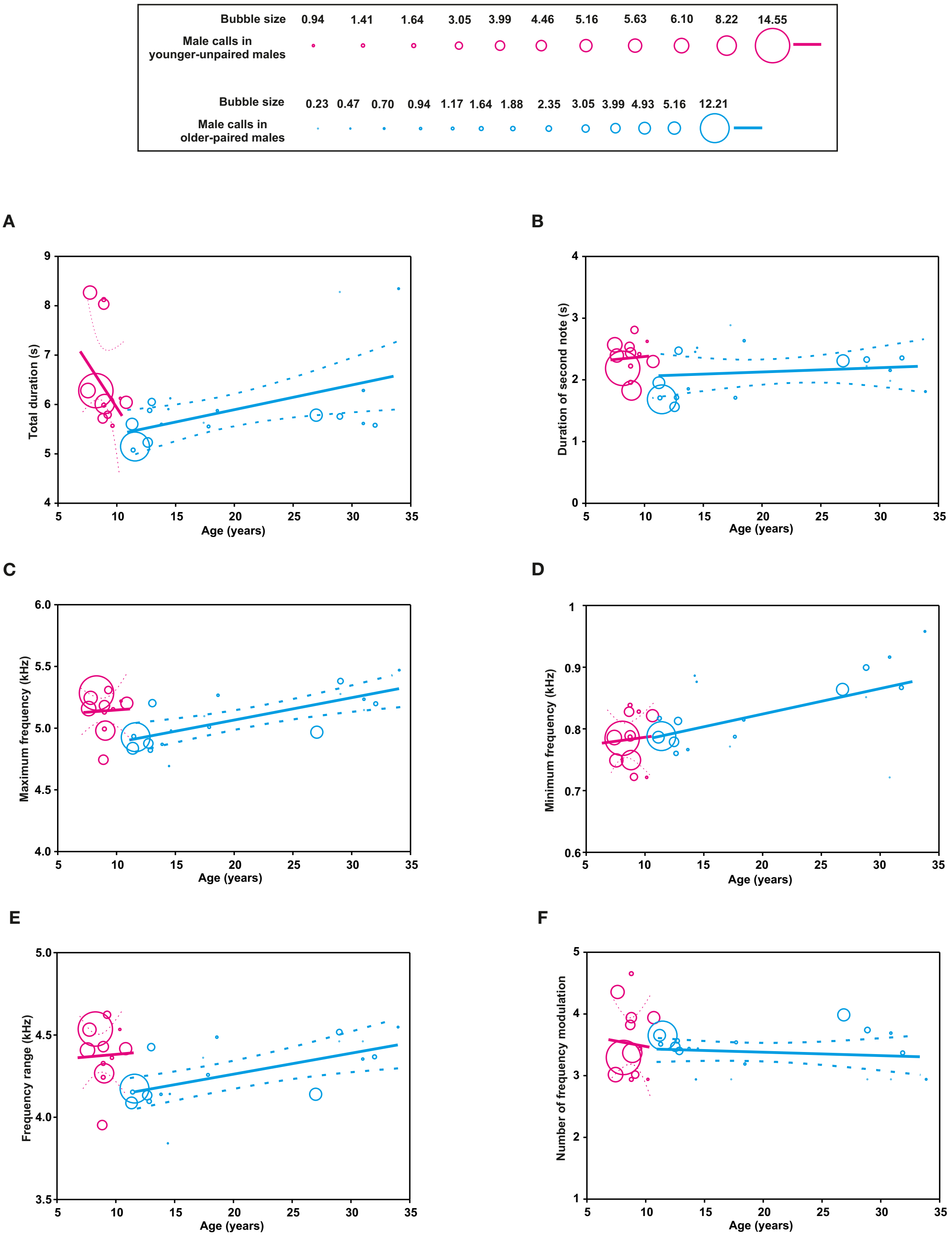


FIGURE 3
Bubble plots showing the influence of age (years) and social status on the acoustic parameters of male calls **(A–F)** in younger-unpaired males (pink bubbles) and older-paired males (blue bubbles). In each panel, the size of the bubbles denotes the proportion of recording samples obtained from each subject at a given age. For each group of subjects, the regression lines are plotted with 95% confidence intervals.

unpaired males) could not be dissociated from the effect of age (younger vs. older males) since both variables were confounded. The breeding of gibbons in European zoological gardens needs to follow the recommendations of the relevant EEP coordinator with respect to the needs of the population as a whole and the specific conditions inherent to each zoo. Due to the long-term surplus of males, the tendency is to extend the stay of a male in his natal group for as long as possible,

TABLE 3 Results based on GLMM showing a comparison between the multi-modulation phrase in younger-unpaired males and the male coda vocalization in older-paired males.

Dependent variables	Social status					Age effect of the multi-modulation phrase in younger-unpaired males			Age effect of the coda vocalization		
	Multi-modulation phrase in younger-unpaired males *(LS mean±SE)*	Coda vocalization in older-paired males *(LS mean±SE)*	Estimate ±SE (Coda vocalization)	F-value	P-value	Estimate ±SE	F-value	P-value	Estimate ±SE	F-value	P-value
Total duration of multi-modulation phrase (s)	6.494 ± 0.061	5.361 ± 0.069	−1.132 ± 0.092	150.40	**<0.0001**	−0.318 ± 0.077	16.83	**<0.0001**	0.018 ± 0.005	10.53	**0.0014**
Duration of second note (s)	2.287 ± 0.027	2.129 ± 0.030	−0.158 ± 0.041	14.56	**0.0025**	−0.026 ± 0.029	0.77	0.3799	0.029 ± 0.003	88.65	**<0.0001**
Maximum frequency (kHz)	5,176.980 ± 17.937	5,090.210 ± 20.243	−86.772 ± 27.047	10.29	**0.0075**	−25.856 ± 19.334	1.79	0.1824	11.722 ± 2.336	25.17	**<0.0001**
Minimum frequency (kHz)	779.750 ± 4.366	809.890 ± 4.928	30.142 ± 6.584	20.96	**0.0006**	9.086 ± 2.983	9.28	**0.0026**	1.659 ± 0.763	4.73	**0.0309**
Frequency range (kHz)	4,480.160 ± 72.316	4,188.240 ± 90.369	−291.920 ± 133.530	300.86	**<0.0001**	−34.943 ± 19.698	3.15	0.0774	10.063 ± 2.249	20.01	**<0.0001**
Number of frequency modulations in the second note	3.599 ± 0.044	3.952 ± 0.050	0.353 ± 0.067	27.40	**0.0002**	0.079 ± 0.052	2.30	0.1308	−0.011 ± 0.005	4.19	**0.0422**

Significant values are highlighted in bold.

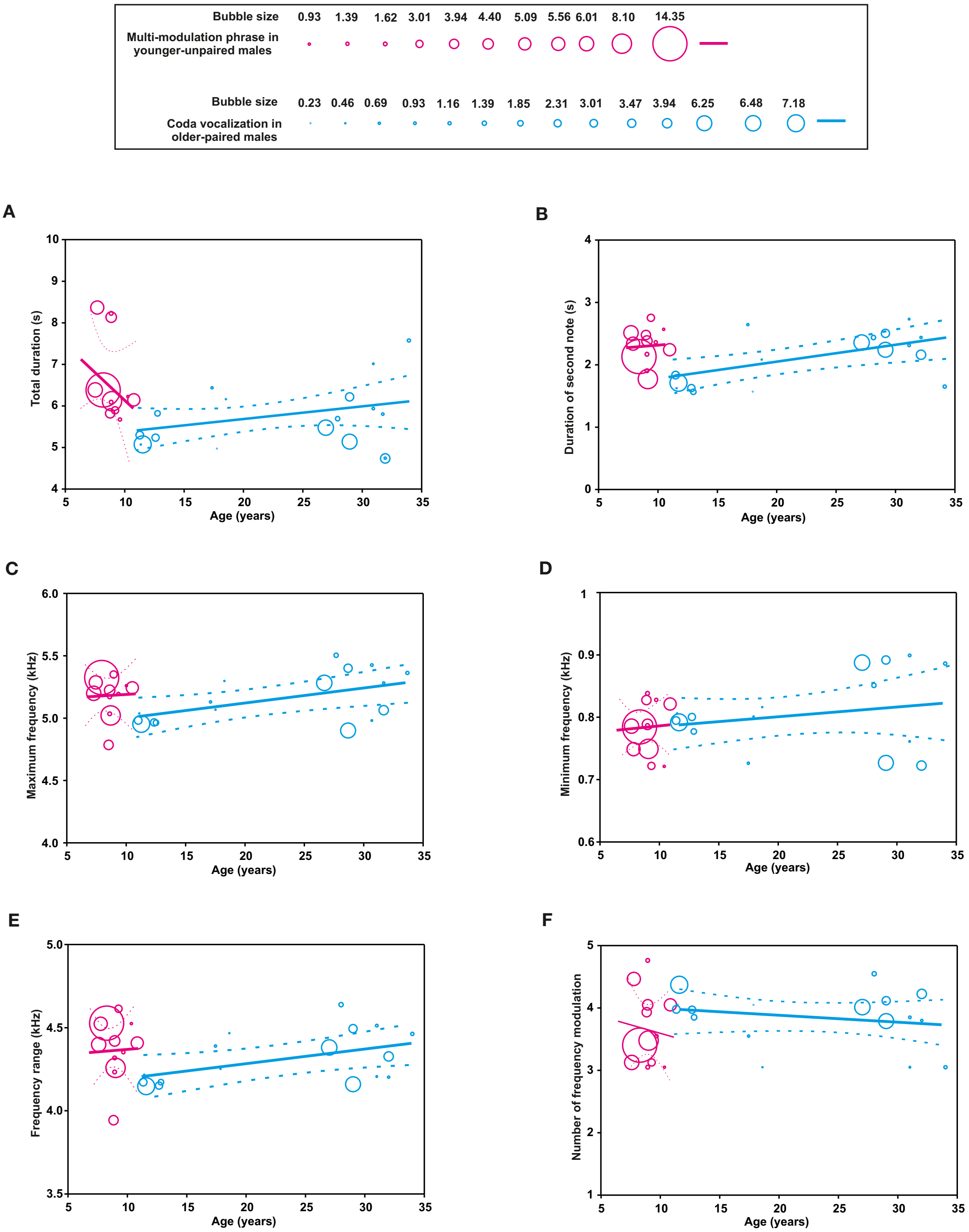


FIGURE 4
Bubble plots showing the influence of age (years) and social status on the acoustic parameters of male calls **(A–F)** in multi-modulation phrase of younger-unpaired males (pink bubbles) and coda vocalization of older-paired males (blue bubbles). In each panel, the size of the bubbles denotes the proportion of recording samples obtained from each subject at a given age. For each group of subjects, the regression lines are plotted with 95% confidence intervals.

ideally up to 10 years of age. The possibility of obtaining vocal samples from both paired and unpaired males that overlap in age distribution is thus limited by the number of animals kept and the EEP breeding recommendations.

Nonetheless, the effect of social status on the song structure might be supported by the fact that male gibbons below 8 years of age have been known to mate and produce offspring (Geissmann, 1991). It is, therefore, possible that younger-unpaired males might

TABLE 4 Results based on GLMM showing a comparison between the multi-modulation phrase and the male coda vocalization in older-paired males.

Dependent variables	Acoustic structure			Age		
	Estimate ±SE (Coda vocalization)	F-value	P-value	Estimate ±SE	F-value	P-value
Total duration of multi-modulation phrase (s)	−0.339 ± 0.092	18.75	**<0.0001**	0.019 ± 0.004	20.06	**<0.0001**
Duration of second note (s)	0.009 ± 0.048	0.04	0.8494	0.028 ± 0.002	113.64	**<0.0001**
Maximum frequency (kHz)	−0.303 ± 28.635	0.00	0.9915	13.049 ± 1.577	68.44	**<0.0001**
Minimum frequency (kHz)	21.315 ± 8.041	7.03	**0.0084**	2.748 ± 0.447	37.79	**<0.0001**
Frequency range (kHz)	21.012 ± 28.010	0.56	0.4536	10.300 ± 1.544	44.50	**<0.0001**
Number of frequency modulations in the second note	0.328 ± 0.070	21.88	**<0.0001**	0.001 ± 0.003	0.15	0.6987

attract females by emitting more potent songs compared with older-paired males.

As mentioned above, the multi-modulation phrase of the male call by younger-unpaired males had fewer extremely rapid frequency modulations in the second note than the coda vocalization in older-paired males. Furthermore, this number of frequency modulations did not differ from that found in the male calls of older-paired males. It might be that younger-unpaired males invest more energy in other acoustic features, which might provide more information to receivers about the singer's attributes than the higher number of frequency modulations in the second note.

The multi-modulation phrase of the male call exhibited minimum frequency values that increased with age, irrespective of the pairing status. Our results are in disagreement with previous findings in adults of non-human primates, showing that the minimum frequency decreases (e.g., Inoue, 1988) or remains similar (Terleph et al., 2016) with increasing age. The proximate mechanisms responsible for the age-dependent increase in the minimum frequency of vocal output that we report are not yet known.

In both older-paired and younger-unpaired males, a series of staccato notes sometimes precede the multi-modulation phrase of the male call. We found that staccato notes had a lower probability of occurrence in younger-unpaired males and exhibited shorter series of notes compared with older-paired males. This is in agreement with Huang et al. (2020), who also observed fewer staccato notes in unpaired Cao-vit gibbons (*Nomascus nasutus*) and black-crested gibbons (*Nomascus concolor*) males than in their paired counterparts. The function of staccato notes is not yet fully known. Given their very subtle structures, low volume, and high numbers in older-paired males, we hypothesize that staccato notes may, for example, play a role in song coordination between males and females during duets.

We found no support for our first prediction in two hylobatid studies, one focusing on social status (Mitani, 1988) and the other one on the effect of age (Barelli et al., 2013). Mitani (1988) compared the vocal structure of unpaired and paired males in agile gibbons. In that report, two of the four unpaired adult males were still resident within their family group and the two others were floaters of unknown age. The study did not reveal any obvious acoustical difference in the solo songs of those males compared to paired males. One possible explanation for the difference in results between our study and that of Mitani (1988) is that paired *Nomascus* gibbon males usually produce the male song only in combination with the female great call, i.e., as part of a non-overlapping duet (Haimoff, 1984a; Geissmann, 2002), whereas paired agile gibbon males produce male songs either as solos or as a contribution to the duet song. It is possible that paired agile gibbon males produce solo songs in different situations, and the vocal structure of their solo songs might differ from that of their duets.

However, our results are in agreement with previous findings, showing that in birds, unpaired males sing higher frequency songs compared to paired males (Staicer, 1996; Martínez and Zuberogoitia, 2002; Staicer et al., 2006; Sung and Handford, 2019). Our findings are also consistent with those of Hennin et al. (2009) in the rufous-and-white wrens (*Thryothorus rufalbus*), in which the trills of bachelor (unpaired) males displayed a higher maximum frequency, broader bandwidths (i.e., frequency range), and longer call durations than paired males. In this study, three solitary males were recorded both before and after pairing with a female, while 13 others were recorded both before and after losing a mate. In both groups, the bachelors produced songs that were more potent than those of paired males. These results suggest that songs of rufous-and-white wrens encode information about their paired status (Hennin et al., 2009).

An effect of age might be another possible explanation for the more salient songs of younger-unpaired males. It has been shown that the development of the male pattern in the southern yellow-cheeked gibbon is completed by the age of 7.1 years (Hradec et al., 2021b). Androgen levels probably contribute to larynx growth (changes in laryngeal muscle) during the development of the male call; androgen receptors are located on the laryngeal cartilage (Newman et al., 2000), which has a permanent impact on the length or tension of the vocal folds (Fitch and Hauser, 1995). Higher levels of androgens in gibbons may persist even after the development of the adult male pattern at the age of 7 years (Barelli et al., 2013), thus further impacting the components of the vocal apparatus—larynx morphology and tension of the vocal folds. This could explain our finding of the wider frequency range and longer duration in the calls of younger-unpaired males. Subsequently, androgen levels in older males may decline and stabilize at a lower level (Barelli et al., 2013).

To our knowledge, there is no other study that focuses on the effect of age on the song structure in unpaired male gibbons. At first glance, visual inspection of the data for younger-unpaired males showed a decline of several acoustical parameters, but subsequent analysis revealed a significant decrease only for the total duration of the multi-modulation phase, not for the other parameters. A study on white-handed gibbons examined the effect of age in paired males but not in unpaired males (Barelli et al., 2013). In that study, two age categories were compared (adult males: 8–25 years of age; senior males: over 25 years of age), revealing an age-related decline in acoustic features of male calls. That study states that senior males emit calls having a lower pitch and a shorter duration than their (younger) adult counterparts. Our results showed the opposite effect for all analyzed acoustical parameters except for the number of frequency modulations in the second note (i.e., there was no age effect). It is difficult to interpret the age effect from our study when compared to that of Barelli et al. (2013) in which only the solo songs of paired males were investigated but not their duet songs. In addition, the white-handed gibbon belongs to the genus *Hylobates*, which differs from the genus *Nomascus* by the absence of frequency modulations (Geissmann, 2002).

It is known that female white-handed gibbons exhibit an age-related decline in the acoustic parameters of the great call, suggesting that call features correlate with physical conditions (Terleph et al., 2016). However, females reproduce up to the age of 40 years in the wild (Reichard et al., 2012). To the best of our knowledge, such data are not available for males, but it can be assumed that 40-year-old males are still fertile and may even reproduce beyond that age. It is not clear how the song features change during senescence in gibbons. It has been shown that acoustical parameters of calls in male baboons (*Papio cynocephalus ursinus*) change with social rank and increasing age (Fischer et al., 2004). Higher-ranking males had a higher call rate and produced longer bouts (Fischer et al., 2002; Kitchen et al., 2003). However, the social system of these primates (multilevel society) differs from the mostly monogamous gibbon species.

Our results are to some extent supported by a study by Wich et al. (2003) in Thomas langurs (*Presbytis thomasi*), which showed that a change in the social environment (unpaired vs. paired males), rather than an age effect, is related to the changes in acoustic parameters of the male call. The male loud call consists of harsh non-tonal elements (N-units) and tonal elements (T-units), but their meaning is not yet known. That study compared males of the same age (about 7 years) and found that in paired males, not in unpaired males, the total duration and number of N-units increased, while the number of T-units decreased. These findings compare to our results despite the differing age compositions of our two groups of subjects (older-paired and younger-unpaired males).

4.2. Structural differences between the coda vocalization and the multi-modulation phrase of the male call in older-paired males

Our study revealed that, in older-paired males, the coda vocalization differed from the male call by a larger number of frequency modulations in the second note. Conversely, the multi-modulation phrase had a longer total duration than the coda vocalization. To our knowledge, this study is the first of its kind that compares the structure of the multi-modulation phrase of the male call to the coda vocalization in older-paired males.

The higher number of frequency modulations found in the second note of the coda vocalization is in agreement with the results obtained from the visual inspection of the coda vocalization in Hainan gibbons (Haimoff, 1984b). It is possible that the emission of trill-like vocalizations, both in the coda and the multi-modulation phrase, requires a significant amount of energy and might be perceived by other males as a reliable signal of the sender's physical quality, as has been suggested in some birds and mammals (Podos, 1997; Illes et al., 2006; Pasch et al., 2011). In contrast, the shorter duration of the coda vocalization relative to the multi-modulation phrase of the male call is inconsistent with the findings in Hainan gibbons (Haimoff, 1984b).

We also found that the acoustical parameters for coda vocalizations and multi-modulation phrases in older-paired males increased significantly with age, with only one exception, i.e., the number of frequency modulations in the second note. It is possible that after the completion of male vocal development, this acoustic feature remains stable in all types of calls (i.e., multi-modulation phrase and coda vocalization) produced in adult males, as found in the current study. This indicates that the number of frequency modulations in the second note may largely be genetically determined and might potentially be included in future studies of the vocal individuality of *Nomascus* gibbons.

The purpose of the coda vocalization and the male call is still not fully understood. Several hypotheses have been proposed for the male song, namely territorial advertisement, resource/mate defense, strengthening (cohesion) pair and/or family bonds (social cohesion), and mate attraction (Cowlishaw, 1992; Geissmann, 2002). It is possible that these two calls do not share the same function due to the difference in spectro-temporal parameters (e.g., frequency range and duration of the second note). It is generally known that the coda vocalization is always emitted immediately following a female great call (Konrad and Geissmann, 2006), which suggests that it functions as a means to strengthen pair and/or family bonds. It is also possible that a higher number of frequency modulations in the second note of the coda vocalization may be an indicator of mate guarding in male gibbons. This is supported by the fact that the male emits the coda vocalization immediately after the female great call, perhaps as a sign of his paired status. The greater ability of older-paired males to rapidly modulate the vocal tract may deter other males from approaching the female. In future studies, it would be useful to determine whether the structure of the multi-modulation phrase in unpaired or paired males is more attractive to females in playback settings.

5. Conclusion

Our study is the first of its kind to provide evidence of structural differences in the male calls and coda vocalizations of southern yellow-cheeked gibbons. Calls uttered by younger-unpaired males differed from both the male calls and coda vocalizations of older-paired males by a higher maximum frequency, wider frequency range, and longer duration of the multi-modulation phrase. In contrast, the male calls of older-paired subjects were preceded by a relatively larger number of staccato notes, and a salient feature of the coda vocalization in those males was the emission of a higher number of frequency modulations in the second note together with a lower minimum frequency. Our findings demonstrate that the male call of younger-unpaired males can be clearly distinguished from the male call and the coda vocalization of older-paired males, but further studies are needed to disentangle the effects of social status and age.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The research conducted herein was approved by the Ethics and Animal Care Committee at the Czech University of Life Sciences, Prague (reference number: CZU/1606) and was performed in accordance with relevant guidelines. This study was fully non-invasive and approved by the management of zoos. All zoological institutions employ rigorous standards for animal welfare and are accredited by the European Association of Zoos and Aquaria (EAZA) and and the Union of Czech and Slovak Zoos (UCSZOO). This study fully complied with the legal requirements of the Czech Republic, the Slovak Republic, and the Republic of Slovenia, and those provided by the European Directive 2010/63/EU. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

MH: conceptualizing the experiment, conducting and collecting all data, writing and editing the manuscript, and constructing the figures and tables. GI: conceptualizing the experiment and writing and editing the manuscript. MP: writing and editing the manuscript and constructing the figures. PB: editing the manuscript. HV-V: statistical analyses and editing the manuscript. All authors contributed to the article and approved the submitted version.

Funding

GI was supported by the Ministry of Agriculture of the Czech Republic (MZE-RO0723).

Acknowledgments

The authors would like to thank Helena Chaloupková and Richard Policht for their helpful comments on an earlier draft of the study. Furthermore, we wish to thank all the curators and keepers from the Czech, Slovak, and Slovenian zoos involved in this study for their support and gracious permission for data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2023.956922/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Distribution of data collection in younger-unpaired and paired males (number and age of subjects).

SUPPLEMENTARY TABLE 1
Acoustic terms and definitions for gibbon song.

SUPPLEMENTARY TABLE 2
Overview of subject males.

SUPPLEMENTARY TABLE 3
Information on the dimensions of the indoor and outdoor enclosures.

SUPPLEMENTARY TABLE 4
Overall composition of family groups.

References

Ballentine, B. (2009). The ability to perform physically challenging songs predicts age and size in male swamp sparrows (*Melospiza georgiana*). *Anim. Behav.* 77, 973–978. doi: 10.1016/j.anbehav.2008.12.027

Banerjee, A., Phelps, S. M., and Long, M. A. (2019). Singing mice. *Curr. Biol.* 29, 190–191. doi: 10.1016/j.cub.2018.11.048

Barelli, C., Mundry, R., Heistermann, M., and Hammerschmidt, K. (2013). Cues to androgens and quality in male gibbon songs. *PLoS ONE.* 8, e82748. doi: 10.1371/journal.pone.0082748

Birot, H. (2022). *16th European Studbook Southern Yellow-Cheeked Gibbon (Nomascus gabriellae)*. Mulhouse: Mulhouse Zoo.

Briefer, E., Vannoni, E., and McElligott, A. G. (2010). Quality prevails over identity in the sexually selected vocalisations of an ageing mammal. *BMC Biol.* 8, 35. doi: 10.1186/1741-7007-8-35

Brockelman, W. Y., and Gittins, S. P. (1984). "Natural hybridization in the *Hylobates lar* species group: Implications for speciation in gibbons," in *The Lesser Apes: Evolutionary and Behavioural Biology,* Preuschoft, H., Chivers, D.J., Brockelman, W.Y., and Creel, N. (eds). (Edinburgh University: Edinburgh Press).

Brockelman, W. Y., and Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature.* 312, 634–636. doi: 10.1038/312634a0

Brockelman, W. Y., and Srikosamatara, S. (1984). "Maintenance and evolution of social structure in gibbons," in *The Lesser Apes: Evolutionary and Behavioural Biology*, Preuschoft, H., Chivers, D.J., Brockelman, W.Y., and Creel, N. (eds). (Edinburgh University: Edinburgh Press).

Caselli, C. B., Mennill, D. J., Gestich, C. C., Setz, E. Z. F., and Bicca-Marques, J. C. (2015). Playback responses of socially monogamous black-fronted titi monkeys to simulated solitary and paired intruders. *Am. J. Primatol.* 77, 1135–1142. doi: 10.1002/ajp.22447

Charlton, B. D., Zhihe, Z., and Snyder, R. J. (2009). The information content of giant panda, Ailuropoda melanoleuca, bleats: Acoustic cues to sex, age and size. *Anim. Behav.* 78, e898. doi: 10.1016/j.anbehav.2009.06.029

Clarke, E., Reichard, U. H., and Zuberbuehler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE.* 1, e73. doi: 10.1371/journal.pone.0000073

Clink, D. J., Carif, R. A., Crofoot, M. C., and Marshall, A. J. (2018). Evidence of vocal performance constraints in a female non-human primate. *Anim. Behav.* 141, 85–94. doi: 10.1016/j.anbehav.2018.05.002

Clink, D. J., Lau, A. R., and Bales, K. L. (2019). Age-related changes and vocal convergence in titi monkey duet pulses. *Behaviour.* 56, 1471–1494. doi: 10.1163/1568539X-00003575

Clink, D. J., Tasirin, J. S., and Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Curr. Zool.* 66, 173–186. doi: 10.1093/cz/zoz035

Clink, D. J., Zafar, M., Ahmad, A. H., and Lau, A. R. (2021). Limited evidence for individual signatures or site-level patterns of variation in male northern gray gibbon (*Hylobates funereus*) duet codas. *Int. J. Primatol.* 42, 896–914. doi: 10.1007/s10764-021-00250-2

Cowlishaw, G. (1992). Song function in gibbons. *Behaviour.* 121, 131–153. doi: 10.1163/156853992X00471

Cramer, E. R. A. (2013). Physically challenging song traits, male quality, and reproductive success in house wrens. *PLoS ONE.* 8, e59208. doi: 10.1371/journal.pone.0059208

De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., et al. (2021). Born to sing! Song development in a singing primate. *Curr. Zool.* 67, 597–608. doi: 10.1093/cz/zoab018

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., et al. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol. Ecol. Evol.* 34, 205–219. doi: 10.1080/03949370.2021.2015451

Demars, C., and Goustard, M. (1978). Le grand chant d'Hylobates concolor leucogenys. Comparison avec les émissions sonores homologues d'H. c. gabriellae et d'H. klossii (iles Mentawai, ouest Sumatra). *Behaviour.* 65, 1–26. doi: 10.1163/156853978X00503

Erb, W. M., Hodges, J. K., and Hammerschmidt, K. (2013). Individual, Contextual, and Age-Related Acoustic Variation in Simakobu (*Simias concolor*) Loud Calls. *Plos One*. 8:e83131. doi: 10.1371/journal.pone.0083131

Ey, E., Pfefferle, D., and Fischer, J. (2007). Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates.* 48, 253–267. doi: 10.1007/s10329-006-0033-y

Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (2002). Acoustic features of male baboon loud calls: influences of context, age, and individuality. *J. Acoust. Soc. Am.* 111:1465–1474. doi: 10.1121/1.1433807

Fischer, J., Kitchen, D. M., Seyfarth, R. M., and Cheney, D. L. (2004). Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behav. Ecol. Sociobiol.* 56, 140–148. doi: 10.1007/s00265-003-0739-4

Fitch, W. T., and Hauser, M. D. (1995). Vocal production in non-human primates—acoustics, physiology, and functional constraints on honest advertisement. *Am. J. Primatol.* 37, 191–219 doi: 10.1002/ajp.1350370303

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., et al. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Fron. Neurosci.* 10, 249. doi: 10.3389/fnins.2016.00249

Geissmann, T. (1984). Inheritance of song parameters in the gibbon song, analyzed in 2 hybrid gibbons (*Hylobates pileatus* x *H. lar*). *Folia Primatol.* 42, 216–235. doi: 10.1159/000156165

Geissmann, T. (1991). Reassessment of age of sexual maturity in gibbons (*Hylobates* spp.). *Am. J. Primatol.* 23, 11–22. doi: 10.1002/ajp.1350230103

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/S1464793101005826

Geissmann, T., and Nijman, V. (2006). Calling in wild silvery gibbons (*Hylobates moloch*) in Java (Indonesia): Behavior, phylogeny, and conservation. *Am. J. Primatol.* 68, 1–19. doi: 10.1002/ajp.20203

Gil, D., and Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends Ecol. Evol.* 17, 133–141. doi: 10.1016/S0169-5347(02)02410-2

Haimoff, E. H. (1984a). "Acoustic and organizational features of gibbon song", in *The Lesser Apes: Evolutionary and Behavioural Biology,* eds H. Preuschoft, W. Y. Chivers, W. Y. Brockelman, and N. Creel (Edinburgh University: Edinburgh Press).

Haimoff, E. H. (1984b). The organization of song in the Hainan black gibbon (*Hylobates concolor hainanus*). *Primates.* 25, 225–235. doi: 10.1007/BF02382394

Hauser, M. D. (1989). Ontogenic changes in the comprehension and production of vervet monkey (*Cercopithecus aethiops*) vocalizations. *J. Comp. Psychol.* 103, 149–158. doi: 10.1037/0735-7036.103.2.149

Hennin, H. L., Barker, N. K. S., Bradley, D. W., and Mennill, D. J. (2009). Bachelor and paired male rufous-and-white wrens use different singing strategies. *Behav. Ecol. Sociobiol.* 64, 151–159. doi: 10.1007/s00265-009-0832-4

Hradec, M., Bolechova, P., and Svobodová, I. (2016). Production of a female-specific great call in an immature male gibbon, the Nomascus genus. *Primates* 57, 445–448 doi: 10.1007/s10329-016-0569-4

Hradec, M., Illmann, G., Bartos, L., and Bolechova, P. (2021b). The transition from the female-like great calls to male calls during ontogeny in southern yellow-cheeked gibbon males (*Nomascus gabriellae*). *Sci. Rep.* 11, 22040. doi: 10.1038/s41598-021-01648-x

Hradec, M., Illmann, G., and Bolechova, P. (2021a). A first report of separation calls in southern yellow-cheeked gibbons (*Nomascus gabriellae*) in captivity. *Primates* 62, 5–10. doi: 10.1007/s10329-020-00870-x

Hradec, M., Linhart, P., Bartos, L., and Bolechova, P. (2017). The traits of the great calls in the juvenile and adolescent gibbon males *Nomascus gabriellae*. *PLoS ONE.* 12, e0173959. doi: 10.1371/journal.pone.0173959

Huang, M. P., Ma, H. G., Ma, C. Y., Garber, P. A., and Fan, P. F. (2020). Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: insights into the origin of human language. *Anim. Behav.* 160, 145–155. doi: 10.1016/j.anbehav.2019.11.017

Illes, A. E., Hall, M. L., and Vehrencamp, S. L. (2006). Vocal performance influences male receiver response in the banded wren. *Proc. R. Soc. B:.Biol. Sci.* 273, 1907–1912. doi: 10.1098/rspb.2006.3535

Inoue, M. (1988). Age gradations in vocalization and body weight in Japanese monkeys (*Macaca fuscat*). *Folia Primatol.* 51:76–86. doi: 10.1159/000156359

Janik, V. M., and Knornschild, M. (2021). Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 376, 20200244. doi: 10.1098/rstb.2020.0244

Kitchen, D. M., Seyfarth, R. M., Fischer, J., and Cheney, D. L. (2003). Loud calls as indicators of dominance in male baboons (*Papio cynocephalus ursinus*). *Behav. Ecol. Sociobiol.* 53, 374–384. doi: 10.1007/s00265-003-0588-1

Konrad, R., and Geissmann, T. (2006). Vocal diversity and taxonomy of Nomascus in Cambodia. *Int. J. Primatol.* 27, 713–745. doi: 10.1007/s10764-006-9042-3

Lau, A. R., Clink, D. J., and Bales, K. L. (2020). Individuality in the vocalizations of infant and adult coppery titi monkeys (*Plecturocebus cupreus*). *Am. J. Primatol.* 82:e23134. doi: 10.1002/ajp.23134

Lau, A. R., Clink, D. J., Crofoot, M. C., and Marshall, A. J. (2018). Evidence for High Variability in Temporal Features of the Male Coda in Muller's Bornean Gibbons (*Hylobates muelleri*). *Int. J. Primatol.* 39, 670–684. doi: 10.1007/s10764-018-0061-7

Martínez, J. A., and Zuberogoitia, I. (2002). Factors affecting the vocal behaviour of eagle owls Bubo bubo: effects of season, density and territory quality. *Ardeola*. 50, 255—258

Mitani, J. C. (1987). Territoriality and monogamy among agile gibbons (*Hylobates agilis*). *Behav. Ecol. Sociobiol.* 20, 265–269. doi: 10.1007/BF00292179

Mitani, J. C. (1988). Male gibbon (*Hylobates agilis*) singing behavior - natural history, song 548 variations and function. *Ethology.* 79, 177–194. doi: 10.1111/j.1439-0310.1988.tb00710.x

Nemeth, E., Kempenaers, B., Matessi, G., and Brumm, H. (2012). Rock sparrow song reflects male age and reproductive success. *PLoS ONE.* 7:e43259. doi: 10.1371/journal.pone.0043259

Newman, S. R., Butler, J., Hammond, E. H., and Gray, S. D. (2000). Preliminary report on hormone receptors in the human vocal fold. *J. Voice* 14, 72–81. doi: 10.1016/S0892-1997(00)80096-X

Owren, M. J., and Rendall, D. (2003). Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human (*Homo sapiens*) listeners. *J. Comp. Psychol.* 117, 380–390. doi: 10.1037/0735-7036.117.4.380

Pasch, B., George, A. S., Campbell, P., and Phelps, S. M. (2011). Androgen-dependent male vocal performance influences female preference in neotropical singing mice. *Anim. Behav.* 82, 177–183. doi: 10.1016/j.anbehav.2011.04.018

Pfefferle, D., and Fischer, J. (2006). Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Anim. Behav.* 72, 43–51. doi: 10.1016/j.anbehav.2005.08.021

Pistorio, A. L., Vintch, B., and Wang, X. (2006). Acoustic analyses of vocal development in a new world primate, the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* 120, 1655–1670. doi: 10.1121/1.2225899

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes:Emberizidae). *Evolution.* 51, 537–551. doi: 10.2307/2411126

Price, T., Arnold, K., Zuberbuhler, K., and Semple, S. (2009). Pyow but not hack calls of the male putty-nosed monkey (*Cercopithcus nictitans*) convey information about caller identity. *Behaviour*. 146, 871–888. doi: 10.1163/156853908X396610

Raemaekers, J. J., and Raemaekers, P. M. (1985). Field playback of loud calls to gibbons (*Hylobates lar*) - Territorial, sex-specific and species-specific responses. *Anim. Behav.* 33, 481–493. doi: 10.1016/S0003-3472(85)80071-3

Reby, D., and McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Anim. Behav.* 65, 519–530. doi: 10.1006/anbe.2003.2078

Reichard, U. H., Barelli, C., Hirai, H., and Nowak, M. G. (2016). "Evolution of gibbons and siamang phylogeny, morphology, and cognition", in *The Evolution of Gibbons and Siamang,* Reichard, U. H., Hirai, H., and Barelli, C. (Springer, New York, NY). doi: 10.1007/978-1-4939-5614-2

Reichard, U. H., Ganpanakngan, M., and Barelli, C. (2012). "White-handed gibbons of Khao Yai: social flexibility, reproductive strategies, and a slow life history," in *Long-term Field Studies of Primates*, Kappeler, P. M., and Watts, D. (eds). (Springer, Berlin). p. 237–258. doi: 10.1007/978-3-642-22514-7_11

Sanvito, S., Galimberti, F., and Miller, E. H. (2008). Development of aggressive vocalizations in male southern elephant seals (*Mirounga leonina*): maturation or learning? *Behaviour*. 145: 137–170. doi: 10.1163/156853907783244729

Schilling, D. (1984). "Song bouts and duetting in the concolor gibbon," in *The Lesser Apes: Evolutionary and Behavioural Biology,* Preuschoft, H., Chivers, D. J., Brockelman, W. Y., and Creel, N. (eds). (Edinburgh University: Edinburgh Press).

Staicer, C. A. (1996). Honest advertisement of pairing status: evidence from a tropical resident woodwarbler. *Anim. Behav*. 51, 375–390. doi: 10.1006/anbe.1996.0036

Staicer, C. A., Ingalls, V., and Sherry, T. A. (2006). Singing behavior varies with breeding status of American redstarts (*Setophaga ruticilla*). *Wilson J. Ornithol.* 118, 439–451. doi: 10.1676/05-056.1

Sun, G. Z., Huang, B., Guan, Z. H., Geissmann, T., and Jiang, X. L. (2011). Individuality in male songs of wild black crested gibbons (*Nomascus concolor*). *Am. J.Primatol.* 73, 431–438. doi: 10.1002/ajp.20917

Sung, H., and Handford, P. (2019). Seasonal and diurnal variation of singing activity: their possible relationship with intra- and inter-sexual function in the savannah sparrow (*Passerculus sandwichensis*). *Zool. Sci.* 36, 173–181. doi: 10.2108/zs180139

Tenaza, R. R. (1976). Song, choruses and countersinging of Klosi gibbons (*Hylobates klossii*) in Siberut island, Indonesia. *Z. Tierpsychol.-J.Compar. Ethol.* 40, 37–52. doi: 10.1111/j.1439-0310.1976.tb00924.x

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2016). Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *Bmc Evol. Biol.* 16,4 doi: 10.1186/s12862-015-0578-8

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018a). Male white-handed gibbons flexibly time duet contributions. *Behav. Ecol. Sociobiol.* 72:16. doi: 10.1007/s00265-017-2432-z

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2018b). An analysis of white-handed gibbon male song reveals speech-like phrases. *Am. J. Phys. Anthropol.* 166, 649–660. doi: 10.1002/ajpa.23451

Thinh, V. N., Hallam, C., Roos, C., and Hammerschmidt, K. (2011). Concordance between vocal and genetic diversity in crested gibbons. *Bmc Evol. Biol.* 11,36 doi: 10.1186/1471-2148-11-36

Thorpe, W. H. (1961). "Bird-Song. The Biology of Vocal Communication and Expression in Birds," in *Cambridge Monographs in Experimental Biology No. 12.* New York: Cambridge University Press.

Tilson, R. L. (1981). Family formation strategies of Kloss's gibbons. *Folis primatol.* 35, 259–297. doi: 10.1159/000155979

Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Randrianarison, R. M., Friard, O., et al. (2017). An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. *Sci. Rep.* 7:10140. doi: 10.1038/s41598-017-10656-9

Torti, V., Gamba, M., Rabemananjara, Z. H., and Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Ital. J.Zool.* 80, 596–607. doi: 10.1080/11250003.2013.845261

Umeed, R., Attademo, F. L. N., and Bezerra, B. (2018). The influence of age and sex on the vocal repertoire of the antillean manatee (*Trichechus manatus manatus*) and their responses to call playback. *Mar. Mammal. Sci*. 34, 577–594. doi: 10.1111/mms.12467

Vanderhoff, E. N., and Hoverud, N. B. (2022). Perspectives on antiphonal calling, duetting and counter-singing in non-primate mammals: an overview with notes on the coordinated vocalizations of bamboo rats (Dactylomys spp., Rodentia: Echimyidae). *Front. Ecol. Evol.* 10:906546. doi: 10.3389/fevo.2022.906546

Vernes, S. C., and Wilkinson, G. S. (2020). Behaviour, biology and evolution of vocal learning in bats. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 375:20190061. doi: 10.1098/rstb.2019.0061

Wanelik, K. M., Azis, A., and Cheyne, S. M. (2013). Note-, phrase- and song-specific acoustic variables contributing to the individuality of male duet song in the Bornean southern gibbon (*Hylobates albibarbis*). *Primates* 54, 159–170. doi: 10.1007/s10329-012-0338-y

Wich, S. A., van der Post, D. J., Heistermann, M., Möhle, U., van Hooff, J. A. R. A. M., and Sterck, E. H. M. (2003). Life-phase related changes in male loud call characteristics and testosterone levels in wild thomas langurs. *Int. J. Primatol.* 24, 1251–1265. doi: 10.1023/B:IJOP.0000005991.97232.2a

frontiers | Frontiers in Ecology and Evolution

TYPE Methods
PUBLISHED 25 August 2023
DOI 10.3389/fevo.2023.1173722

OPEN ACCESS

EDITED BY
Ulrich H. Reichard,
Southern Illinois University Carbondale,
United States

REVIEWED BY
Emmanuel Dufourq,
Stellenbosch University, South Africa
Cristian Pérez-Granados,
University of Alicante, Spain

*CORRESPONDENCE
Silvy M. van Kuijk
✉ silvyvankuijk@utexas.edu

RECEIVED 25 February 2023
ACCEPTED 31 July 2023
PUBLISHED 25 August 2023

CITATION
van Kuijk SM, O'Brien S, Clink DJ, Blake JG and Di Fiore A (2023) Automated detection and detection range of primate duets: a case study of the red titi monkey (*Plecturocebus discolor*) using passive acoustic monitoring.
Front. Ecol. Evol. 11:1173722.
doi: 10.3389/fevo.2023.1173722

Automated detection and detection range of primate duets: a case study of the red titi monkey (*Plecturocebus discolor*) using passive acoustic monitoring

Silvy M. van Kuijk[1]*, Sun O'Brien[2], Dena J. Clink[3], John G. Blake[4,5] and Anthony Di Fiore[1,5]

[1]Department of Anthropology, The University of Texas at Austin, Austin, TX, United States, [2]Department of Computer Science, The University of Texas at Austin, Austin, TX, United States, [3]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States, [4]Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, United States, [5]Estación de Biodiversidad Tiputini, Universidad San Francisco de Quito, Quito, Ecuador

Passive acoustic monitoring (PAM) – an approach that uses autonomous acoustic recording units (ARUs) – can provide insights into the behavior of cryptic or endangered species that produce loud calls. However, extracting useful information from PAM data often requires substantial human effort, along with effective estimates of the detection range of the acoustic units, which can be challenging to obtain. We studied the duetting behavior of pair-living red titi monkeys (*Plecturocebus discolor*) using PAM coupled with an open-source automated detection tool. Using data on spontaneous duetting by one titi pair, combined with recordings from two Song Meter SM2 ARUs placed within their home range, we estimated that the average source level of titi duets was ~105 dB re 20 µPa at 1 m with an attenuation rate of 8 dB per doubling of distance, and we determined that the detection radius for manual annotation of duets in audio recordings was at least 125 to 200 m, depending on the approach used. We also used a supervised template-based detection algorithm (binary point matching) to evaluate the efficacy of automated detection for titi duets in audio recordings using linear arrays of ARUs within a ~2 km^2 area. We used seven titi duet templates and a set of "off-target" howler monkey (*Alouatta seniculus*) templates to reduce false positive results. For duets with a signal-to-noise (SNR) ratio > 10 dB (corresponding to a detection radius of ~125 m) our detection approach had a recall (the number of all duets that are correctly detected) of 1.0. Performance decreased when including duets with a lower SNR (recall = 0.71, precision = 0.75). The fact that multiple lines of evidence suggest an effective detection radius of 125 to 200 m for titi duets across upland terra firme and seasonally flooded forest lends support to our findings. We suggest that PAM studies of other cryptic but vocally active species would benefit from following similar experimental and analytic procedures to determine an ARU's effective detection radius and to improve the performance of automated detection algorithms.

KEYWORDS

automated detection, autonomous recorders, duetting, attenuation, passive acoustic monitoring, *Plecturocebus*, primate vocalizations

1 Introduction

Duetting is an interactive form of acoustic communication that is relatively common in some pair-living species of tropical birds, such as *Thryothorus* wrens and *Cercomacra* antbirds (Hilty, 2003; Slater and Mann, 2004; Hall, 2009), but can also be found in pair-living mammals, particularly primates (Méndez-Cárdenas and Zimmermann, 2009). In primates, duets are considered "loud calls", i.e., species-specific intergroup vocal signals that are structured to propagate over long distances (Waser and Waser, 1977; Mitani and Stuht, 1998). Duetting has been found in all examined species in the subfamily Callicebinae (genera *Callicebus*, *Cheracebus*, *Plecturocebus*: Caselli et al., 2014; Adret et al., 2018), all but two species of the south-east Asian gibbons (family Hylobatidae, Geissmann, 2002), the lemurs *Indri indri* (Pollock, 1986) and *Lepilemur edwardsi* (Méndez-Cárdenas and Zimmermann, 2009), the genus *Tarsius* (Groves and Shekelle, 2010), and in the colobine *Presbytis potenziani* (Tilson and Tenaza, 1976; Sangchantr, 2004). The structure of duets can vary considerably among primate taxa. For example, pair mates may coordinate their vocalizations to overlap with one another, as seen in titi monkeys (Müller and Anzenberger, 2002; Adret et al., 2018), or to alternate with the partner's vocalizations, as seen in gibbons (Fan et al., 2009). Numerous functional hypotheses have been suggested to explain the evolution of duetting behavior in such a diverse set of primate species. These hypotheses can be categorized based on the presumed audience for the vocalization: the pair mate versus other nearby conspecifics. As a pair-mate focused behavior, duetting has been suggested to have evolved as a mechanism to initially form and later maintain the pair-bond (Geissmann, 1999; Fan et al., 2009). As behavior directed at nearby solitary individuals or neighboring pairs, duetting is proposed to have evolved as a mechanism for mate-guarding (Fan et al., 2009), as a mechanism for communicating with neighboring groups to maintain intergroup spacing (Robinson, 1981; Dolotovskaya and Heymann, 2022), or as behavior associated with collective resource or territory defense (Koloff and Mennill, 2013; Caselli et al., 2015; Dolotovskaya and Heymann, 2022). These hypotheses are, of course, not mutually exclusive.

Despite much interest from scientists, however, empirical tests regarding the evolutionary origin and function of duet calls are difficult to perform. This is in part because – although duetting is a very conspicuous signal – it can be difficult to study in elusive species that hide in vegetation while calling and that are sometimes hard to habituate to human presence (Souza-Alves and Ferrari, 2010; Pinto et al., 2013). The use of passive acoustic monitoring (PAM) provides a method that can circumvent some of these difficulties in studying duetting in cryptic yet vocal animals. PAM is an ecological survey tool that makes use of autonomous recording units (ARUs) programmed to automatically record at a set schedule and deployed at an ecologically appropriate temporal and spatial scale (Deichmann et al., 2018; Sugai et al., 2019). Collecting data with ARUs means human presence in the field can be limited. This method has recently proven to be useful in monitoring taxa that are rare or elusive such as Geoffroy's spider monkeys (*Ateles geoffroyi*, Lawson et al., 2023), Hainan gibbons (*Nomascus hainanus*, Dufourq et al., 2021) and black lion-tamarins (*Leontopithecus chrysopygus*, Zambolli et al., 2023). Although the use of ARUs has great potential for longitudinal monitoring at relatively low cost, it requires careful consideration to calibrate the methodology used to the species of interest.

One important consideration pertains to the propagation and attenuation of vocalizations in different habitats and under different ecological conditions, which can have a major but often overlooked impact on the detection ranges of ARUs across a landscape. Attenuation is commonly discussed in terms of two factors: spherical spreading and excess attenuation. Spherical spreading refers to the natural decrease in sound intensity as soundwaves propagate outward in all directions (e.g., in the shape of a sphere) from a source. In addition to spherical spreading, excess attenuation encompasses additional factors such as scattering (resulting from the interruption of soundwave paths by objects in the environment) and absorption (where sound energy is absorbed by another medium like soil or water) (Bradbury and Vehrencamp, 1998). Source level, frequency range, and call duration are variables that can affect the propagation of vocalizations (Waser and Brown, 1984; Nemeth et al., 2006). For a given frequency, the louder the source level (typically operationalized as the intensity of the sound at 1 meter from the source) of the primates' vocalizations, the further the call will propagate (Bradbury and Vehrencamp, 1998). However, source levels are known or can be estimated for only a few primates (Table S1, Supplementary Material). Lower frequency sounds attenuate less quickly due to their longer wavelengths (Bradbury and Vehrencamp, 1998), which is why primate loud calls tend to have lower frequencies than other vocalization types (Mitani and Stuht, 1998). Habitat characteristics, including canopy density, ambient noise levels and topography, as well as weather conditions such as temperature, humidity and wind speed also affect propagation distances of calls (Ellinger and Hödl, 2003; Darras et al., 2016; Gibb et al., 2019). Excess attenuation disproportionately affects vocalizations in heavily forested environments like rainforests, where dense vegetation and complex structures contribute to increased absorption and scattering (Brown and Waser, 2017). However, certain frequency ranges may be less affected by attenuation, creating a "sound window" that is more suitable for long-distance communication (Waser and Brown, 1984). Consequently, selection pressures may lead to the emergence of long-distance vocalizations that have much of their intensity within these sound windows of lower attenuation, even resulting in acoustic differences between populations of the same primate species inhabiting different habitats (Sugiura et al., 2006). When recording vocalizations with an ARU, various factors such as device settings, height above the ground of the recorder and microphone (Padgham, 2004; Rempel et al., 2013; Darras et al., 2020), as well as the direction of an animal's vocalizations (Pérez-Granados et al., 2019) also influence the distance at which calls can be recorded.

The use of ARUs allows us to expand the spatiotemporal scale of our research, but often results in the accumulation of vast amounts of audio data that need to be processed. Manually processing audio recordings to identify the start and stop times of vocalizations of interest (an approach known as annotation) is time-consuming and prone to error and bias. For example, if recordings are being reviewed by multiple observers, then differences in experience or perceptual abilities may be an additional source of variation that needs to be accounted for in analyses (Swiston and Mennill, 2009). Similarly, if recordings are all reviewed manually by the same listener, order effects or listener fatigue could introduce error. For ARUs to be a valuable resource in our scientific endeavors, automated tools for screening recordings to detect and classify calls can be important for addressing this bottleneck in data processing. Classification tools predict categories of a signal of interest, such as its species or vocalization type (Stowell, 2022). Detection tools either generate a binary classification (presence or absence of a signal) for audio files or generate the location of a signal of interest by listing its start and end times within an audio file (Stowell, 2022). Moreover, automated detection tools are often "deterministic algorithms", meaning that they can improve reproducibility in that, when well documented, the same automated audio file processing pipeline, run on the same dataset should yield exactly the same results. Still, automated detection tools present their own set of errors and biases (Digby et al., 2013), thus the tools are most valuable if the increase in time and space that can be surveyed outweighs the limitations of the detection algorithm and pipeline, or if the errors are more predictable and consistent in nature than those associated with manual annotation; however, error associated with human annotations is often ignored (Swiston and Mennill, 2009; Digby et al., 2013).

Methods for automated detection of acoustic signals within audio recordings are often machine learning-based classification algorithms that can be divided, conceptually, into "supervised", "semi-supervised", and "unsupervised" approaches. Whereas supervised and semi-supervised methods use a set of training data that is labeled by an observer (e.g., marking which audio files do and do not have duets after manual inspection), an unsupervised algorithm looks for patterns in the provided data without any prior information given by a human observer. Some of the commonly used supervised detection algorithms include Support Vector Machines (SVM; Noble, 2006; Heinicke et al., 2015; Clink et al., 2020), Gaussian Mixture Models (GMM; Bishop, 2006; Janvier et al., 2013; Heinicke et al., 2015), and K-nearest neighbors (Janvier et al., 2013; Bayestehtashk et al., 2014; Taunk et al., 2019). Some other methods, like Hidden Markov Models (HMM; Eddy, 2004; Porcaro, 2015) and artificial neural networks (ANN; Krogh, 2008; Pozzi et al., 2012) can be used for supervised, semi-supervised, or unsupervised classification. A variety of automated methods have been used to detect primate acoustic signals for at least 16 different primate species (Table S2, Supplementary Material), although very few of these studies have focused on primates of the Americas. In addition, automated detection methods have rarely been used for identifying duetting behavior (e.g., Schroeder and McRae, 2020; Szymański et al., 2021), and, to our knowledge, only three studies have used automated approaches with duetting primates: band-limited energy detection in gibbons (*Hylobates funereus*, Clink et al., 2023) and convolutional neural networks in gibbons (*Nomascus hainanus*, Dufourq et al., 2021) and indris (*Indri indri*: Ravaglia et al., 2023).

In this paper, we describe our approach using PAM and an open-source automated detection tool to study the duetting behavior of red titi monkeys (*Plecturocebus discolor*). Red titis are small-bodied, pair-living primates found in Colombia, Ecuador, and Perú (Vermeer and Tell-Alvarado, 2015). Though mostly cryptic, titi individuals produce loud calls of various types (e.g., solo calls, duets, and choruses) often in the early morning (Robinson, 1981; Kinzey and Becker, 1983; Aldrich, 2006; Van Kuijk, 2013; Dolotovskaya and Heymann, 2022). In the field, the duet of one pair is often followed by response duets from neighboring pairs (Caselli et al., 2015). The titis' cryptic behavior means that many duets are sung from within hard-to-observe areas, such as vine tangles or the dense vegetation of sleeping trees (Kinzey and Becker, 1983; De Luna et al., 2010). In addition, unhabituated groups will often cut their calls short when people or other potential threats are nearby. Using PAM, however, it is possible to record duets and other loud calls from multiple groups without interfering with the primates' natural behavior.

Here we use multiple complementary datasets to explore a number of important methodological issues relevant to using PAM to study titi duetting behavior. These issues have general relevance for PAM studies of other cryptic but vocally active species. First, we determined the average source level of titi monkey duets using data on spontaneous duetting by a titi pair, in combination with recordings from two Song Meter SM2 ARUs within their home range (hereafter "home range dataset"). Second, we examined the detection radius of duets (the radius around a recorder in which duets can be reliably recognized) on Song Meter recording devices using standardized playbacks at known distances from an ARU (hereafter "playback dataset"). Variation in ambient noise levels and other variables can change the detectability of a duet. Therefore, our third aim was to determine the detection probability of duets within the detection radius of the recorder by calculating the proportion of all duets that were recorded by the ARU. Fourth, the ability to detect vocalizations (by ear or on an ARU) is also influenced by the intensity of the call in comparison to the intensity of ambient noise and the amount of excess attenuation that affects the call as it propagates through the environment. Therefore, we examined how the signal-to-noise ratio (SNR) of duets decreases with distance using a combination of the home range dataset, playback dataset and audio recordings collected systematically during the early morning hours (05:45 to 08:10) from ARUs placed along a series of linear N–S transects from across the study area (hereafter "transect dataset"). We also estimated the amount of excess attenuation the duets experience on top of spherical spreading. Then, using a supervised template-based detection algorithm (binary point matching), we automatically detected the duets in the audio of the transect dataset. We also used our data to evaluate the SNR of duets that are detected by the automated algorithms and used this result to estimate the detection radius of duets on the Song Meter ARU. Finally, we highlight some methodological issues that must be considered when using PAM and automated algorithms to study duetting primates.

2 Methods

2.1 Study location

We conducted this study at the Tiputini Biodiversity Station (TBS, 00°37'05" S, 76°10'19" W, 190–270 m a.s.l.), located on the northern bank of the Tiputini River, in the province of Orellana in Ecuador (Figure 1A). TBS is adjacent to the Yasuní National Park and is part of the larger Yasuní Biosphere Reserve. The station's presence preserves a ~700 ha tract of primary tropical rainforest consisting of mostly *terra firme* forest (lowland evergreen forest) along with some várzea and igapó regions (two types of flooded lowland evergreen forest) near streams, rivers, and a small lake. Annual precipitation averages 2924 mm ± SD 267 mm and temperature has a monthly average ranging from 23–25°C throughout the year (Van Belle et al., 2018).

2.2 Study species

Three pairs of red titi monkeys and their offspring (groups K, L, and B) were habituated and have been studied regularly at TBS since 2003 as part of a long-term comparative study of sympatric pair-living primates (Van Belle et al., 2021). Our data on home range use and vocal behavior (used to determine duet source levels, to examine the detection probability of calls, and to measure decrease of SNR with distance) were collected in the range of group L in June and July of 2016. This group consisted of a male/female pair that shared this range from at least March 2009 through the period of this study. The pair had two offspring at the time recordings were collected: a sub-adult male born in January/February of 2014 and a juvenile female born in December of 2014. During the time group L was studied, in June/July of 2016, the subadult male was seen chorusing with his parents and singing solo calls. The juvenile female did not participate in any duets or produce any loud calls on her own.

2.3 Audio data collection

2.3.1 Home range dataset

Our first dataset, the "home range dataset", was collected with the aim of estimating the source level of titi monkey duets, characterizing duet propagation loss over distance, and determining the detection radius of duets (the radius around the Song Meter SM2 ARUs in which duets are recorded). We placed two Song Meter SM2 ARUs (Wildlife Acoustics, Inc., Maynard, MA, USA) equipped with two SMX-II omnidirectional microphones each (frequency response 20–20,000 Hz) within group L's home range at a height of 12 m for 34 days from mid-June to late July of 2016. Gain settings of the devices were left at the default 48 dB. As the recorders had to be placed in the canopy with the use of a slingshot and ropes, a combination of logistics and home range knowledge determined the deployment location of the recorders: we needed enough open space to use the slingshot, yet this had to be a location in which group L frequently spent time. Once a suitable location was found (Figures 1B, C), we recorded the location of the SM2 ARUs using a Garmin 76Cx GPS. The ARUs were attached perpendicular to one another to a metal frame so that the four microphones on the two ARUs were all spaced apart equally. The ARUs were set to record 24 hours per day at 16-bit resolution with a 44 kHz sampling rate. Audio files were saved every 60 minutes in high-quality uncompressed (lossless) waveform format (WAV). Batteries and SD cards were exchanged roughly every 5 days.

2.3.2 Playback dataset

With our second dataset, the "playback dataset", we aimed to examine the detection radius, detection probability, and propagation loss of duets in a more standardized way and at larger distances than our home range dataset allowed for. We played a 3-second clip of a duet previously recorded of group K at different specified distances from a stationary recorder along two trails in July of 2016. We attached a single SM2 ARU at a height of 12 m to a permanent canopy tower that was built around an emergent kapok tree (*Ceiba pentandra*). The ARU was configured to record at 16-bit resolution using a 44 kHz sampling rate. With the use of a Garmin 76Cx GPS, we then created one trail with GPS-mapped locations at 10, 25, 50, 75, 100, and 125 m from the base of the tower and a second trail with GPS mapped locations at the same distances plus additional locations at 150, 175 and 200 m from the tower. The length of these trails was limited to 125 m and 200 m due to geographical changes in the terrain.

At each of the locations along the trail SvK played the group K duet using a Sony ICD-UX533 digital voice recorder connected to a

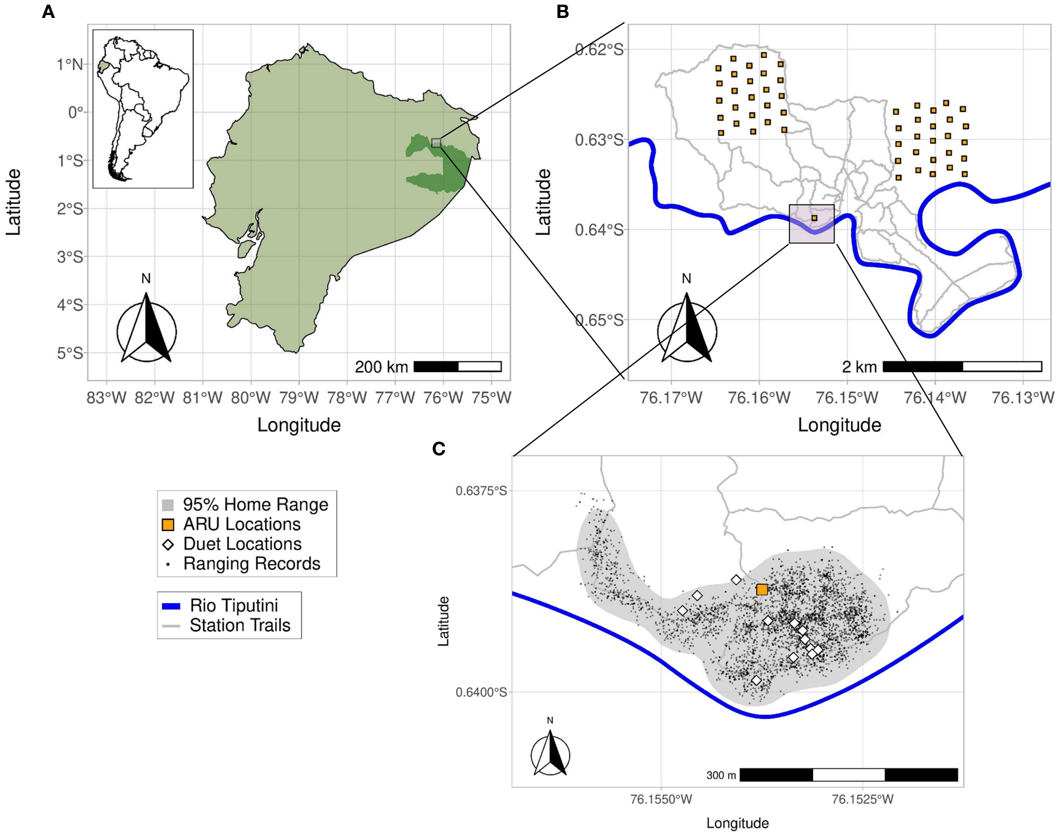


FIGURE 1
(A, B) The study site at the Tiputini Biodiversity Station is located adjacent to Yasuní National Park (dark green inset), Ecuador, on the left bank of the Rio Tiputini. Recorders for the transect dataset were located in two 1x1 km plots located within or along the trail system of the station. **(C)** The gray area marks the 95% kernel density home range for titi group L and the white diamonds are the locations of 12 duets recorded during behavioral follows of group L that were also recorded by the SM2 ARU (orange square). The small black dots are GPS ranging locations collected at 20-minute intervals while observers followed group L during field seasons from Oct to Jan 2015.

Pignose Legendary 7-100 portable amplifier at a height of ~2 m. Because the source level of titi monkey duets was unknown at the onset of this part of the study, we repeated these recordings at 80, 90 and 100 dB re 20 µPa at 1m (A-weighted) along both trails so we could analyze, *post hoc*, the data from the amplitude that is the closest match to the estimated source level of titi duets. We calibrated the three amplitude levels at 1 m from the Pignose amplifier with an American Recorder Technologies sound level meter. We repeated our playback recordings of all three amplitude levels twice at each distance and did so on two consecutive mornings: one morning with clear sunny weather and one overcast yet dry morning.

2.3.3 Transect dataset

Our last audio dataset, the "transect dataset", is mainly used to evaluate the efficacy of a supervised template-based detection algorithm (binary point matching) for automated detection of titi duets in audio recordings. In addition, we use these data to examine how distance influences the detection of duets using multiple ARUs spaced at regular intervals along transects. The transect dataset comprises simultaneous recordings collected by one of us (JB) between late January and early March of 2013–2017 and originally intended for use in ornithological studies (Blake, 2021). For this third dataset, JB collected simultaneous recordings using a set of five Song Meter SM2+ ARUs placed at 200 m intervals along 10 different N–S transects within the TBS study area. Overall, these recorders were placed at a total of 50 locations, 25 in each of two approximately 100-ha research plots (Figure 1B), i.e., they were spread over an area of roughly 2 km^2 and covered a range of microhabitats and topographies. These plots were established in 2001 and contain trails every ~100 m from east to west and every ~200 m from north to south. The grid is marked and GPS mapped every 50 m. For each day of recording, JB deployed five devices along one of the north–south grid lines in the plots at the intersections with the east–west transects, leading to a total distance of 800 m between recorders at opposite ends of the transect. The ARUs were positioned at 1.5 to 2 m off the ground and configured to record 10-min audio files with 16-bit resolution and using a 16 kHz sampling rate. The resultant 3600 second audio files were saved in high-quality uncompressed (lossless) waveform format (WAV).

The ARUs were programmed to start recording at 05:45 am and stop recording at 08:10 am, as the vocal activity of birds typically declines rapidly in the 2 hours following sunrise (Blake, 1992). Each 10-minute recording was separated by a 5-minute break, leading to a collection of ten recordings per device for each morning of deployment. Although the intermittent recording schedule and restricted time frame of sampling did not allow us to capture all titi monkey loud calls emitted in range of the recorders on any given morning, our data show that this time frame is appropriate for this study as 67.7% of all duets are heard before 07:30 (Figure 2). The devices were left in their locations until they had collected two full mornings of recordings without rainy weather and were then moved to a new N–S transect within the plot until all 25 locations per plot (5 N–S transects × 5 ARUs per transect) had been sampled. Batteries and SD cards were replaced as needed.

2.4 How loud are titi monkey duets and what is their detection probability?

To evaluate when titi monkeys were most likely to sing duets, we analyzed data collected by six observers that followed titi monkeys between 2007 and 2016. From our long-term behavioral database we determined the total number of hours these observers spent in the field and the total number of duets they heard (either from monkeys they were following or from other groups) during this time. From these data, we calculated the duetting rate (number of duets heard per 100 observer field hours) for each hour of the day from 05:30 to 18:30, or from just before sunrise to just after sunset. Most duets occur during the early morning (Figure 2). During the 2016 field season, we regularly followed group L from 05:30 to ~11:00, after which time duetting is infrequent. On mornings with heavy rain, we would follow the group as soon as rain let up, usually locating them still in their sleep tree. During follows, we used the same model GPS to record the group's travel path. As the monkeys generally do not travel when duetting, we also recorded the locations of all duet vocalizations emitted during the follow. We also recorded data on the timing and duration of duet sequences as well as any vocal responses of nearby groups.

To determine the detection probability of duets on the SM2 ARU in our home range dataset, we generated spectrograms of the audio files from both ARUs from the same days and hour as each duet in Raven Pro 1.6 (K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA) by applying a fast Fourier transform with a 1024-point Hann window (3 dB filter bandwidth = 61.9 Hz), 50% overlap, and a 1024-point DFT, with time and frequency measurement precision of 32 ms and 15.6 Hz. We then reviewed the sections of the recording where we expected, based on follow data, to find a duet. We matched the timing and duration of the duet sequences to ensure we could distinguish between group L's duets and other duets. We calculated the SNR for the duet on all four spectrograms (one for each microphone) and used only the recording from the microphone with the highest SNR value as that presumably represents the microphone oriented most directly towards the monkeys.

From the home range dataset, we also estimated the source level of the titi duets using the R package 'PAMGuide' (Merchant et al., 2015). The absolute received level of the duet was measured from a 5-second segment at the start of the duet. PAMGuide allows for calculating the absolute received level of the signal when relevant hardware specifications are known. In our case, we provided PAMGuide with information on transducer sensitivity (−36 dB re 1 V/μPa at 1 kHz), gain settings (+48 dB), and the voltage of the analogue-to-digital converter (1.414 V). Then, using the resulting calibrated amplitude measurements for the duets as recorded on the ARU and the distance between the recorder and the vocalizing monkeys (calculated using the GPS coordinates of both points), we applied the inverse square law to estimate the source level of the duets at 1 m from the monkeys. Importantly, the average source level resulting from this calculation fails to account for the effect of

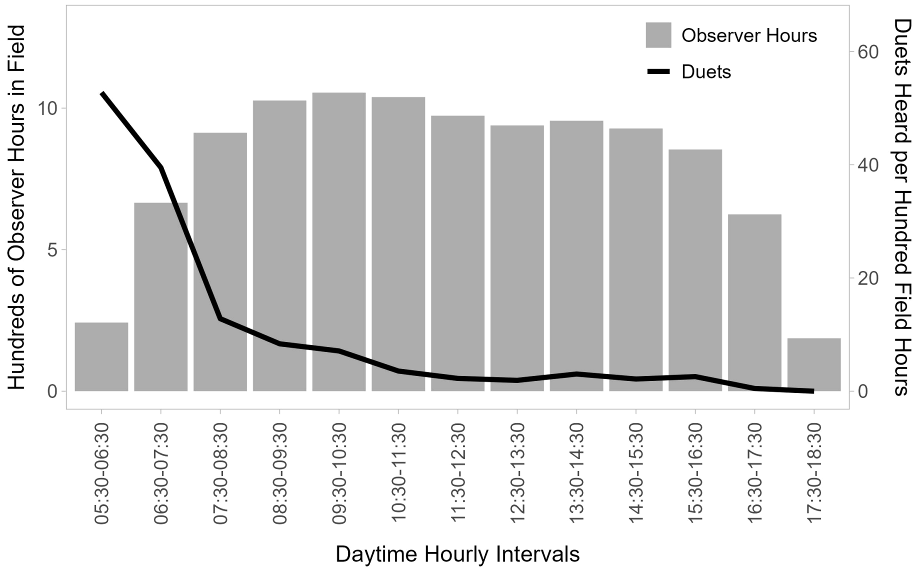


FIGURE 2
Temporal distribution of duet calls. Gray bars show the distribution of total observer field time (hundreds of observation hours between 2007 and 2015 for six observers focused on following titi monkeys) per hour of the day between 05:30 and 18:30, and the black line shows the number of duets recorded per 100 observer hours in field in each hourly block. Duetting is concentrated in the window from shortly before sunrise to the first few hours after sunrise (05:30 to 08:30).

excess attenuation and how this covaries with distance between the source and a recorder. We therefore calculated the average source level for the two duets closest to the recorder and used this value as a minimum estimate of the actual source level, from which we inferred the theoretical decay in SPL with distance assuming spherical spreading only. We then compared theoretical to observed SPL values at different distances between source and recorder to estimate the effect of excess attenuation on titi duets.

2.5 What is the detection radius of titi duets using our ARU setup?

We used both the home range and playback datasets to determine the detection radius of duets on our ARU setup. Because the duets in our home range dataset were limited in their distance from the recorder, the playback dataset allows us to expand upon the home range dataset results by increasing the tested distances. Spectrograms for the playback dataset audio files were created using identical specifications as discussed in Section 2.4 for the home range dataset. We only created spectrograms from the microphone most directly pointed towards the playback locations. The spectrograms were inspected by two observers who had no field experience with titi monkeys, but who were trained in annotating duets in audio recordings in Raven Pro. The observers were naive with respect to the timing of the duets. The observer-generated data was used to determine the proportion of playback stimuli at each playback distance that was manually recognized and annotated.

2.6 How does distance influence detection of titi monkey duets?

To determine at what distance from an ARU duets can still be detected in audio recordings, either by human observers or machine learning algorithms, we examined the relationship between the signal-to-noise ratio (SNR) of duets extracted from recordings and distance between the caller and the recorder in all three audio datasets (see Section 2.3).

For our transect dataset, we initially reviewed only data from one randomly chosen ARU from each series of five to avoid pseudo-replication by having the same duet reflected in the dataset more than once. This left us with a total of 230 hours of acoustic data. To locate duets within the audio files, SvK and one other observer generated spectrograms of the recordings in Raven Pro 1.6 using the same settings described in Section 2.4. The resulting spectrograms were again manually inspected by paging through 75 second windows at a time and playing back portions of the files to identify duets that were hard to see in the spectrogram yet still audible. Each detected duet was then assigned a subjective "quality score" ranging from 1 through 4, based on the listener's qualitative perception of the recording (Figure 3). We scored duets as quality 1 when the monkeys were calling close enough to the recorder that they could be heard moving through the vegetation or giving quiet contact vocalizations. Duets were assigned quality score 2 when the titis could no longer be heard moving, but the image of the call in the spectrogram showed evidence of multiple harmonics. Quality score 3 was assigned to duets that were audible to the listener and had clearly visible fundamental frequencies but showed far fewer harmonics than calls of quality score 1 or 2. Quality score 4 was assigned to duets that were barely audible and had only faint fundamental frequencies visible in spectrograms.

For a subset of the duets detected through the manual procedure described above, we then investigated whether the same duet could be heard on any of the other four devices that were recording at the same time. For this subset, we investigated all duets for which we had assigned qualitative quality scores of 1 through 3 (N=17 duets) plus an additional 16 duets assigned a quality score of 4. We then calculated the SNR of the calls appearing simultaneously on each of the recorders to determine which of the five recorders the group was closest to. SNRs were calculated using Raven Pro by drawing a 5-second selection box with a frequency range of 700–1400 Hz around the start of a duet. In case of interference of other organisms' vocalizations at the start of a duet, we used the first 5-second section thereafter without interrupting calls of other animals. A second selection box with an identical duration and frequency range was drawn around a section of ambient noise shortly before the duet or, in case of interfering noise, directly after the end of the duet (e.g., Figure 3A). The 700–1400 Hz frequency range was chosen as it encompasses the peak frequencies (i.e., the loudest spectral components) of a duet (which are most likely to propagate furthest through the environment). The duration was chosen such that it is fairly easy to find a suitable area of ambient noise in the recording without other loud vocalizations or other interfering noises.

For the 5-second selection boxes around the duets and ambient noise, we used Raven to calculate the "Inband Power (dB)". We converted the inband power measurements for these two selections from dB to linear units using the formula $y = 10^{x/10}$, where x is the inband power in dB and y is the inband power in linear units. We then calculated the SNR in linear units using the formula $\mathrm{SNR}_{\mathrm{linear}} = (y\ Signal - y\ Noise)/y\ Noise$, where *y Signal* and *y Noise* are the inband power in linear units for the duet and ambient noise selections, respectively. Last, we turned the SNR from linear units back into decibel units by using $\mathrm{SNR}_{\mathrm{decibels}} = 10 \times \log_{10} (\mathrm{SNR}_{\mathrm{linear}})$ (K. Lisa Yang Center for Conservation Bioacoustics, 2022). Because the distance between titi monkeys and ARUS in the transect dataset is unknown, we used the SNRs of the different datasets to determine at what distance from the ARU the SNR is too low for a duet to reliably be detected by the algorithm.

2.7 How well do automated detection algorithms perform?

We examined the efficacy of automated detection of duet calls in audio files from our transect dataset using the binary point matching (BPM) template matching algorithm implemented in the monitoR package (Katz et al., 2016) for the R statistical programming environment (version 4.1.2; R Core Team, 2022). Because titi monkey duets are diverse and change in pattern throughout the duration of the song, we chose a random selection

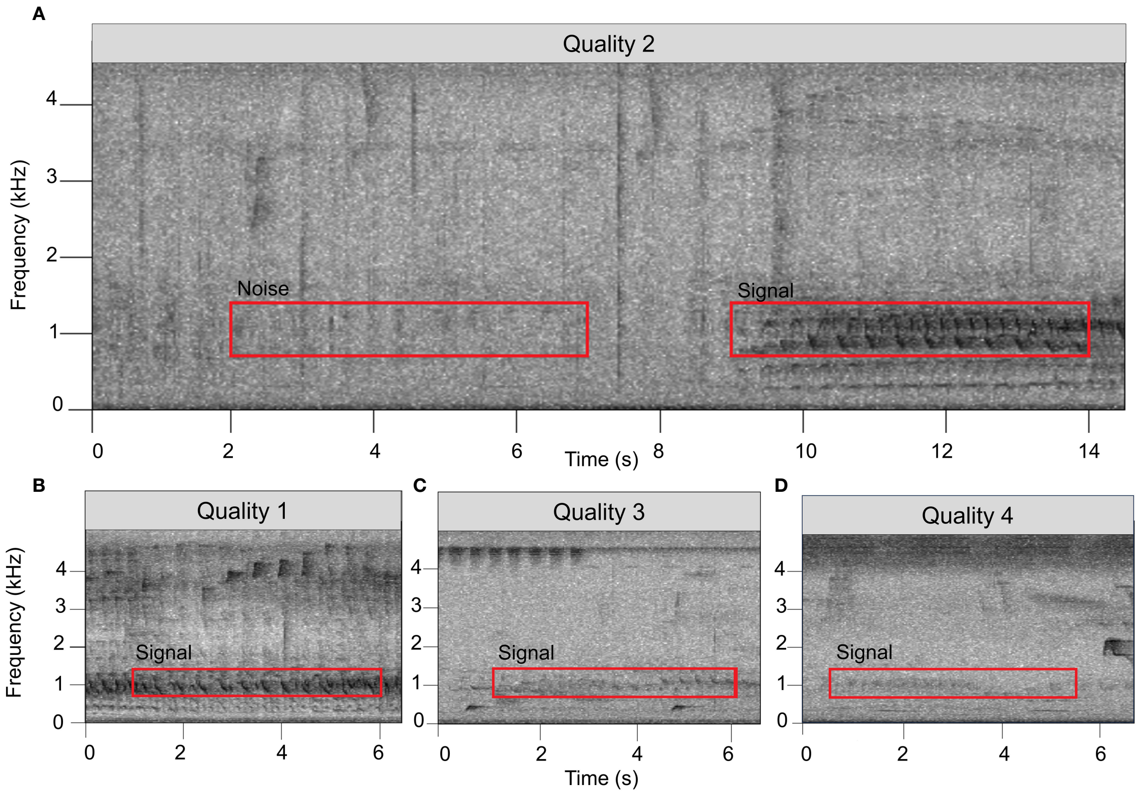


FIGURE 3
The SNR of a signal was determined by comparing the root mean square amplitude of a 5-sec, 700–1400 Hz window at the onset of a signal of interest (red box labeled "Signal") to a comparably-sized window with the same frequency range in a low-noise area of the recording (red box labeled "Noise"). The onset of a duet was determined by locating the first loud note (the "bellow") after the fainter introductory notes. **(A–D)** show examples of duets assigned by listeners to four subjective "quality" scores.

of duets to create templates from in order to capture the full variation of these complex vocalizations. To increase the odds of detecting duets at short and long distances from the recorder, we also chose high-quality duets at a range of distances to build templates from. We created seven different templates that we used to screen each audio file: one duet was taken from a short-distance recording made with a Marantz PMD 660 digital recorder and Sennheiser ME67 shotgun microphone with K6 power module, and 6 duets with various assigned quality scores extracted from the transect dataset. BPM templates are created from these reference calls by specifying a frequency range, time range, and amplitude cutoff as multidimensional parameters. We chose a −25 dB amplitude cutoff and 10-second time frame for all templates because this created the clearest distinction in similarity scores between test files containing duets and test files that did not contain duets. In addition, the frequency range was specific to each template, to accommodate differences in SNR of the duets and surrounding ambient noise. The minimum frequency ranged from 200 to 300 Hz and the maximum frequency ranged from 1400 Hz (duets with lower SNR) to 1800 Hz (duets with higher SNR) depending on the number of visible harmonics.

Duet templates were initially tested against a validation dataset. This was a sample from the transect dataset that included 8 days of data (i.e., 80 10-minute audio files), and included the presence of two known titi duets, one titi burst gobble (a loud call of shorter duration but similar frequency range to the duet, similar to Caselli et al.'s (2014) "type 1 call" and Robinson's (1979) "short sequence"), five howler monkey (*Alouatta seniculus*) roars, and a multitude of other common loud vocalizations such as bird calls in a similar frequency to titi duets that could potentially cause false-positive detections. We then tested multiple similarity cut-off scores as the detection threshold for a titi monkey, identifying a cut-off score of 4.0 as one that led to a low number of false negatives and a low number of false positives in our test data.

We then applied this detection pipeline to a large test dataset of 1099 additional 10-minute audio files from the transect dataset (and excluding the eight days used for training the data) and compared the results of that detection process to manual annotation of the same 1099 files. For each audio file inspected, we calculated an average maximum similarity score for the set of seven titi templates (and always excluding scores for when a particular template was applied to the file it was extracted from). The automated detection algorithm's performance can be visualized using a confusion matrix, which classifies true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Novaković et al., 2017; Flach, 2019). That is, a true positive is when the algorithm correctly detected a titi duet in an audio file, whereas a false positive is when a different vocalization or ambient noise was mistaken for a titi duet. True negatives are files that the algorithm correctly classified as not containing a titi duet and false negatives are files containing a titi duet that the algorithm failed to detect.

The initial analysis with this set of templates (Pipeline 1) used only a mean similarity score cut-off of 4.0. Given that a low SNR value for a signal of interest in recordings is known to lead to decreased algorithm performance (Spillman et al., 2017), to evaluate the role of SNR on algorithm performance for titi duets, we modified our dataset to exclude quality score 4 duets. Pipeline 2 then used the mean similarity score of 4.0 on this modified dataset. Because large numbers of howler monkey roars were erroneously classified as titi duets (i.e., false-positive detections), we created an additional three templates specific to howler monkeys and screened recordings with these templates as well. Pipeline 3 used our original test dataset and scored a detection when two conditions were met: (1) the titi template score was higher than 4.0 and (2) when the titi template score exceeded the howler template score for the audio file. Last, Pipeline 4 used the detection rule including the howler monkey templates and the modified dataset that excluded quality score 4 duets. More detailed description of algorithm performance and the four pipelines can be found in the Supplementary Material (Data Sheet 1).

We used common metrics to evaluate the success of the four pipelines: precision, recall, balanced accuracy, and the F1 score. Precision is the portion of all positive detections that are correct detections. Recall (also called sensitivity) is the proportion of all duets in the manually scored dataset that are correctly detected. Recall suggests how well the algorithm detects duets, whereas precision stipulates the reliability of the algorithm. Accuracy is the proportion of all predictions that were correct, whether those are detections or nondetections. Because this number would be skewed by a large number of true negatives, we use balanced accuracy instead, which is calculated as the average of recall and specificity (i.e., the proportion of true negatives divided by all negatives). The F1 score is the harmonic mean of precision and recall. One major point of criticism of automated detection procedures such as those implemented in monitoR is that there is often a high number of false-positive detections (Barclay, 1999; Swiston and Mennill, 2009). As red titis at our study site duet infrequently, false negatives are of larger concern as this would reduce the already small number of duets that naturally occur in any dataset. However, we attempted to find a balance that keeps both the number of false negatives as well as false positives reasonable. Last, as the detection probability of a duet strongly depends on its SNR, we calculated the SNR of all duets to investigate how SNR influences detection probability of the duets.

3 Results

3.1 How loud are titi monkey duets and what is their detection probability?

Five observers recorded 820 duets in 10,439 hours in the field between 2007 and 2016. Only 28 (3.3%) of these duets were recorded before 06:00 am, yet between 05:30 and 06:30 we recorded the highest number of duets per 100 hours at 55.6. We recorded 41.5 duets per 100 hours between 06:30 and 07:30, after which the number of duets per 100 hours strongly declines (Figure 2).

A total of 13 duets were recorded in the home range dataset for group L during behavioral follows on nine of the 34 days that the ARU was active, meaning they sang roughly once every 3 to 4 days. Of the 13 duets, two were spontaneous calls (i.e., either the first calls of the day detected by the observer or the first ones visible in the audio recordings) that often elicited responses from neighboring groups. The other 11 duets were responses to nearby groups' duets as we heard calls of other groups shortly before the onset of group L's duet. The earliest duet from group L was recorded at 06:18 and the latest duet occurred at 09:35. The duration of duets varied from 33 seconds to 6 minutes and 53 seconds, with an average length of 3 minutes and 20 seconds.

We were able to locate 12 of these 13 duets in the ARU recordings (Figure 1C). *Post-hoc* inspection of the data revealed that the missing (not recorded) duet resulted from technological difficulties with the recorder. The calculated distance between the GPS points of the ARU and duetting locations ranged from 38 to 125 m. Mean source level of the 12 duets as estimated based on spherical spreading only was 84.6 dB re 20 μPa at 1 m ± SD 4.9 dB. The mean amplitude estimated on the basis of the two closest duets only was 91.9 dB re 20 μPa at 1 m (see also Table S1, Supplementary Material). This increase in estimated source level in comparison to the overall average demonstrates that duets are affected by excess attenuation in addition to spherical spreading. Figure 4A demonstrates the difference in the theoretical decay in SPL with only spherical spreading (red line) as compared to the observed decay in our dataset (blue line), assuming an SPL at 1m of 91.9 dB re 20 μPa as estimated above. Figure 4B shows the estimated excess attenuation in relation to distance by plotting the difference in SPL between the theoretical and observed curves of Figure 4A. Additionally, using our playback dataset, it was possible to estimate the amount of excess attenuation that duets experience at different distances from a speaker by comparing theoretical versus observed received levels for a given source level (100 dB re 20 μPa at 1 m, Figures 4C, D). We estimated that the two closest calls recorded in our home range dataset would have an average excess attenuation ~12.9 dB (Figure 4D), suggesting that the source level for these calls would actually be ~105 dB re 20 μPa at 1 m (91.9 + 12.9 = 104.8) (Table S1, Supplementary Material). Moreover, using our playback dataset, we found that excess attenuation averaged 2 dB per doubling of distance, in addition to the 6 dB attenuation due to spherical spreading. This leads to an estimated total attenuation rate of 8 dB per doubling of distance for titi duets at TBS.

3.2 What is the detection radius of titi duets using our ARU setup?

In the home range dataset for group L, the furthest duet from the ARU location, at 125 m, was easily visually detected in the

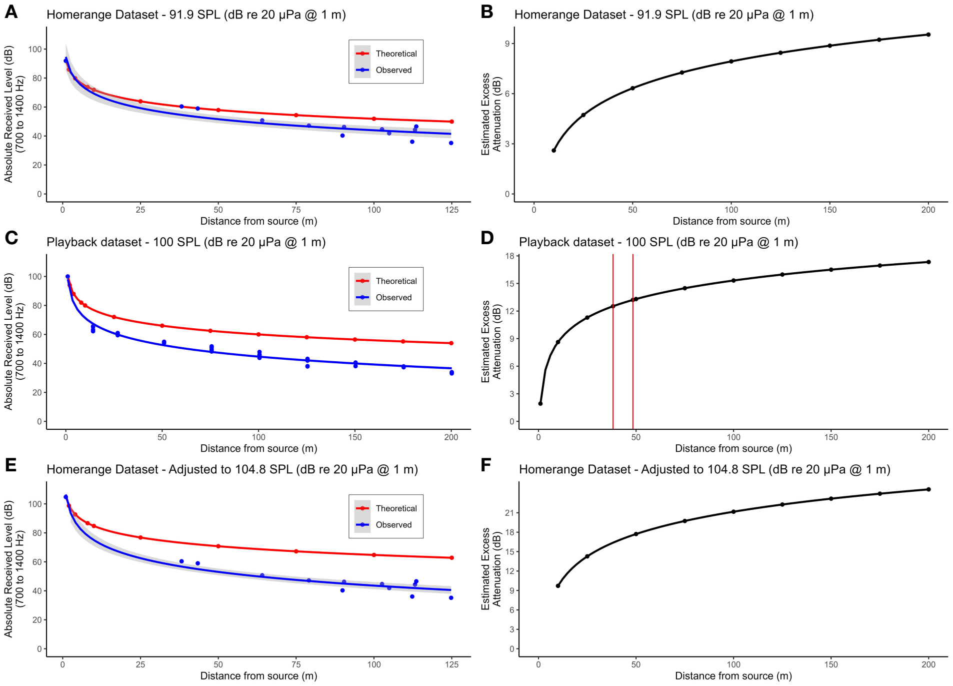


FIGURE 4

(A) uses the average source level of the two duets recorded closest to the ARU to calculate the theoretical decay in SPL over distance assuming spherical spreading only (red dots). The blue dots show the observed SPL of the recorded duets. Both sets of data points are fitted to a line of best fit of exponential decay (red and blue lines). **(B)** shows the estimated effect of excess attenuation on the titi duets by calculating the difference in SPL of the theoretical and observed values at 25 m intervals between 0 and 200 m. **(C, D)** show the decay in SPL and estimated effect of excess attenuation, respectively, for our standardized playback experiments. The vertical red lines in **(D)** reflect the distances of the two duets recorded closest to the ARU (38.2 and 48.4 m). The excess attenuation values for those two duets were then used to adjust our home range dataset values for excess attenuation, reflected in **(E, F)**. For the home range and playback datasets, ARUs were positioned 12 m above ground.

spectrogram, suggesting that SM2 recorders have a minimum detection radius of 125 m. Our playback experiment allows us to expand on these findings. We focused our analysis on the playbacks conducted at 100 dB re 20 µPa at 1 m because this most closely matches our average estimated source level for titi monkey duets. The visibility of duets in spectrograms was comparable to duets with a quality score of 1 through 3 in our transect dataset. Duets were clearly detectable by visual inspection of the spectrograms, showing little loss of harmonic structure for the closest duets and some loss of the highest harmonics for more distant duets. That is, we never had to listen to audio files to confirm the presence of a duet, even at our furthest tested distance of 200 m. These results suggest we can increase our estimated detection radius from 125 to at least 200 m.

3.3 What is the detection probability of a duet within the detection radius?

For our home range dataset, all the 12 duets recorded during behavioral follows were also captured on the ARU audio files (one additional duet recorded during behavioral follows took place during a time when the Song Meter malfunctioned). Based on this limited sample, titi monkey duets have a detection probability of 100% when given within a 125-meter radius around the Song Meter ARU.

In manual review of audio files generated in our playback experiments, naïve observers annotated all playback duets at all distances up to our maximum distance of 200 m. Again, this result suggests that playback recordings conducted at 100 dB have a detection probability of 100% up to at least 200 m.

3.4 How does distance influence detection of titi monkey duets?

For our transect dataset, duets were never detected on all five recorders simultaneously. Five duets (15.2%) were only heard on one recorder; 18 duets (54.5%) were heard on at least one recorder 200 m away from the ARU where the call was recorded with the highest SNR; 10 duets (30.3%) were heard on at least one ARU 400 m away; and no duets were heard on ARUs 600 or 800 m away from the one where the highest SNR was found. The highest SNR values seen in any set of five simultaneous recordings (i.e., at the recorder that the monkeys were presumed to be closest to) ranged from 1.2 to 32.5 dB. In Figure 5A, we plot the relationship between SNR and distance from the recorder with the highest SNR value, splitting these data into three different sets of boxplots based on the listeners' assignment of a quality score to the recording on the ARU with the highest SNR. Though we cannot relate SNR to precise distances between the recorder and primates, our data nonetheless clearly show how SNR decreases with distance. Notably, for all quality-

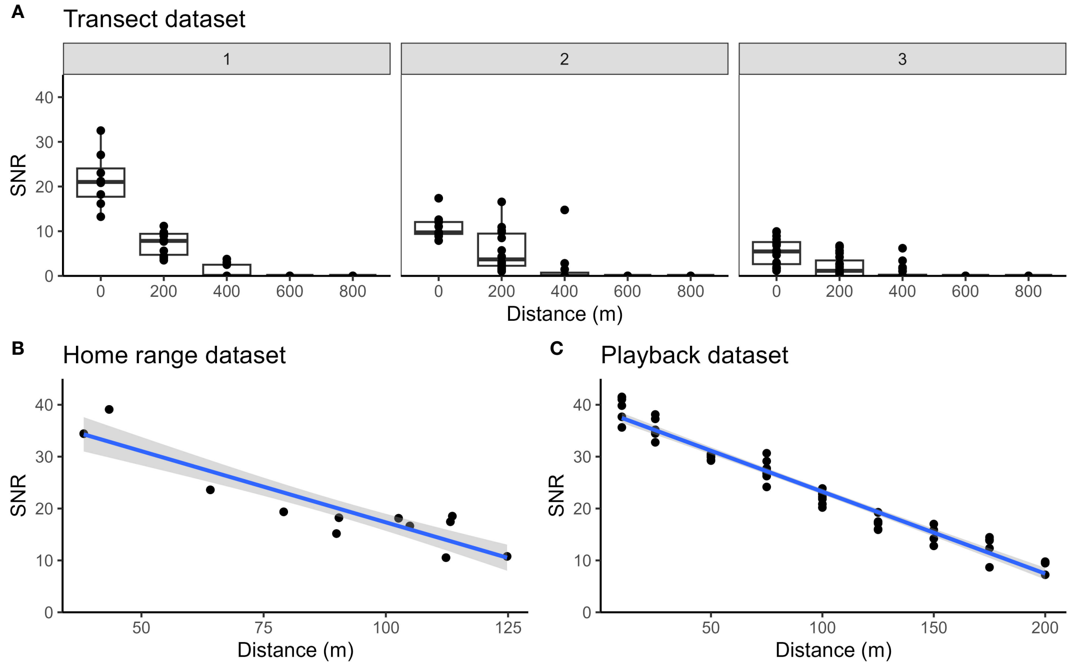


FIGURE 5
Changes in SNR with distance from the ARU, plotted for **(A)** transect dataset, **(B)** home range dataset, and **(C)** playback dataset. For the transect dataset, the data points represent each detected duet on each recorder. Data from the recorder with the highest SNR for each duet is plotted as distance 0, and points for distances at 200 m, 400 m, etc. represent the same duets detected on increasingly distant recorders. The first panel of boxplots contains duets with a maximum quality score of 1 or 2 at distance 0, the second panel contains duets with a maximum distance score or 3, and the third panel contains boxplots with a maximum quality score of 4. The boxplot's upper and lower boundary represent the first and third quartiles of the SNR and the bold line within the box represents the median SNR of the data points for that recorder. For the home range and playback datasets, the data points represent each duet we found in the audio data. The gray area around the blue regression line demonstrates the 95% confidence interval. For the transect dataset, ARUs were positioned 1.5 to 2 m above ground whereas for the home range and playback datasets the ARUs were located 12 m above ground. Lower SNR values measured at close distance from the source in **(A)**, panel 1 may be explained by the height of the ARUs which were positioned closer to the ground as compared with **(B, C)**.

score categories assigned to the call on the recorder where the SNR was highest, the SNR of that same call on the next closest recorder (200 m away) is dramatically lower, and calls were virtually never detected 400 m away.

The home range dataset (Figure 5B) also documents a similarly rapid decline in SNR with distance. The closest two calls at 38 and 43 m from the ARU have a SNR of 34.4 and 39 dB, respectively. The furthest duet was recorded at 125 m and has a SNR of 10.8 dB. The playback dataset (Figure 5C) shows a similar trend, though the SNR does not decrease as rapidly as in our home range dataset.

3.5 How well do automated detection algorithms perform?

Our observers manually annotated 93 titi monkey loud calls in the transect dataset and classified 83 as duets, nine as burst gobbles, and one solo call. Because duets, solo calls, and burst gobbles likely have different functions (Robinson, 1979; Caselli et al., 2015) and possibly different detection radii, we only use duets in this study. Of the 83 duets we manually annotated in the transect dataset, two were used in training the algorithm, leaving 81 duets in the test dataset for the algorithm to detect. Of these duets, zero were assigned a qualitative quality score of 1, three were scored as quality 2, 14 were scored as quality 3, and 64 were scored as quality 4. Both howler monkey roars and titi duets that occasionally overlapped with each other were found in 19 audio files. Of the 1016 files that did not contain titi duets, 231 contained howler monkey roars.

We set the similarity cut-off score at 4.0 as the detection threshold for a titi monkey duet. Using this threshold, we ran four different detection pipelines using different combinations of the test dataset and binary matching templates. Pipeline 1 utilized the full test dataset (N = 81 duets) with titi duet templates and registered a "positive" detection when the mean titi template score was greater than 4.0. Duets categorized as quality 4 ranged in SNR from 0.1 to 11.2 dB with a mean and SD of 3.8 ± 2.9 dB. Quality 2 and 3 duets ranged in SNR from 5.9 to 27.1 dB with a mean of 12.9 ± 5.6 dB. Of all the true-positive detections, the SNR ranged from 8.5 to 27.1 dB with a mean of 13.2 ± 5.4 dB. The duets that were not detected ranged in SNR from 0.1 to 9.9 dB with a mean of 3.7 ± 2.7 dB. Duets with a SNR > 10 dB were detected correctly 100% (10 out of 10) of the time, whereas only 12.6% of duets with a SNR < 10 dB were detected (9 out of 71). For Pipeline 2, we used a modified dataset that excluded the 64 quality score 4 duets (Figure 3D) as manual detections. This improved the balanced accuracy, recall and F1-score of the algorithm (Table S3, Supplementary Material). In Pipeline 3, we used the full test dataset of 81 duets and additionally included the howler monkey templates in the detection rule such that a positive detection was recorded when the mean titi template score was greater than 4.0 and the mean titi template score exceeded mean howler template score. Pipeline 3 had a much higher precision rate than the first two pipelines, but slightly reduced accuracy and recall. Pipeline 4 used the modified test dataset that excluded quality 4 calls as well as the howler monkey templates. Here, the algorithm detected 12 of 17 duets (70.6%) and recorded 5 false negatives (Table S3, Supplementary Material). The algorithm also successfully distinguished other titi loud calls from duets; the one solo call and nine burst gobbles in our testing data were now correctly identified as negative duet detections. The false positives remained identical to Pipeline 3. Pipeline 4 had a precision of 0.75, a recall of 0.71, and an F1-score of 0.72. More detailed descriptions of the results of all four Pipelines, including confusion matrix results, are available in the Supplementary Material.

4 Discussion

4.1 Study limitations

In this study, we show that PAM in combination with automated detection of duets can be a useful method of data collection for duetting primates provided that appropriate attention is paid to understanding particulars of the system under investigation, e.g., about source level and attenuation with distance and about how other animals' vocalizations might interfere with successful detection. We determined that the source level of titi duets is ~105 dB re 20 µPa at 1m with an attenuation rate of 8 dB per doubling of distance. The detection radius for manual annotation of duets in audio recordings using our ARU setup is at least 200m but is much lower for the automated detection algorithm. This result is explained by the poor detectability of duets with a low SNR. Algorithm recall for duets with an SNR > 10 dB was 100%, corresponding to a detection radius of ~125m. The SNR of most duets drops below 10 dB for distances beyond 125 m and leads to poor recall. These results highlight important considerations to be made in PAM research design when choosing to analyze data using manual annotation versus automated detection of calls of interest.

Our study has three main limitations, however. First, we used ARUs, Wildlife Acoustics' Song Meter SM2 and SM2+, which are no longer commercially available. Though this is a problem that is likely to occur with many recorders over time as technological advancements are made, it does make replicability of studies and generalization of results more difficult due to differences in recorder and microphone specifications. Second, some of our datasets are rather small. For example, our observational data on home range use and vocal activity consists of data from only one focal pair of titi monkeys. We detected 13 duets on nine of 34 observation days, which puts the singing rate of this group at once every 3 to 4 days. The small dataset in combination with the infrequent nature of duets in this species means that our results cannot be extrapolated with confidence to other populations or groups of *Plecturocebus discolor*, because it does not consider intergroup variation or potential changes in singing rate throughout the year (Dolotovskaya and Heymann, 2022). On the other hand, our transect dataset were collected across a much larger geographic area (roughly 2 km^2), and with titi home range size estimates from the TBS region ranging from 4.1 ha or 0.041 km^2 (Van Belle et al., 2021) to 6.1 ha or 0.061km^2 (Dacier et al., 2011), this area should contain an estimated 33 to 49 titi monkey pairs. Undoubtedly, a similar study with an extended temporal or spatial scale could lead to more robust results, especially considering the infrequent nature of duet calls in this region. Third, the trails along which we conducted our playback experiments were not long enough to document when the detection probability of duets decreased as calls were always detectable even at the largest playback distance of 200 m. Therefore, the full extent to which distance affects the detectability of duets in recordings relying either on manual annotation or automated detection remains an empirical question, though our results suggest a minimum detection radius of 125 to 200 m.

4.2 Passive acoustic monitoring

Using 12 naturally occurring duets recorded within the home range that were picked up by two ARUs, we estimated that the source level of the titi duets is ~92 dB re 20 µPa at 1 m when assuming only spherical spreading and ~105 dB re 20 µPa at 1 m when we correct for excess attenuation. The source level of titi monkey duets is similar to the estimated source levels of other primate loud calls such as howler monkey roars and gibbon great calls (Sekulic, 1983; Whitehead, 1995; Terleph et al., 2016). However, detailed comparisons are complicated by inconsistencies in the reporting of excess attenuation levels (Table S1, Supplementary Material). For example, sound pressure levels for howler monkeys have been reported as 90 dB at 5 m (Whitehead, 1995) and 70 dB at 50 m (Sekulic, 1983). Using the inverse square law, both estimates would result in a source level of 104 dB at 1 m, but this does not account for excess attenuation. This means that when the attenuation rate (the combination of both spherical spreading and excess attenuation) of a call of interest in a specific environment is unknown, source levels are likely to be underestimated. This can lead to incorrect assumptions on a call's detection radius and detection probability. However, it bears noting that excess attenuation does not always lead to an increased decay in SPL. In some cases, reflection of sound waves can cause constructive interference, leading to less attenuation than predicted even through spherical spreading alone (Hedwig et al., 2018). Here, the attenuation rate of duets was estimated to be 8 dB per doubling of distance in a primary tropical rainforest: 6 dB due to spherical spreading plus an additional 2 dB of excess attenuation per doubling of distance. These results are supported by similar findings for comparable habitats. Ellinger and Hödl (2003) estimated excess attenuation to be 10 dB at 50 m in lowland rainforest in Venezuela and Waser and Brown (1986) estimated excess attenuation at evergreen rainforests in Kenya and Uganda to be ~7 dB at 50 m, both slightly lower than our estimate of 13 dB at 50 m (Figure 4). Excess attenuation levels vary depending on variables such as call frequency, height above the ground, and time of day (Waser and Brown, 1986; Ellinger and Hödl, 2003; Sugiura et al., 2006).

The duet furthest from the ARU in our home range dataset was 125 m away. This limits any conclusions for an estimated detection radius and probability to this distance. When analyzed manually by observers, all duets were easily visible in spectrograms and no duets were missed, placing the detection radius at 125 m with a detection probability of 1. In our playback dataset, duets were also consistently detected by human observers at all tested distances up to the maximum of 200 m with a detection probability of 1, suggesting the actual detection radius for the Song Meter SM2 ARUs when relying on manual annotation by human observers is greater than 200 m. We did not conduct our playbacks at distances beyond 200 m because of strong changes in the slope of the terrain that we suspected would have significantly altered the habitat acoustics such that data beyond 200 m would have been unreliable. However, a recent study shows that slope does not affect the ability of ARUs to detect signals of interest (Shaw et al., 2022). Instead, the orientation of microphones and vocalizing primates plays a more important role in ARU detection radius (Shaw et al., 2022). Indeed, our playback dataset may slightly overestimate the detection radius because we always oriented the speaker directly towards the ARU when conducting playbacks, while this would not necessarily be the case for naturally occurring duets, thereby influencing the likelihood of an ARU picking up the calls. This becomes apparent when we compare confidence intervals around the regression lines in Figures 5B, C; the range of variation in SNR of naturally occurring titi duets in the home range dataset leads to a larger confidence interval around the relationship between SNR and distance, whereas playback dataset show less variation in SNR data due to the consistent orientation of the speaker, leading to a much narrower confidence interval around the inferred relationship. Combining our different datasets and taking into account their limitations, the detection radius within which the detection probability is close to 1 when audio data are verified by human observers is ~200 m.

4.3 Automated detection

Overall, our automated detection protocol frequently missed duets with a SNR below 10 dB, but performed relatively well at identifying calls with a higher SNR, which is perhaps not surprising. Difficulties with detecting low SNR signals is a common limitation of automated detection (Spillmann et al., 2017), particularly in complex environments like rainforests. This outcome is to be expected as the power variations that the algorithm looks for may be barely distinguishable from ambient noise. To increase the odds that a call has a SNR sufficiently high to be detected, several approaches could be taken. When using an array of recorders, narrowing the distance between recorders increases the odds that a singing group of primates is closer to a recorder. However, this can significantly increase the cost of a project as well as amount of data collected that needs to be analyzed, and it can be logistically challenging to manage large numbers of recorders. Instead (though not explored in this paper) noise reduction techniques such as noise spectral subtraction method (Bayestehtashk et al., 2014) or adaptive level equalization (Towsey, 2013) might increase the SNR of calls. Though noise reduction techniques can enhance the signal of interest, they can also result in a loss of detail in the signal (Towsey et al., 2014).

In our initial run of the binary point matching algorithm on 1099 test data files, we identified 29 false positives, yielding low precision and recall values. High rates of false positives are common when the focus of the detection algorithm is on minimizing false negatives, in noisy environments, and when aiming to detect complex vocalizations (Marques et al., 2009; Swiston and Mennill, 2009; Heinicke et al., 2015; Bobay et al., 2018). The number of false positives seen in any given dataset is also determined by the relative abundance of the species contributing to false positive detections, which can change across seasons. The rate of false positives can partially be mitigated by adjusting the detection threshold, but this comes at the cost of increasing the number of false negatives. In our dataset, most false positives seen in initial runs of our algorithm were caused by howler monkey roars. Howler roars and titi duets overlap significantly in frequency range, and the complexity of the duets and noisiness of the howler roars makes it difficult for the detection algorithm to differentiate between the two. However, we were able to mitigate this problem by creating an additional set of howler monkey templates that we applied to each file and then using a detection rule that considers the similarity score of both titi and howler templates when deciding whether or not a titi duet is detected. This change in the detection rule successfully allowed us to remove as positive detections all howler monkey roars, but also resulted in the loss of a few titi detections. This typically happened when both species vocalized within the same 10-minute audio file. Whether this is problematic or not heavily depends on the study design and research question. Our human observers misclassified five burst gobbles as duets, demonstrating that false positives are not unique to automated detection algorithms. However, the rate of false positives is generally much lower for manual annotation than automated detection (Swiston and Mennill, 2009). Depending on the nature of the study, if the rate of false positives is less of a concern than the rate of false negatives, the choice could be made to manually validate all detections of an algorithm (Knight et al., 2020). In monitoR this can be done with the function ‘showPeaks’.

By creating templates for non-titi vocalizations that were found, in initial runs, to cause large numbers of false-positive detections and then including scores on those “off-target” templates as part of a more complex detection rule, we greatly decreased the number of false positives and thereby increased the precision rates of the algorithm. In addition, recall (the true positive rate) was largely influenced by the quality of the calls in the data. Large numbers of distant calls with a low SNR are difficult to detect by algorithms as they do not stand out as clearly against background noise. When the expectation is that these calls should be detected by the algorithm, the outcomes are rather poor (e.g., algorithm iterations 1 and 3). However, when we focus only on calls within a smaller detection radius around the recorder (algorithm iterations 2 and 4), the algorithm performs considerably better. This result suggests that automated detection algorithms can and should be tailored to each specific species and research question. Among others, differences in call structures, habitat structure, number of other vocally active animals in the habitat, algorithm type, and study goals determine the required specifications of the algorithm, and thus its outcomes.

The need for detection algorithm parameters and decision rules to be tailored to specific sites and call types also creates challenges for comparing results between studies. The fourth iteration of our algorithm had a precision of 0.67. When we compare this to other studies that have used various automated detection procedures to locate primate vocalizations within audio files, we find that our precision score is lower than found in studies of Guianan red howler monkeys (*Alouatta macconnelli*), black-fronted titi monkeys (*Callicebus nigrifrons*), Hainan gibbons (*Nomascus hainanus*), and northern grey gibbons (*Hylobates funereus*) (Versteegh et al., 2016; Do Nascimento et al., 2021; Clink et al., 2023; Ravaglia et al., 2023), but higher than found for several African primates (Heinicke et al., 2015). Our recall of 0.71 is similar to that found for northern grey gibbons (Clink et al., 2023), higher than that of Guianan red howler monkeys and African primates (Heinicke et al., 2015; Do Nascimento et al., 2021), and lower than recall values for black-fronted titi monkeys, Hainan gibbons, and indris (Versteegh et al., 2016; Dufourq et al., 2021; Ravaglia et al., 2023), though some of these more recent studies with high recall used a deep learning approach. However, it is difficult to compare the performance of the algorithms using these metrics as some differences cannot easily be quantified (e.g., vocalization and habitat structure, the number of vocally active animals in an environment that can trigger false positives). In addition, other important details are often unknown. For example, if the calls in the test dataset are predominantly calls with a high SNR, algorithm performance will be high, but when calls with lower SNRs are also included, performance of the algorithm may drop significantly. Clearly, understanding the quality of the calls in the test dataset is important, as it allows us to modify the design of future studies.

4.4 Recommendations for studying duetting with PAM

In this study, we demonstrate that passive acoustic monitoring together with an automated detection algorithm can be an effective method to study duetting behavior in titi monkeys and potentially other duetting animals. However, due to the complex temporal and spectral structures of duets, a series of steps need to be taken to ensure a study design that fits the desired outcome. Here, we describe such a workflow.

First, as recorder specifications and both structure and source level of duets greatly affect the detection radius, empirical determination of the detection radius and detection probability for the combination of the chosen recorder type and vocalizations of interest is a critical first step. Because habitat characteristics also play a role in the detection radius of duets, detection radius should be determined for each site at which the methodology is implemented. In this study, we used two ways of determining the detection radius of an ARU model for a vocalization of interest. If the source level of the primate's loud call is known or can be determined, playback studies can be used to determine the detection radius and detection probabilities. If the source level cannot be determined, using one or multiple ARUs in primate home ranges can be used in combination with location data on loud calls from behavioral follows of habituated groups. However, this alternative is less time efficient than playback experiments.

Second, a trade-off must be made between choosing whether to manually annotate the collected audio data with the use of spectrograms, or to use an automated detection algorithm to locate the duets in the audio data. This will also influence evaluation of the detection radius of the recorder, as we demonstrated that duets can typically be annotated at larger distances by human observers than by the detection algorithm. When ARUs are used on a small scale, manual annotation of duets is feasible. The potential for significant numbers of false positive and false negative detections means that automated detection algorithms come with their own set of errors and biases, but with the benefit that these biases are measurable and more consistent than observer errors and biases.

Third, if the amount of audio data collected for the purposes of the study is too large to allow manual annotation within a reasonable time frame, a mix of both techniques can be used if high accuracy is desired. If a detection algorithm is designed to reduce the number of false negatives, the resulting large number of positive detections – which will be a combination of true positives and false positives – can still be manually verified and corrected. It should be kept in mind that when the animal of interest duets only infrequently, the number of false positives will likely increase at a much higher rate than the number true positives. However, although a high false positive rate means that more putative detections need to be verified, it does not negatively affect the performance of the algorithm on detecting true positives. When the animals under study vocalize infrequently, a deep learning approach might yield benefits over template matching as it allows for data augmentation (i.e., artificially increasing the training data by creating modified duets from existing ones) when training the algorithm (for example studies, see Table S2, Supplementary Material). Deep learning commonly requires more expert knowledge to implement, though alternatives geared towards increasing accessibility for less experienced scientists are in development (Arthur et al., 2021).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Annotated code used in the manuscript is available on https://github.com/SoundScience/Automated-detection-and-detection-range-of-titi-monkey-duets.

Ethics statement

Approval for JB's research was obtained from the Institutional Animal Care and Use Committee (IACUC), University of Florida Non-Regulatory Animal Research Committee (#201710065). Approval for research by SvK and AD was obtained from the

IACUC, The University of Texas at Austin Animal Research Committee (AUP-2017-00260). Work at the Tiputini Biodiversity Station was conducted in accordance with research permit number 025-2019-IC-PNY-DPAO and earlier ones (JB) and MAE-DNB-CM-2015-0009-M-0001 (AD), Ministerio del Ambiente, Puerto Francisco de Orellana.

Author contributions

SvK designed the research with guidance from AD. JB and SvK collected data, SO and SvK created templates and tested the algorithms. SvK and SO conducted the analyses with guidance and statistical programming support from AD and DC. SvK and AD led the writing of the manuscript, and all authors contributed to the article and approved the submitted version.

Funding

This research was supported by the University of Florida and The University of Texas at Austin.

Acknowledgments

We extend our sincere gratitude to Patrice Adret for his insightful feedback and conversations. His continued support and encouragement were pivotal in the development of this paper. We would like to thank Ministerio del Ambiente de Ecuador (MAE), the Universidad San Francisco de Quito, and the Tiputini Biodiversity Station and its staff, especially Consuelo Barriga de Romo and Tomi Sugahara, for their assistance with planning, logistics and permits. We thank Anna Kurtin and Kaleb Judd for helping with manual annotation of the duets for the playback and transect datasets. We are also grateful to Shawn McCracken for providing the Song Meter SM2 recorders for the summer of 2016. Finally, we would also like to thank editor Ulrich H. Reichard and two reviewers for their valuable feedback that helped improve this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2023.1173722/full#supplementary-material

References

Adret, P., Dingess, K. A., Caselli, C. B., Vermeer, J., Martínez, J., Luna Amancio, J. C., et al. (2018). Duetting patterns of titi monkeys (Primates, Pitheciidae: Callicebinae) and relationships with phylogeny. *Animals* 8, 178. doi: 10.3390/ani8100178

Aldrich, B. C. (2006). A song-based survey of the Andean titi monkey (*Callicebus oenanthe*) at Tarangue, with notes on its vocalisations (Oxford: Oxford Brookes University). MSc thesis.

Arthur, B. J., Ding, Y., Sosale, M., Khalif, F., Kim, E., Waddell, P., et al. (2021). SongExplorer: A deep learning workflow for discovery and segmentation of animal acoustic communication signals. *bioRxiv*. doi: 10.1101/2021.03.26.437280

Barclay, R. M. (1999). Bats are not birds—a cautionary note on using echolocation calls to identify bats: a comment. *J. Mammal.* 80, 290–296. doi: 10.2307/1383229

Bayestehtashk, A., Shafran, I., Coleman, K., and Robertson, N. (2014). "Detecting vocalizations of individual monkeys in social groups," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL. 4775–4779 (IEEE).

Bishop, C. M. (2006). *Neural networks for pattern recognition* (New York, NY: Springer).

Blake, J. G. (1992). Temporal variation in point counts of birds in a lowland wet forest in Costa Rica. *Condor* 94, 265.275. doi: 10.2307/1368816

Blake, J. G. (2021). Acoustic monitors and direct observations provide similar but distinct perspectives on bird assemblages in a lowland forest of eastern Ecuador. *PeerJ* 9, e10565. doi: 10.7717/peerj.10565

Bobay, L. R., Taillie, P. J., and Moorman, C. E. (2018). Use of autonomous recording units increased detection of a secretive marsh bird. *J. Field Ornithol.* 89, 384–392. doi: 10.1111/jofo.12274

Bradbury, J. W., and Vehrencamp, S. L. (1998). *Principles of animal communication* (Sunderland, MA: Sinauer Associates Inc).

Brown, C. H., and Waser, P. M. (2017). "Primate habitat acoustics," in *Primate hearing and communication*. Eds. R. M. Quam, M. A. Ramsier, R. R. Fay and A. N. Popper (Cham, Switzerland: Springer International Publishing), 79–107. doi: 10.1007/978-3-319-59478-1_4

Caselli, C. B., Mennill, D. J., Bicca-Marques, J. C., and Setz, E. Z. (2014). Vocal behavior of black-fronted titi monkeys (*Callicebus nigrifrons*): Acoustic properties and behavioral contexts of loud calls. *Am. J. Primatol.* 76, 788–800. doi: 10.1002/ajp.22270

Caselli, C. B., Mennill, D. J., Gestich, C. C., Setz, E. Z., and Bicca-Marques, J. C. (2015). Playback responses of socially monogamous black-fronted titi monkeys to simulated solitary and paired intruders. *Am. J. Primatol.* 77, 1135–1142. doi: 10.1002/ajp.22447

Clink, D. J., Ahmad, A. H., and Klinck, H. (2020). Brevity is not a universal in animal communication: evidence for compression depends on the unit of analysis in small ape vocalizations. *R. Soc. Open Sci.* 7, 200151. doi: 10.1098/rsos.200151

Clink, D. J., Comella, I., Ahmad, A. H., and Klinck, H. (2023). A workflow for the automated detection and classification of female gibbon calls from long-term acoustic recordings. *Front. Ecol. Evol.* 11. doi: 10.3389/fevo.2023.1071640

Dacier, A., de Luna, A. G., Fernandez-Duque, E., and Di Fiore, A. (2011). Estimating population density of Amazonian titi monkeys (*Callicebus discolor*) via playback point counts. *Biotropica* 43, 135–140. doi: 10.1111/j.1744-7429.2010.00749.x

Darras, K. F., Deppe, F., Fabian, Y., Kartono, A. P., Angulo, A., Kolbrek, B., et al. (2020). High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ* 8, e9955. doi: 10.7717/peerj.9955

Darras, K., Pütz, P., Rembold, K., and Tscharntke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biol. Conserv.* 201, 29–37. doi: 10.1016/j.biocon.2016.06.021

Deichmann, J. L., Acevedo-Charry, O., Barclay, L., Burivalova, Z., Campos-Cerqueira, M., d'Horta, F., et al. (2018). It's time to listen: there is much to be learned from the sounds of tropical ecosystems. *Biotropica* 50, 713–718. doi: 10.1111/btp.12593

De Luna, A. G., Sanmiguel, R., Di Fiore, A., and Fernandez-Duque, E. (2010). Predation and predation attempts on red titi monkeys (*Callicebus discolor*) and equatorial sakis (*Pithecia aequatorialis*) in Amazonian Ecuador. *Folia Primatol.* 81, 86–95. doi: 10.1159/000314948

Digby, A., Towsey, M., Bell, B. D., and Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods Ecol. Evol.* 4, 675–683. doi: 10.1111/2041-210X.12060

Dolotovskaya, S., and Heymann, E. W. (2022). Coordinated singing in coppery titi monkeys (*Plecturocebus cupreus*): Resource or Mate Defense? *Front. Ecol. Evol.* 457. doi: 10.3389/fevo.2022.898509

Do Nascimento, L. A., Pérez-Granados, C., and Beard, K. H. (2021). Passive acoustic monitoring and automatic detection of diel patterns and acoustic structure of howler monkey roars. *Diversity* 13, 566. doi: 10.3390/d13110566

Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., et al. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sens. Ecol. Conserv.* 7, 475–487. doi: 10.1002/rse2.201

Eddy, S. R. (2004). What is a hidden Markov model? *Nat. Biotechnol.* 22, 1315–1316. doi: 10.1038/nbt1004-1315

Ellinger, N., and Hödl, W. (2003). Habitat acoustics of a neotropical lowland rainforest. *Bioacoustics* 13, 297–321. doi: 10.1080/09524622.2003.9753503

Fan, P. F., Xiao, W. E. N., Huo, S., and Jiang, X. L. (2009). Singing behavior and singing functions of black-crested gibbons (*Nomascus concolor jingdongensis*) at Mt. Wuliang, Central Yunnan, China. *Am. J. Primatol.* 71, 539–547. doi: 10.1002/ajp.20686

Flach, P. (2019). "Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward," in *Proceedings of the AAAI conference on artificial intelligence*, Honolulu, Hawaii. 9808–9814.

Geissmann, T. (1999). Duet songs of the siamang, *Hylobates syndactylus*: II. Testing the pair-bonding hypothesis during a partner exchange. *Behaviour* 136, 1005–1039. doi: 10.1163/156853999501694

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* 77, 57–76. doi: 10.1017/s1464793101005826

Gibb, R., Browning, E., Glover-Kapfer, P., and Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10 (2), 169–185. doi: 10.1111/2041-210X.13101

Groves, C., and Shekelle, M. (2010). The genera and species of Tarsiidae. *Int. J. Primatol.* 31, 1071–1082. doi: 10.1007/s10764-010-9443-1

Hall, M. L. (2009). A review of vocal duetting in birds. *Adv. Study Behav.* 40, 67–121. doi: 10.1016/S0065-3454(09)40003-2

Hedwig, D., DeBellis, M., and Wrege, P. H. (2018). Not so far: attenuation of low-frequency vocalizations in a rainforest environment suggests limited acoustic mediation of social interaction in African forest elephants. *Behav. Ecol. sociobiol.* 72, 1–11. doi: 10.1007/s00265-018-2451-4

Heinicke, S., Kalan, A. K., Wagner, O. J., Mundry, R., Lukashevich, H., and Kühl, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods Ecol. Evol.* 6, 753–763. doi: 10.1111/2041-210X.12384

Hilty, S. L. (2003). *Birds of Venezuela. 2nd ed.* (New Jersey: Princeton University Press).

Janvier, M., Horaud, R., Girin, L., Berthommier, F., Boë, L. J., Kemp, C., et al. (2013). "Supervised classification of baboon vocalizations," in *Neural Information Processing Scaled for Bioacoustics: NIPS4B*, Lake Tahoe: NV.

Katz, J., Hafner, S. D., and Donovan, T. (2016). Tools for automated acoustic monitoring within the R package monitoR. *Bioacoustics* 25, 197–210. doi: 10.1080/09524622.2016.1138415

Kinzey, W. G., and Becker, M. (1983). Activity pattern of the masked titi monkey, *Callicebus personatus*. *Primates* 24, 337–343. doi: 10.1007/BF02381979

K. Lisa Yang Center for Conservation Bioacoustics. (2022). *Signal-to-noise (SNR)–User protocol*. Available at: https://ravensoundsoftware.com/knowledge-base/signal-to-noise-ratio-snr (Accessed May 23, 2023).

Knight, E. C., Sòlymos, P., Scott, C., and Bayne, E. M. (2020). Validation prediction: a flexible protocol to increase efficiency of automated acoustic processing for wildlife research. *Ecol. Appl.* 30, e02140. doi: 10.1002/eap.2140

Koloff, J., and Mennill, D. J. (2013). The responses of duetting antbirds to stereo duet playback provide support for the joint territory defence hypothesis. *Ethology* 119, 462–471. doi: 10.1111/eth.12084

Krogh, A. (2008). What are artificial neural networks? *Nat. Biotechnol.* 26, 195–197. doi: 10.1038/nbt1386

Lawson, J., Rizos, G., Jasinghe, D., Whitworth, A., Schuller, B., and Banks-Leite, C. (2023). Automated acoustic detection of Geoffroy's spider monkey highlights tipping points of human disturbance. *Proc. R. Soc. B* 290, 20222473. doi: 10.1098/rspb.2022.2473

Marques, T. A., Thomas, L., Ward, J., DiMarzio, N., and Tyack, P. L. (2009). Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales. *J. Acoustical Soc. America* 125, 1982–1994. doi: 10.1121/1.3089590

Méndez-Cárdenas, M. G., and Zimmermann, E. (2009). Duetting—A mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi*)? *Am. J. Phys. Anthropol.* 139, 523–532. doi: 10.1002/ajpa.21017

Merchant, N. D., Fristrup, K. M., Johnson, M. P., Tyack, P. L., Witt, M. J., Blondel, P., et al. (2015). Measuring acoustic habitats. *Methods Ecol. Evol.* 6, 257–265. doi: 10.1111/2041-210X.12330

Mitani, J. C., and Stuht, J. (1998). The evolution of nonhuman primate loud calls: acoustic adaptation for long-distance transmission. *Primates* 39, 171–182. doi: 10.1007/BF02557729

Müller, A. E., and Anzenberger, G. (2002). Duetting in the titi monkey *Callicebus cupreus*: Structure, pair specificity and development of duets. *Folia Primatol.* 73, 104–115. doi: 10.1159/000064788

Nemeth, E., Dabelsteen, T., Pedersen, S. B., and Winkler, H. (2006). Rainforests as concert halls for birds: are reverberations improving sound transmission of long song elements? *J. Acoustical Soc. America* 119, 620–626. doi: 10.1121/1.2139072

Noble, W. S. (2006). What is a support vector machine? *Nat. Biotechnol.* 24, 1565–1567. doi: 10.1038/nbt1206-1565

Novaković, J. D., Veljović, A., Ilić, S. S., Papić, Ž., and Tomović, M. (2017). Evaluation of classification models in machine learning. *Theory Appl. Math. Comput. Sci.* 7, 39.

Padgham, M. (2004). Reverberation and frequency attenuation in forests—implications for acoustic communication in animals. *J. Acoustical Soc. America* 115, 402–410. doi: 10.1121/1.1629304

Pérez-Granados, C., Bota, G., Giralt, D., Albarracín, J., and Traba, J. (2019). Cost-effectiveness assessment of five audio recording systems for wildlife monitoring: differences between recording distances and singing direction. *Ardeola* 66, 311–325. doi: 10.13157/arla.66.2.2019.ra4

Pinto, L. P., Barnett, A. A., Bezerra, B. M., Boubli, J. P., Bowler, M., Cardoso, N. D. A., et al. (2013). "Why we know so little: the challenges of fieldwork on the Pitheciids," in *Evolutionary biology and conservation of titis, sakis and uacaris*. Eds. L. M. Vega, A. A. Barnett, S. F. Ferrari and M. A. Norconk (Cambridge: Cambridge University Pres), 145–150. doi: 10.1017/CBO9781139034210.019

Pollock, J. I. (1986). The song of the indris (*Indri indri*; Primates: Lemuroidea): Natural history, form, and function. *Int. J. Primatol.* 7, 225–264. doi: 10.1007/BF02736391

Porcaro, L. (2015).). Modeling Lemur vocalizations from a signal processing perspective (Torino: University of Torino). MSc thesis.

Pozzi, L., Gamba, M., and Giacoma, C. (2012). "Artificial neural networks: A new tool for studying lemur vocal communication," in *Leaping Ahead*. Eds. M. Masters, M. Gamba and F. Génin (New York, NY: Springer), 305–313. doi: 10.1007/978-1-4614-4511-1_34

Ravaglia, D., Ferrario, V., De Gregorio, C., Carugati, F., Raimondi, T., Cristiano, W., et al. (2023). There you are! Automated detection of indris' songs on features extracted from passive acoustic recordings. *Animals* 13, 241. doi: 10.3390/ani13020241

R Core Team (2022) *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org (Accessed May 24, 2023).

Rempel, R. S., Francis, C. M., Robinson, J. N., and Campbell, M. (2013). Comparison of audio recording system performance for detecting and monitoring songbirds. *J. Field Ornithol.* 84, 86–97. doi: 10.1111/jofo.12008

Robinson, J. G. (1979). An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Z. für Tierpsychol.* 49, 381–405. doi: 10.1111/j.1439-0310.1979.tb00300.x

Robinson, J. G. (1981). Vocal regulation of inter-and intragroup spacing during boundary encounters in the titi monkey, *Callicebus moloch*. *Primates* 22, 161–172. doi: 10.1007/BF02382607

Sangchantr, S. (2004). *Social organization and ecology of Mentawai leaf monkeys (Presbytis potenziani)* (PhD dissertation. New York: Columbia University).

Schroeder, K. M., and McRae, S. B. (2020). Automated auditory detection of a rare, secretive marsh bird with infrequent and acoustically indistinct vocalizations. *Ibis* 162, 1033–1046. doi: 10.1111/ibi.12805

Sekulic, R. (1983). The effect of female call on male howling in red howler monkeys (*Alouatta seniculus*). *Int. J. Primatol.* 4, 291–305. doi: 10.1007/BF02735551

Shaw, T., Müller, S., and Scherer-Lorenzen, M. (2022). Slope does not affect autonomous recorder detection shape: Considerations for acoustic monitoring in forested landscapes. *Bioacoustics* 31, 261–282. doi: 10.1080/09524622.2021.1925590

Slater, P. J., and Mann, N. I. (2004). Why do the females of many bird species sing in the tropics? *J. Avian Biol.* 35, 289–294. doi: 10.1111/j.0908-8857.2004.03392.x

Souza-Alves, J. P., and Ferrari, S. F. (2010). Responses of wild titi monkeys, Callicebus coimbrai (Primates: Platyrrhini: Pitheciidae), to the habituation process. *Zool. (Curitiba)* 27, 861–866. doi: 10.1590/S1984-46702010000600004

Spillmann, B., van Schaik, C. P., Setia, T. M., and Sadjadi, S. O. (2017). Who shall I say is calling? Validation of a caller recognition procedure in Bornean flanged male orangutan (*Pongo pygmaeus wurmbii*) long calls. *Bioacoustics* 26, 109–120. doi: 10.1080/09524622.2016.1216802

Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* 10, e13152. doi: 10.7717/peerj.13152

Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W. Jr., and Llusia, D. (2019). Terrestrial passive acoustic monitoring: review and perspectives. *BioScience* 69, 15–25. doi: 10.1093/biosci/biy147

Sugiura, H., Tanaka, T., and Masataka, N. (2006). Sound transmission in the habitats of Japanese macaques and its possible effect on population differences in coo calls. *Behaviour* 143, 993–1012. doi: 10.1163/156853906778623617

Swiston, K. A., and Mennill, D. J. (2009). Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *J. Field Ornithol.* 80, 42–50. doi: 10.1111/j.1557-9263.2009.00204.x

Szymański, P., Olszowiak, K., Wheeldon, A., Budka, M., and Osiejuk, T. S. (2021). Passive acoustic monitoring gives new insight into year-round duetting behaviour of a tropical songbird. *Ecol. Indic.* 122, 107271. doi: 10.1016/j.ecolind.2020.107271

Taunk, K., De, S., Verma, S., and Swetapadma, A. (2019). "A brief review of nearest neighbor algorithm for learning and classification," in *International Conference on Intelligent Computing and Control Systems*, Seconderabad, India. 1255–1260, ICCS.

Terleph, T. A., Malaivijitnond, S., and Reichard, U. H. (2016). Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *BMC Evolution. Biol.* 16, 1–13. doi: 10.1186/s12862-015-0578-8

Tilson, R. L., and Tenaza, R. R. (1976). Monogamy and duetting in an Old World monkey. *Nature* 263, 320–321. doi: 10.1038/263320a0

Towsey, M. W. (2013). *Noise removal from wave-forms and spectrograms derived from natural recordings of the environment* (Brisbane: Queensland University of Technology).

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., and Roe, P. (2014). Visualization of long-duration acoustic recordings of the environment. *Proc. Comput. Sci.* 29, 703–712. doi: 10.1016/j.procs.2014.05.063

Van Belle, S., Porter, A., Fernandez-Duque, E., and Di Fiore, A. (2018). Ranging behavior and potential for territoriality in equatorial sakis (*Pithecia aequatorialis*) in Amazonian Ecuador. *Am. J. Phys. Anthropol.* 167, 701–712. doi: 10.1002/ajpa.23645

Van Belle, S., Porter, A. M., Fernandez-Duque, E., and Di Fiore, A. (2021). Ranging behavior and the potential for territoriality in pair-living titi monkeys (*Plecturocebus discolor*). *Am. J. Primatol.* 83, e23225. doi: 10.1002/ajp.23225

Van Kuijk, S. (2013). Living on the edge: Critically endangered San Martin titi monkeys (*Callicebus oenanthe*) show preference for forest edges in C.C. Ojos de Agua, Peru. MSc thesis: Oxford Brookes University.

Vermeer, J., and Tello-Alvarado, J. C. (2015). The distribution and taxonomy of titi monkeys (*Callicebus*) in central and southern Peru, with the description of a new species. *Primate Conserv.* 29, 1–21. doi: 10.1896/052.029.0102

Versteegh, M., Kuhn, J., Synnaeve, G., Ravaux, L., Chemla, E., Cäsar, C., et al. (2016). Classification and automatic transcription of primate calls. *J. Acoustical Soc. America* 140, EL26–EL30. doi: 10.1121/1.4954887

Waser, P. M., and Brown, C. H. (1984). Is there a "sound window" for primate communication? *Behav. Ecol. Sociobiol.* 15, 73–76. doi: 10.1007/BF00310219

Waser, P. M., and Brown, C. H. (1986). Habitat acoustics and primate communication. *Am. J. Primatol.* 10, 135–154. doi: 10.1002/ajp.1350100205

Waser, P. M., and Waser, M. S. (1977). Experimental studies of primate vocalization: Specializations for long-distance propagation. *Z. für Tierpsychol.* 43, 239–263. doi: 10.1111/j.1439-0310.1977.tb00073.x

Whitehead, J. M. (1995). Vox Alouattinae: a preliminary survey of the acoustic characteristics of long-distance calls of howling monkeys. *Int. J. Primatol.* 16, 121–144. doi: 10.1007/BF02700156

Zambolli, A. H., Manzano, M. C. R., Honda, L. K., Rezende, G. C., and Culot, L. (2023). Performance of autonomous recorders to detect a cryptic and endangered primate species, the black lion-tamarin (*Leontopithecus chrysopygus*). *Am. J. Primatol.* 85, e23454. doi: 10.1002/ajp.23454

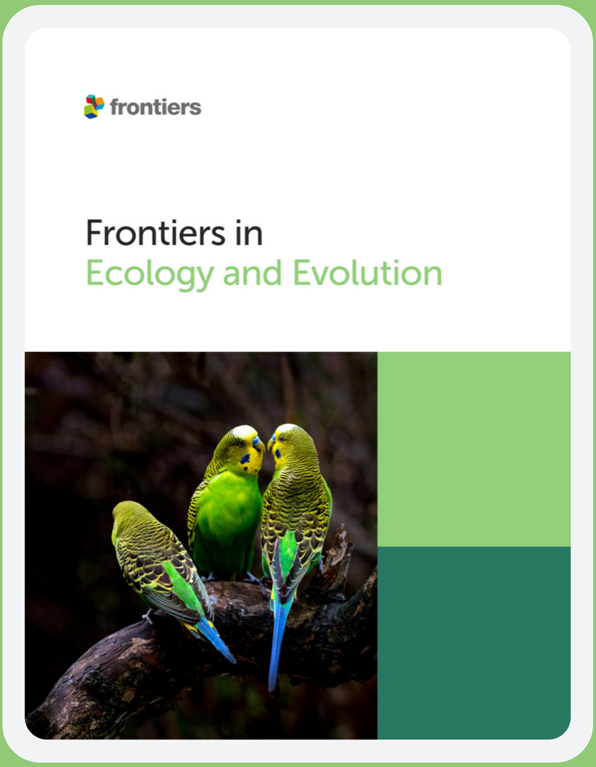
Frontiers in
Ecology and Evolution

Ecological and evolutionary research into our natural and anthropogenic world

This multidisciplinary journal covers the spectrum of ecological and evolutionary inquiry. It provides insights into our natural and anthropogenic world, and how it can best be managed.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

frontiers | Research Topics