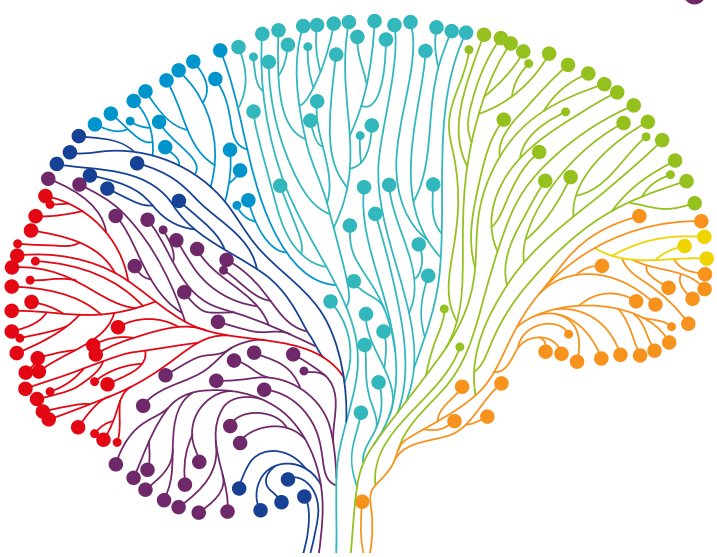
ACOUSTIC COMMUNICATION ANALYSIS FOR UNDERSTANDING ANIMAL BEHAVIOR

EDITED BY: Sara Anna Bonini and Susanna Pietropaolo

PUBLISHED IN: Frontiers in Behavioral Neuroscience



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-946-9
DOI 10.3389/978-2-88976-946-9

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

ACOUSTIC COMMUNICATION ANALYSIS FOR UNDERSTANDING ANIMAL BEHAVIOR

Topic Editors:
Sara Anna Bonini, University of Brescia, Italy
Susanna Pietropaolo, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Citation: Bonini, S. A., Pietropaolo, S., eds. (2022). Acoustic Communication Analysis for Understanding Animal Behavior. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-946-9

Table of Contents

04 ***Editorial: Acoustic Communication Analysis for Understanding Animal Behavior***
Sara Anna Bonini and Susanna Pietropaolo

07 ***LMT USV Toolbox, a Novel Methodological Approach to Place Mouse Ultrasonic Vocalizations in Their Behavioral Contexts—A Study in Female and Male C57BL/6J Mice and in* Shank3 *Mutant Females***
Fabrice de Chaumont, Nathalie Lemière, Sabrina Coqueran, Thomas Bourgeron and Elodie Ey

25 ***Toward a Computational Neuroethology of Vocal Communication: From Bioacoustics to Neurophysiology, Emerging Tools and Future Directions***
Tim Sainburg and Timothy Q. Gentner

49 ***Maturation of Social-Vocal Communication in Prairie Vole (*Microtus ochrogaster*) Pups***
Megan R. Warren, Drayson Campbell, Amélie M. Borie, Charles L. Ford IV, Ammar M. Dharani, Larry J. Young and Robert C. Liu

64 ***Response Calls Evoked by Playback of Natural 50-kHz Ultrasonic Vocalizations in Rats***
Annuska C. Berz, Markus Wöhr and Rainer K. W. Schwarting

83 ***HybridMouse: A Hybrid Convolutional-Recurrent Neural Network-Based Model for Identification of Mouse Ultrasonic Vocalizations***
Yizhaq Goussha, Kfir Bar, Shai Netser, Lior Cohen, Yacov Hel-Or and Shlomo Wagner

95 ***Identification, Analysis and Characterization of Base Units of Bird Vocal Communication: The White Spectacled Bulbul (*Pycnonotus xanthopygos*) as a Case Study***
Aya Marck, Yoni Vortman, Oren Kolodny and Yizhar Lavner

109 ***Inhibition of miR-128 Enhances Vocal Sequence Organization in Juvenile Songbirds***
Caitlin M. Aamodt and Stephanie A. White

118 ***Semi-Automated Training of Rat Ultrasonic Vocalizations***
Aaron M. Johnson, Charles Lenell, Elizabeth Severa, Denis Michael Rudisch, Robert A. Morrison and Adrianna C. Shembel

126 ***From Mating to Milk Access: A Review of Reproductive Vocal Communication in Mice***
Sara Capas-Peneda, Yolanda Saavedra Torres, Jan-Bas Prins and I. Anna S. Olsson

136 ***Effects of Congenital Blindness on Ultrasonic Vocalizations and Social Behaviors in the ZRDBA Mouse***
Nouhaila Bouguiyoud, Elena Morales-Grahl, Gilles Bronchti, Johannes Frasnelli, Florence I. Roullet and Syrina Al Aïn

148 ***Ultrasonic Vocalizations in Adult C57BL/6J Mice: The Role of Sex Differences and Repeated Testing***
Marika Premoli, Valeria Petroni, Ronald Bulthuis, Sara Anna Bonini and Susanna Pietropaolo

frontiers | Frontiers in Behavioral Neuroscience

TYPE Editorial
PUBLISHED 03 August 2022
DOI 10.3389/fnbeh.2022.991573

OPEN ACCESS

EDITED AND REVIEWED BY
Walter Adriani,
National Institute of Health (ISS), Italy

*CORRESPONDENCE
Sara Anna Bonini
sara.bonini@unibs.it

SPECIALTY SECTION
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

RECEIVED 11 July 2022
ACCEPTED 21 July 2022
PUBLISHED 03 August 2022

CITATION
Bonini SA and Pietropaolo S (2022) Editorial: Acoustic communication analysis for understanding animal behavior.
Front. Behav. Neurosci. 16:991573.
doi: 10.3389/fnbeh.2022.991573

Editorial: Acoustic communication analysis for understanding animal behavior

Sara Anna Bonini[1]* and Susanna Pietropaolo[2]

[1]Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy, [2]Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), University of Bordeaux, CNRS, UMR 5287, Bordeaux, France

KEYWORDS

acoustic communication, ultrasonic vocalizations, animal behavior, social context, communication analysis, automatic USVs analysis

Editorial on the Research Topic
Acoustic communication analysis for understanding animal behavior

Animals of most species emit vocalizations to communicate each other and to convey their emotional states. Intraspecific communication consists of the transfer of information from one or more animals [sender(s)] to one or more conspecifics [receiver(s)] in order to affect the current or future behaviors of the receiver(s). Vocal communication may occur both in the audible and ultrasonic frequency ranges and it is subjected by a complex brain modulation involving multiple areas and neural circuits. Communication is also impaired in a variety of pathological conditions, such as neurodevelopmental disorders (autism spectrum disorders, ASD, in particular), and these alterations can be tackled by pharmacological therapeutic strategies that have been proposed in several preclinical studies.

One of the most ambitious challenges of animal vocal communication concerns the study of ultrasonic vocalizations (USVs) in rodents and it consists of fully automatizing the procedures for spectrographic USV analysis. This would allow bypassing the operator-dependent and time-consuming analyses that are necessary for the quantitative and qualitative studies of USVs. This challenge was successfully faced by de Chaumont et al. in this special issue, as they developed one open-access software to automatically study mouse USVs. Their system also allowed coupling the spectrographic analysis of USVs with the automatic labeling of several behaviors scored over a long time-period (i.e., 3 days), thus providing an additional evaluation of the behavioral value of USVs in laboratory mice. The method was validated in C57BL/6J mice and also in mutants lacking the Shank3 gene, i.e., a mouse model of autism, in both sexes and at multiple ages. Also Goussha et al. developed a new software, the HybridMouse, that is an audio analysis tool for automatically identifying, labeling, and extracting recorded USVs. The methodological challenges of automatic processing a large amount of data from audio files does not concern only mouse ultrasonic communication. Here Marck et al. developed an automatic system based on audio signal processing algorithms and deep

learning and applied to the vocal repertoire of the White Spectacled Bulbul (*Pycnonotus xanthopygos*), a bird species with a complex vocal communication system. Other modern computational methods for bioacoustics have been developed for the analysis of vocal communication in several animal species (including fish, amphibians, and insects), as described in detail in the review by Sainburg and Gentner.

Beside the analysis of the vocalizations emitted by the sender, the response of the receiving animal to vocalizations deserves to be studied. In rats, two main types of vocalizations are typically distinguished (Brudzynski, 2013; Wöhr and Schwarting, 2013): vocalizations with frequencies around 22 kHz, that are referred to as aversive or distress calls, presumably representing a negative affective state (Blanchard et al., 1991; Fendt et al., 2018), and vocalizations with frequencies around 50 kHz, that are thought to represent a positive affective state usually emitted during appetitive situations like play or mating (Knutson et al., 1998; Panksepp, 2005). In this context, Berz et al. investigated the emission of USVs in response to the playback of natural 50 kHz USVs in male juvenile rats of different strains. Their data demonstrated that most rats emitted response calls specifically linked to 50-kHz USV playback and these response calls were mostly characterized by a frequency range of 20–32 kHz and a mean duration of approximately 300 ms in all rat strains. The possible function of these type of calls has yet to be clarified, although it was unaffected by the pharmacological blockade of dopamine D2 receptors by haloperidol administration. In rats, the emission of USVs can be also promoted by operant conditioning during multiple weeks: this approach was employed in the study by Johnson et al. where a semi-automated method for training rats to increase their rate of USV production was introduced.

The impact of social context on animal communication is an issue of critical relevance that is far from being fully elucidated. Here, Warren et al. studied the ontogeny of prairie vole pup ultrasonic vocalizations when isolated or when the mother was present, but physically unattainable. They demonstrated a developmental maturation in all features of pup vocalizations and the impact of the different social contexts in modifying vocal emission. Capas-Peneda et al. reviewed instead the role of mouse USVs in the reproductive context; indeed, they summarized the most recent evidence demonstrating that USVs have an important role in parental cooperation, inducing both maternal and paternal behaviors. Bouguiyoud et al. investigated the effects of visual deprivation in acoustic communication and social behaviors using a mouse model of congenital blindness. They demonstrated here that congenital visual deprivation had no effect on the number of USVs emitted in pups and juveniles, but affected the USV emission in adult males and the social behaviors in juvenile and adult mice. The role of social isolation on adult USVs and social behaviors was instead assessed in both male and female mice by Premoli et al. Their study contributed to provide new guidelines for assessing ultrasonic communication in inbred mice, demonstrating also several sex differences in ultrasonic communication and social behaviors that were mostly unaffected by pre-testing social isolation.

Finally, another interesting aspect that has been investigated in a study included in this special issue concerns the molecular mechanisms underlying vocal communication. In particular, Aamodt and White examined whether microRNA-128 is behaviorally regulated in Area X (the striatopallidal song control nucleus) in juvenile zebra finches and found that its levels decline with singing. Furthermore, they demonstrated that inhibition of miR-128 in young birds enhanced the organization of learned vocal sequences, thus suggesting an important role for miR-128 in vocal communication and as a potential therapeutic target for autism spectrum disorders.

Author contributions

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

Funding

SP received funding from Bordeaux University, CNRS, Association Autour de Williams and Fondation pour l'Audition (FPA-RD-2020-8). SAB received funding from the University of Brescia.

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

Blanchard, R. J., Blanchard, D. C., Agullana, R., and Weiss, S. M. (1991). Twenty-two kHz alarm cries to presentation of a predator, by laboratory rats living in visible burrow systems. *Physiol Behav.* 50, 967–972. doi: 10.1016/0031-9384(91)90423-l

Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Curr Opin Neurobiol.* 23, 310–317. doi: 10.1016/j.conb.2013.01.014

Fendt, M., Brosch, M., Wernecke, K. E. A., Willadsen, M., and Wöhr, M. (2018). Predator odour but not TMT induces 22-kHz ultrasonic vocalizations in rats that lead to defensive behaviours in conspecifics upon replay. *Sci Rep.* 8, 11041. doi: 10.1038/s41598-018-28927-4

Knutson, B., Burgdorf, J., and Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *J Comp Psychol.* 112, 65–73. doi: 10.1037/0735-7036.112.1.65

Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* 14, 30–80. doi: 10.1016/J.CONCOG.2004.10.004

Wöhr, M., and Schwarting, R. K. (2013). Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation. *Cell Tissue Res.* 81–97. doi: 10.1007/s00441-013-1607-9

frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fnbeh.2021.735920

LMT USV Toolbox, a Novel Methodological Approach to Place Mouse Ultrasonic Vocalizations in Their Behavioral Contexts—A Study in Female and Male C57BL/6J Mice and in *Shank3* Mutant Females

Fabrice de Chaumont, Nathalie Lemière, Sabrina Coqueran, Thomas Bourgeron† and Elodie Ey*†*

Human Genetics and Cognitive Functions, Institut Pasteur, UMR 3571 CNRS, Université de Paris, Paris, France

OPEN ACCESS

Edited by:
Susanna Pietropaolo, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Francesca R. D'Amato, National Research Council (CNR), Italy
Özge Sungur, University of Marburg, Germany

***Correspondence:**
Fabrice de Chaumont
fabrice.de-chaumont@pasteur.fr
Elodie Ey
eye@igbmc.fr

†These authors share last authorship

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *03 July 2021*
Accepted: *20 September 2021*
Published: *13 October 2021*

Citation:
de Chaumont F, Lemière N, Coqueran S, Bourgeron T and Ey E (2021) LMT USV Toolbox, a Novel Methodological Approach to Place Mouse Ultrasonic Vocalizations in Their Behavioral Contexts—A Study in Female and Male C57BL/6J Mice and in Shank3 Mutant Females. Front. Behav. Neurosci. 15:735920. doi: 10.3389/fnbeh.2021.735920

Ultrasonic vocalizations (USVs) are used as a phenotypic marker in mouse models of neuropsychiatric disorders. Nevertheless, current methodologies still require time-consuming manual input or sound recordings clean of any background noise. We developed a method to overcome these two restraints to boost knowledge on mouse USVs. The methods are freely available and the USV analysis runs online at https://usv.pasteur.cloud. As little is currently known about usage and structure of ultrasonic vocalizations during social interactions over the long-term and in unconstrained context, we investigated mouse spontaneous communication by coupling the analysis of USVs with automatic labeling of behaviors. We continuously recorded during 3 days undisturbed interactions of same-sex pairs of C57BL/6J sexually naive males and females at 5 weeks and 3 and 7 months of age. In same-sex interactions, we observed robust differences between males and females in the amount of USVs produced, in the acoustic structure and in the contexts of emission. The context-specific acoustic variations emerged with increasing age. The emission of USVs also reflected a high level of excitement during social interactions. We finally highlighted the importance of studying long-term spontaneous communication by investigating female mice lacking *Shank3*, a synaptic protein associated with autism. While the previous short-time constrained investigations could not detect USV emission abnormalities, our analysis revealed robust differences in the usage and structure of the USVs emitted by mutant mice compared to wild-type female pairs.

Keywords: mouse, ultrasonic vocalization (USV), social behavior analysis, mouse model, autism, age, sex, Shank3

INTRODUCTION

Social communication regulates major biological functions under strong selective pressure, such as finding reproductive partners, raising progeny, group coordination for territory advertisement, and protection from predators (Bradbury and Vehrencamp, 2011). It is not yet clear whether the mechanisms underlying these functions are shared between species, but an increasing number of genes and brain circuits related to social interaction and communication have been identified

(e.g., Arriaga et al., 2012; Chen and Hong, 2018; Tu et al., 2019; Kelley et al., 2020), suggesting that at least certain social brain circuits are conserved across species. Several genes associated with autism have been mutated in animal models and lead to atypical social interactions (e.g., mice: Ey et al., 2011; Crawley, 2012; rats: Modi et al., 2018; Drosophila: Coll-Tané et al., 2019; monkeys: Tu et al., 2019). This suggests that animals can be used as models to better understand the causes of neuropsychiatric conditions affecting social interactions and communication, keeping in mind their limits (e.g., the absence of convincing proofs of vocal learning, a major characteristic of human communication, in the above-cited model species; Jarvis, 2019). Hereafter, we will focus on mice, due to their broad use as models for neuropsychiatric disorders.

Mice are social animals, naturally living in *demes*, with a single dominant male, occasionally a few subordinate males, and several females occupying contiguous nests, but only a fraction of them reproduce (Palanza et al., 2005). This social organization leads mice to use tactile, olfactory, visual, and vocal (mostly in the ultrasonic range) signals in same-sex social interactions, male-female socio-sexual interactions, and mother-infant relationships (Latham and Mason, 2004; Brennan and Kendrick, 2006; Portfors, 2007). Tactile, visual, and olfactory cues are investigated through the observation of contacts, body postures, and marking behavior (e.g., Mathis et al., 2018; de Chaumont et al., 2019). Ultrasonic vocalizations (USVs) are interpreted as a proxy for vocal communication (Zippelius and Schleidt, 1956; Sewell, 1970; Portfors, 2007). USVs may also represent emotional reflectors of social interactions, reflecting the positive or negative valence of the interaction and the excitement of the interacting mice. Our understanding of mouse USVs nevertheless is still poor compared to our knowledge of vocal communication in other species, such as birds or primates (reviewed in Naguib et al., 2009; for non-human primates in Fischer and Hammerschmidt, 2010).

Mouse USVs are generally investigated in contexts in which the motivation of the animals is controlled, through either social deprivation or the introduction of a sexual component to trigger a maximum quantity of USVs (reviewed in Portfors, 2007). Indeed, after the developmental period when isolated mouse pups emit USVs that trigger approach and retrieval from the mother (Zippelius and Schleidt, 1956; Sewell, 1970), the amount of USVs emitted by juveniles and adults is maximized by social deprivation (Ey et al., 2018). In these contexts, USVs may be used to regulate close contacts and hierarchy (Moles et al., 2007; Ey et al., 2018). Finally, an estrous female, or at least its odor, stimulates adult males to vocalize (Holy and Guo, 2005; Chabout et al., 2015). Such male USVs are suspected to facilitate close body contact with the female (Pomerantz et al., 1983; Hammerschmidt et al., 2009; reviewed in Egnor and Seagraves (2016). Examining USVs in these specific contexts allowed to suggest some of their functions, but, to refine them, linking USVs emission to behavioral contexts is required.

Under constrained conditions, previous studies highlighted that USVs are mostly emitted during close contacts and approach behaviors in male-male (Ferhat et al., 2015) and female-female (Ferhat et al., 2016) interactions (including when one is socially deprived). Specific phases of courtship interactions trigger different call types, which might therefore reflect the different behaviors expressed in these phases (Nyby, 1983; Hanson and Hurley, 2012). It was also shown that male-female and female-female interactions trigger the longest USVs sequences, with the most complex types of USVs, compared to male-male interactions (Matsumoto and Okanoya, 2021). In a group of four adult mice (two males and two females, aged 3–5 months) interacting for 5 h after at least 2 weeks of social deprivation, both the females and males vocalized, more specifically during chasing (Neunuebel et al., 2015). The lowest call rates occurred when the animals were isolated from one another, whereas the highest call rates occurred when the mice were sniffing each other's ano-genital region. Vocalizations are therefore emitted in different proportions depending on the behavioral context (Sangiamo et al., 2020). In paired interactions, USVs trigger no behavioral variations in female-female interactions, but males run faster when the females accelerate while vocalizing in male-female interactions (Warren et al., 2018a,b, 2020). Altogether, linking USVs with behavioral contexts remains focused on constrained interactions and does not shed light on the usage of spontaneous USVs.

In the present study, we developed a method using the Live Mouse Tracker (LMT; de Chaumont et al., 2019), which allows a detailed description of the behavior, synchronized with USV recordings (**Figures 1A,B**). We delivered the complete pipeline of recording and analyses, as well as an original online application[1] (**Figure 1C**) to automatically segment and analyze the USVs. Using this method, we characterized the spontaneous social and vocal behavior of C57BL/6J (hereafter B6) mice in same-sex pairs over an undisturbed period of 3 days and nights. We tested whether the usage and structure of USVs vary according to their behavioral contexts of emission. We focused on same-sex interactions between sexually naive mice to avoid mixing sexual motivation in the factors that influenced vocal behavior. We compared the vocal behavior of male and female pairs at three different ages (5 weeks and 3 and 7 months) to examine age-related maturation. In addition, we investigated whether spontaneous USVs are perturbed in mice lacking Shank3, a glutamatergic synaptic scaffolding protein that we previously associated with autism (Durand et al., 2007; Leblond et al., 2014; deletion of exon 11; Schmeisser et al., 2012; Vicidomini et al., 2017). This first in-depth investigation of spontaneous mouse social communication and the resources presented here will be helpful for the community to design more sensitive tests to better investigate the natural abilities of mice (Warburton et al., 1988; Gerlai and Clayton, 1999), as well as to detect and interpret social communication phenotypes in mouse models of neuropsychiatric disorders.

MATERIALS AND METHODS

Animals

We tested eight male and eight female C57BL/6J mice (hereafter B6 mice; Charles River Laboratories, Ecully, France). Mice

[1] https://usv.pasteur.cloud

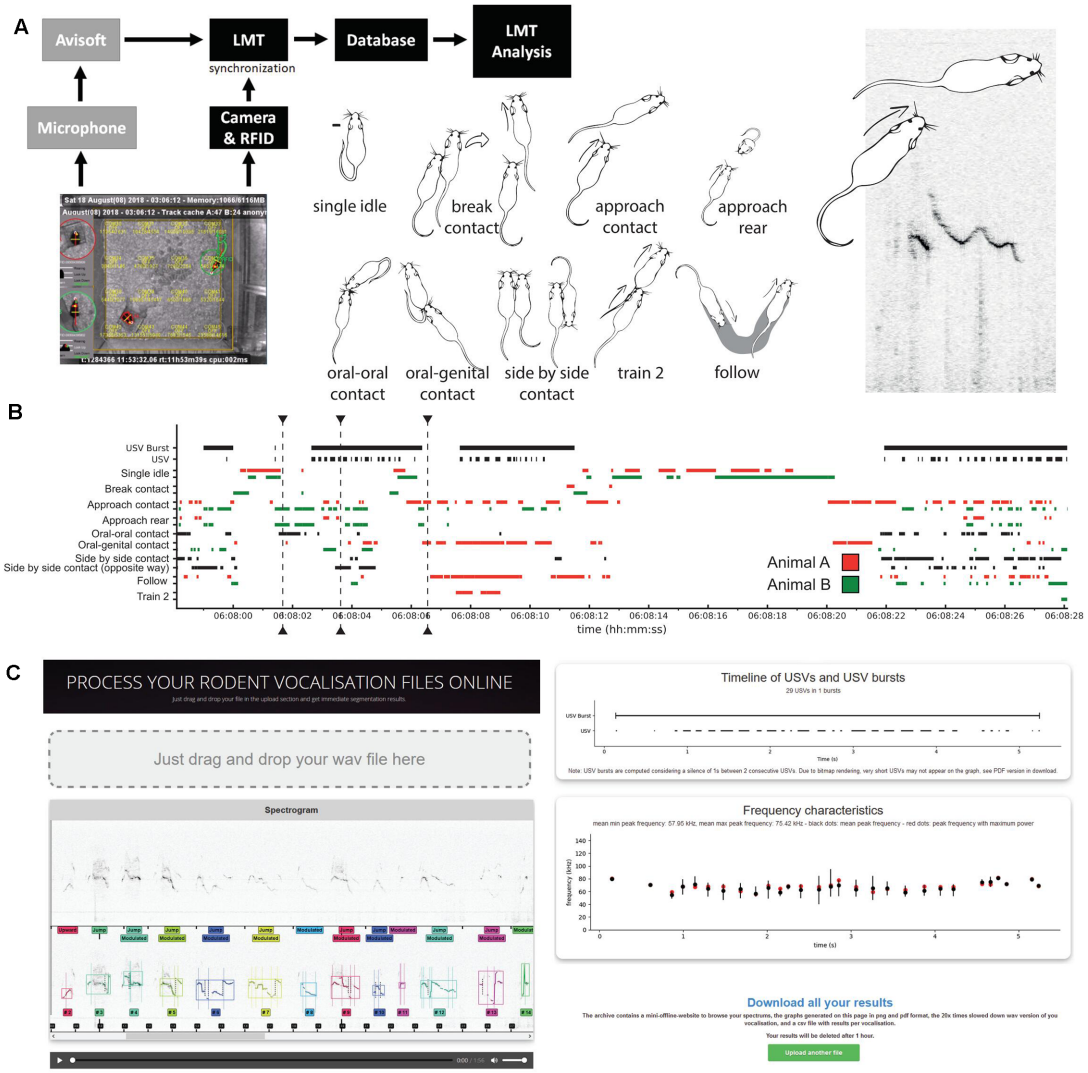


FIGURE 1 | Overview of the study of spontaneous ultrasonic vocalizations (USVs) in same-sex pairs of mice. **(A)** Set-up used to synchronize the spontaneous USVs and behaviors of a same-sex pair of mice using AviSoft and Live Mouse Tracker (LMT). Examples of behaviors automatically extracted using LMT; see definitions in "Materials and Methods" Section and in de Chaumont et al. (2019). **(B)** Timeline of 30 s of the experiment (day 1, 6 a.m.), with USVs, USV bursts, and the major behavioral events. **(C)** Screenshots of the online processing application (https://usv.pasteur.cloud) with an example set. Spectrogram and automatically extracted USVs in which each color labels a different USV. The segmentation follows the shape of the frequency of maximum amplitude. USV and USV burst timeline extracted, frequency characteristics. "Download all results" button providing figures and data for each acoustic feature extracted for each voc.

arrived at 3 weeks of age as a group and were directly housed in same-sex pairs in the experimental facility. *ProSAP2/Shank3* mutant mice (deletion of exon 11; Schmeisser et al., 2012) were generated onsite from heterozygous parents on a C57BL/6J background (>10 backcrosses). Twelve *Shank3*$^{-/-}$ females arrived at 1–1.5 month of age in the experimental facility. Before arrival, these mice were housed within mixed-genotype groups, littermates or not. *Shank3*$^{+/+}$ females were not available in sufficient number to conduct a similar study with these control mice, and therefore B6 females aged 3 months were used as control group given their similar genetic background. *Shank3*$^{-/-}$ females were directly housed in age-matched pairs upon arrival, at least 3 weeks before the recording session to let them stabilize social relationships. In this mouse model, we focused on a single age class (3 months of age) since it is the classical age for behavioral characterization. We focused only on females given the low amount of USVs spontaneously emitted by B6 males, which is the genetic background used to generate the *Shank3* mutant mice tested here. In the experimental facility, mice were housed in classical laboratory cages with food and water *ad libitum* and 11 h/13 h dark/light rhythm (lights off at 08:00 p.m.).

We inserted RFID chips (12 × 2.12 mm; Biolog-Id, Bernay, France) subcutaneously behind the left ear and pushed it down to the flank at 4 weeks of age in B6 mice and between 5 and 7 weeks in *Shank3*$^{-/-}$ mice. After this operation conducted under gas

(isoflurane) anesthesia and local analgesia (<0.05 ml lidocaïne at 20 mg/ml), mice were left at least 1 week to recover.

We did not control for the sexual status of females in our experiments. Nevertheless, as the experiments lasted 3 days, we cover a large proportion of the sexual cycle (see also von Merten et al., 2014). In addition, we wanted to avoid manipulating animals throughout the recording session and following the estrus cycle would have necessitated daily handling.

Behavioral and USV Recordings

For the recordings, each pair was placed in the LMT setup (de Chaumont et al., 2019), a Plexiglas cage of 50 × 50 cm with transparent walls furnished with fresh bedding, food dispersed on the ground in the center, a water bottle at the down right side of the cage, a house (width: 100 mm, depth: 75 mm, height: 40 mm) in red Plexiglas in the down left corner and nesting material (six dental cottons) spread on the bedding in the center of the cage. Each LMT setup was isolated in an experimental room, with no other mouse in the same room. Recordings were launched between 03:00 and 04:00 p.m. (20–23°C, 80–100 lux when lights were on). Once the recordings were started, we did not disturb the mice anymore for 3 days (11 h/13 h dark/light rhythm, with lights off at 08:00 p.m.). Recordings were stopped after 71 h and setups were cleaned with soap water, dried, and re-furnished before launching another recording session. In WT mice, the first recordings occurred at 5 weeks of age. We next recorded in identical conditions the same mice at 3 and 7 months of age. Between the recording sessions, the pairs of mice were left undisturbed, with only a weekly change of the bedding. The six pairs of *Shank3*$^{-/-}$ mice were recorded only once, at 2.5–3 months of age, the age class classically recorded in behavioral phenotyping of mouse models.

Launching recordings consisted in monitoring both behavior and USV emission. Behavioral monitoring occurred through the LMT system (Kinect camera on the top, plugins 465, 524, 600, or 705; de Chaumont et al., 2019), in which mice were automatically identified and tracked throughout the recording session. All spontaneous ultrasonic vocalization sequences were recorded using an Avisoft-Bioacoustics CM16/CMPA microphone (located 50 cm above the bottom right corner of the LMT cage, oriented toward the center of the cage) and the Avisoft UltraSoundGate Recorder system (Avisoft Bioacoustics, Glienicke, Germany; 300 kHz sampling rate, 16-bit format) using the trigger function (trigger: level of this channel; pre-trigger: 1 s; hold time: 1 s; duration >0.005 s; trigger event: 2% energy in 25–125 kHz with entropy <50%). Both systems were synchronized within LMT (see hereafter). Behaviors were attributed to individuals while USVs were measured at the scale of the pair of mice since we cannot identify the emitter.

Behavioral Events

The spontaneous behavior of the mice was automatically labeled using the LMT system. Both social and non-social behaviors were used, as in (de Chaumont et al., 2019; recapitulated in **Table 1**).

USV Segmentation Method

Vocalization Waveform to Spectrum

We process wav files recorded at a sampling rate of 300 kHz with a resolution of 16 bits per sample. We first apply an FFT (overlap of 0.75% and FFT-size = 1,024 points) to the original audio signal to get its spectrum.

The signal recorded presents three main problems: (1) The continuity of the USVs can be interrupted if the signal gets too low. (2) The animals produce a lot of noise by interacting with their environment. (3) The animal facility environment itself produces interfering ultrasonic noise. To overcome these problems which might prevent correct classification, we need to filter the signal of the spectrum. The processing steps are detailed in the **Supplementary Methods** section (**Supplementary Methods**—USV segmentation method: Method parameters, Filtering spectrum data, Constant and blinking frequency canceler, Vocalization segmentation, Spectrum signal extraction). The USV segmentation was validated on a set of manually annotated USVs (**Supplementary Methods**—USV detection validation; **Supplementary Table I**). As recorded files might contain USVs or noise, we sorted them automatically (**Supplementary Methods**—Filtering out wave files containing only noise) and double-check them manually.

Acoustic Feature Extracted

The acoustic traits were computed for each USV (**Supplementary Table II**) and each USV burst (**Supplementary Table III**).

Avisoft Burst Record and Synchronization With Live Mouse Tracker

The system is designed to work for an unlimited duration. As USVs are infrequent events, we do not record the sound continuously. We instead use the automatic record trigger functionality of Avisoft-RECORDER. The automatic trigger of Avisoft-RECORDER monitors the sound level and starts recording a sound if the current sound level within a given frequency range is over a given threshold. The sound is recorded as long as the sound level is over the threshold. This function takes a hold time parameter: if Avisoft-RECORDER detects another signal during the hold period, the record is not interrupted. The hold period also adds a record period around the first and the last signal over the threshold. In our experiment, we use a hold time of 1 s.

To synchronize USV recording with the tracking, we use the "Trigger control" of Avisoft-RECORDER. This function allows launching of an external program at each start and end of records. We use the free software PacketSender[2] to perform communication between Avisoft-RECORDER and Live Mouse Tracker (LMT). Through PacketSender, we send a UDP string packet containing the file number currently recorded by Avisoft-RECORDER. This information is recorded by LMT within the database as an "USV event." The goal of the synchronization is to match the USV record with the current data frame number recorded by LMT.

[2]https://packetsender.com/

TABLE 1 | Definitions of behavioral events extracted by Live Mouse Tracker.

Behavioral events	
Name	**Description**
Single idle	The mouse is stopped (speed lower than 5 pixels per frame) and not in contact with any other mouse.
Single move	The mouse is moving (speed higher than 5 pixels per frame) and not in contact with any other mouse.
Break contact	The mouse breaks a contact with another mouse and moves away from this mouse.
Social approach	The mouse is approaching another one, i.e., the distance between the two animals shortens, the speed of the mouse is higher than the speed of the approached mouse, and the distance between the two animals is shorter than two mean body lengths (of the approached animal). This approach does not necessarily lead to a contact.
Approach contact	The mouse is approaching another one and makes contact with it.
Contact	The mouse is in contact (i.e., the two masks have one common pixel) with the other mouse.
Oral-oral contact	The mouse is sniffing the oral region of another mouse, i.e., the two nose points are less than 15 pixels (26 mm) from one another.
Oral-genital contact	The mouse is sniffing the ano-genital region of another mouse, i.e., the nose point of the first mouse is within 15 pixels (26 mm) from the tail point of the other mouse.
Side by side contact	The side of a mouse is within 30 pixels (52 mm) from the side of the other mouse; both animals are oriented in the same direction.
Side by side contact (head-to-tail)	The side of a mouse is within 30 pixels (52 mm) from the side of the other mouse; both animals are oriented in opposite directions.
Follow	The mouse is walking alone directly in the path of another mouse. Both mice move at a speed higher than 5 pixels per frame, with an angle between their direction vectors smaller than 45°, and the distance between their centers of mass shorter than two mean body lengths.
Train2	The mouse is following another mouse (speed higher than 5 pixels per frame) while sniffing its ano-genital region.

LMT USV Toolbox, an Open-Source, Free, and Online USV Analysis Pipeline

The currently available methods to detect and analyze mouse USVs need specific installations and software. To facilitate the testing of our own algorithm, we provide a website to test the method or to process data online[1]. The user simply drags and drops his/her wave file, waits a few seconds (depending on the length of the sample file), and finally evaluates the quality of the USV segmentation and the data extracted from the sound file. The goal of this website is to provide immediate access to the method without installing any software.

The first panel of the website is dedicated to evaluating USV detection. The first spectrogram represents the original data and the second one provides the annotated data. The player under this spectrogram allows to listen to the sound file slowed down by twenty times. The other panels display:

- the length of the wave file given as input.
- the number of USVs detected within the wave file.
- a timeline displaying the USVs detected over the whole file and their temporal organization in USV bursts, in which the intervals between USVs are shorter than 750 ms.
- the frequency characteristics of each USV within the sound file (in kHz). Each vertical black bar displays the min/max peak frequency of the USV (and therefore also the frequency range) while the black dot displays the mean peak frequency and the red dot displays the peak frequency with the maximum amplitude in each USV.
- the duration of each USV (in ms).
- the power (i.e., amplitude) of each USV, depicted in arbitrary unit.
- the proportion of USVs with frequency modulations.
- the proportion of USVs containing one or more frequency jump(s).
- a table gathering all the acoustic variables extracted on each USV of the sound file.

The user can download all these results for his/her own sound file. These results are deleted after 1 h. Data downloaded from this web page can be directly used with the scripts that we provide in the present study. To perform the analysis on thousands of files, we also provide the desktop version of the analysis program, working in batch mode (download through the Live Mouse Tracker website: http://livemousetracker.org).

USV Analysis Toolbox

For Live Mouse Tracker, we provided a full API in Python to process event classification and to process queries. As for Live Mouse Tracker, we provide an API in Python for the biologists to process USVs[3]. This package allows one to re-create all data representations used in this study with its own data. This API is available on gitHub[3].

Analyses and Statistical Tests

The whole data processing, computations, statistical analyses, and figure generation were conducted using Python 3.8. *P*-values are presented as uncorrected for multiple testing, with a distinctive sign (°) if they survived after Bonferroni correction indicated hereafter for each analysis.

[3]https://github.com/fdechaumont/LMT-USV-Toolbox

Behavioral Profiles of Mice

We built a behavioral profile for each mouse considering the following behavioral event types: single move, single idle, get away, break contact, approach within the social range, approach contact, nose-nose contact, nose-anogenital contact, side-by-side contact, side-by-side contact/head-to-tail, follow, train2. We also examined the total distance traveled. For each behavioral event type, we compared the total time spent in this behavioral event type, the number of events, and the mean duration of these events.

Each of these variables was tested for the effect of age within each sex separately. We used a non-parametric Kruskal–Wallis test followed by paired Wilcoxon tests to compare age classes within each sex. When applying a correction for multiple testing, we corrected it by the number of tests conducted (2). The effect of sex was tested within each age class separately using Mann–Whitney *U*-tests. When applying a correction for multiple testing, we corrected it by the number of age classes (3). Variables for *Shank3*$^{-/-}$ mice were compared to variables for B6 mice aged of 3 months using Mann–Whitney *U*-tests (no correction applied).

Comparison Between Age Classes, Sexes, or Genotypes of USVs and USV Bursts

To examine the effect of age class, we used nonparametric paired Wilcoxon tests to compare the amount of USVs between age classes within each sex. When applying a correction for multiple testing, we corrected it by the number of age classes (3). We compared the amount of USVs between *Shank3*$^{-/-}$ and B6 mice using nonparametric Mann–Whitney *U*-tests (no correction applied).

To detect an age or a sex effect, we used a linear mixed model (fixed factor: age or sex, random factor: pair) to compare the number of USVs per USV burst between sexes within each age class (5 weeks, 3 months, and 7 months). When applying a correction for multiple testing, we corrected it by the number of age classes (3). To detect an effect of the *Shank3* mutation on the number of USVs per USV burst, we used a linear mixed model (fixed factor: genotype, random factor: pair) to compare the number of USVs per USV burst between *Shank3*$^{-/-}$ pairs and B6 pairs. The same analyses were conducted within each behavioral context (single idle, break contact, approach contact, nose-nose contact, nose-anogenital contact, side-by-side contact, side-by-side contact/head-to-tail, follow, train2). When applying a correction for multiple testing, we corrected it by the number of behavioral contexts examined (9).

Acoustic features of USVs were compared between sexes within age classes, between age classes within each sex, and also between sexes within age classes and within each behavioral context using Linear Mixed Models using the pairs as a random factor. When applying a correction for multiple testing, we corrected it by the number of age classes (3) multiplied by the number of acoustic variables tested (16) when testing sex effect, by two tests multiplied by the number of acoustic traits (16) when testing age effect. When comparing acoustic traits across contexts, corrections included the number of acoustic traits (16) and the number of contexts combinations. We also compared the acoustic features between B6 and *Shank3*$^{-/-}$ mice using Linear Mixed Models with genotype as a fixed factor and genotype as a random factor. When applying a correction for multiple testing, we corrected it by the number of acoustic variables (16).

Context-Specific Acoustic Features

We first selected representative acoustic variables (**Supplementary Methods**—Selection of representative acoustic variables) for the main text but nevertheless depicted all of them in the **Supplementary Figures**. We tested whether acoustic features of USVs (duration, frequency characteristics, frequency range, modulations, harshness, slope) varied according to the contexts in which USVs were emitted. For that purpose, after computing the acoustic features of all USVs, we compared acoustic features of USVs between the different behavioral contexts (single idle, break contact, approach contact, nose-nose contact, nose-anogenital contact, side-by-side contact, side-by-side contact/head-to-tail, follow, train2) using linear mixed models (fixed factor: context, random factor: pair) with a Bonferroni correction (combination of two within nine behavioral events × 16 acoustic features). Acoustic variations between contexts in *Shank3*$^{-/-}$ mice were tested similarly (Figure M1c-d). Heatmaps represented significance levels after Bonferroni correction (size of the points) as well as the effect size and direction of variations (cold/warm colors).

Relationship Between the Occurrence of USV Bursts and the Speed of Mice as Well as the Duration of Social Events

We focused on the following sample of behavioral events: break contact, approach contact, contact, nose-anogenital contact, follow, and Train2. First, we tested whether animals displaying these behaviors accompanied by USVs had a higher speed than when they displayed the same behaviors without USVs. For that purpose, for each behavioral event of each type, we computed the mean speed of the mouse displaying this behavior over the whole event. We performed a paired Wilcoxon test (alternative = "greater") to compare the mean speed of the animal with and without USVs at the group level. In addition, to better understand inter-individual variations, we compared, within each individual and for each behavioral event type, the speeds during events overlapping with USVs to the speeds during events not overlapping with USVs using Mann–Whitney *U*-tests (unpaired, one-sided).

Second, we tested whether one type of behavioral event was longer when emitted concomitantly with USVs than when it was not concomitant with USVs. For that purpose, within each individual and for each behavioral event type, we gathered the durations of events overlapping with USVs as well as the duration of events not overlapping with USVs. We compared these two lists of durations using Mann–Whitney *U*-tests (unpaired, one-sided) within each individual. We also performed a paired Wilcoxon test (alternative = "greater") to compare at the group level the mean duration of the event with USVs and the mean duration of the same event without USVs. To control for the fact that longer behavioral events were more likely to include

USVs, we used a random process in which USVs would occur at a random time. In this case, the longer USVs would have more chances to overlap with a USV. We simulated 1,000 different random distributions of the USVs, and we kept the same proportions of USVs overlapping with behavioral events as in the original data. We performed this test for each behavioral event in each experiment separately (data not shown). In each simulation, we performed the same statistical test (Mann–Whitney *U*-test) as in the original data, and we kept the *p*-value. The ratio of the number of *p*-values below the original one provided the probability to obtain the same effect randomly. In our study, this probability was less than 3%, suggesting that the effects of the presence of USVs on the duration of behavioral events could be found randomly with only a weak probability, i.e., longer events were not more likely to occur with USVs than shorter ones.

simulation code:

Compute_Duration_Events_With_USV_Simulation.py

We conducted these tests for both animals of each WT female pair at 5 weeks, 3 months, and 7 months of age as well as for both animals of *Shank3*$^{-/-}$ pairs aged of 3 months for all behavioral events above cited. Overall, we adjusted our statistical analyses for multiple testing accordingly (Bonferroni correction).

Intonation Within USV Bursts

The intonation graph displays a representation of the burst where the duration of each burst is stretched from 0 to 100. 0 for the start of the burst and 100 for its end. Corresponding frames are computed and if a USV is present at this frame, we report the value for the given acoustic parameter and display −/+ its STD. The same method is performed for the number of USVs.

Corresponding code and graph generation:

HarshLMocationInBurst3.py

RESULTS

A Platform to Analyze and Synchronize USVs With Behaviors

We needed to automatically detect USVs despite the background noise (i.e., animals moving in the bedding, continuous electronic or ventilation noises) and to extract their acoustic features, all these aspects being synchronized with behavioral monitoring (**Supplementary Methods**—Motivation to create a new recording and analysis pipeline). In our system, sounds were recorded using Avisoft SASLab Pro Recorder and synchronized with behavioral monitoring using LMT (Methods—Avisoft record and synchronization with Live Mouse Tracker). Mechanical noises (e.g., amplitudes at continuous frequencies from electronic devices or ventilation) were removed (**Supplementary Methods**—Filtering spectrum data) and the USVs extracted (**Supplementary Methods**—USV segmentation method/USV detection validation). The USVs were then synchronized with behaviors and the acoustic characteristics extracted and stored as metadata in the database (**Supplementary Tables II, III**—Description of the acoustic variables measured for each USV and burst). We provide a library in Python to access and query these data. We developed an online application that allows the segmentation and analysis of wave files without software installation[1]. Annotated spectrograms are generated online to allow visual inspection of the results. We also provide online examples. The desktop version of this resource is also available and open source.

Age- and Sex-Related Variations in Spontaneous Social Behaviors

We examined the spontaneous social interactions of male and female same-sex pairs of B6 mice over 3 days. The previously observed reduced activity in males compared to females (de Chaumont et al., 2019) was not significant in juvenile males but only in mature males (i.e., aged 3 months or more) compared to females (**Figure 2A**). The reduced time spent in social contacts displayed by males compared to females was visible already in immature males (example of nose-anogenital contact in **Figure 2B**; see also **Supplementary Figure 1**). Surprisingly, the mean duration of non-specific contacts (i.e., contact between the two mice without specifying any body parts) tended to be longer in mature males than in females (**Figure 2C**), while, as soon as specific contact are considered, the mean duration of these specific contacts such as nose-nose (**Figure 2D**) and nose-anogenital contacts (**Figure 2E** and **Supplementary Figure 2**) were shorter in males than in females. Therefore, even if contact events are globally longer in males than in females, these two very specific contact types seem more prominent in females than in males. Males also displayed significantly shorter approach (**Figure 2F**) or escape (**Figure 2G**) behaviors compared to females in adulthood, suggesting that their contacts initiated and terminated more abruptly, possibly reflecting aggressive interactions. As expected, these sex-related differences were less visible when the animals were immature (i.e., aged 5 weeks; **Supplementary Figures 1, 2**). In both sexes, the total time spent in nose-nose and nose-anogenital contacts tended to decrease with increasing age (**Figure 2B** and **Supplementary Figure 1**).

Compared to age-matched pairs of wild-type B6 female mice, *Shank3*$^{-/-}$ pairs displayed atypical organization of their social contacts. Indeed, while they spent typical total time in social interactions (**Figure 2H** and **Supplementary Figure 3**), *Shank3*$^{-/-}$ pairs displayed oral-oral (**Figure 2I**) and side-side (**Figure 2J**) contact events of significantly longer mean durations compared to age-matched B6 females (see also **Supplementary Figure 4**).

Major Sex-Related Variations in the Usage of Spontaneous USVs

Quantification of USVs

Over these 3 days of monitoring, spontaneous USVs were mostly (around 80%) emitted during the nocturnal activity period in both B6 males and females of any age (**Figure 3A** and **Supplementary Figures 5, 6**). Interestingly, there was a strong sex-related difference in USV emission, with pairs of B6 males emitting significantly fewer USVs than pairs of B6 females across

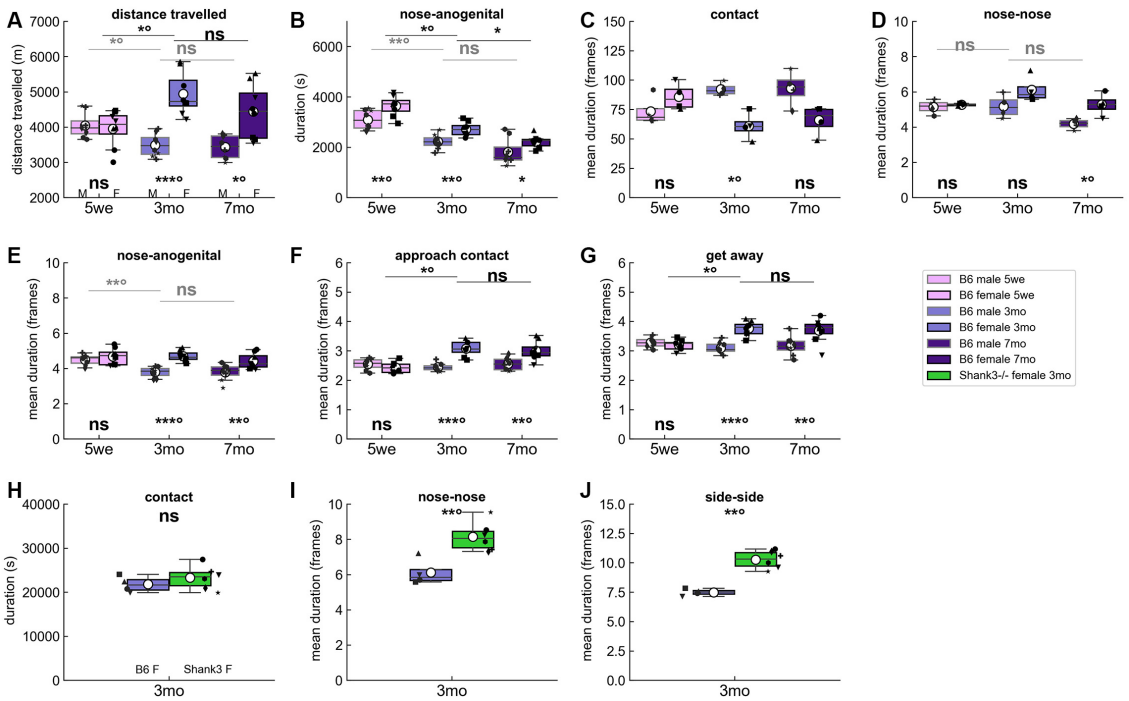


FIGURE 2 | Spontaneous behaviors in B6 male and female pairs as well as in *Shank3*$^{-/-}$ female pairs over three nights. **(A)** Total distance traveled over the three nights of recording in B6 male (M) and female (F) pairs at the three age classes tested. Age- and sex-related variations in the **(B)** total time spent in nose-anogenital contact, **(C)** mean duration of contact events (i.e., contacts in a broad sense without any specific body part involved), **(D)** mean duration of nose-nose contact events, **(E)** mean duration of nose-anogenital contact events, **(F)** mean duration of social approach events, and **(G)** mean duration of get away events over the three nights of recording. Each black dot corresponds to an individual in **(A,B,E–G)** with a similar shape for the two individuals of the pair, while in **(C,D)** each dot corresponds to a pair since the behaviors are symmetrical (Mann–Whitney *U*-tests used to test for differences between sexes within each age class; Kruskal–Wallis test followed by Wilcoxon paired test if significant to test for differences between 5 weeks and 3 months and between 3 months and 7 months within each sex). Comparisons between B6 female pairs aged 3 months and age-matched *Shank3*$^{-/-}$ female pairs of the **(H)** total time spent in contact, **(I)** mean duration of nose-nose contact events, and **(J)** mean duration of side-side contact events over the three nights. Each black dot corresponds to a pair since the behaviors are symmetrical (Mann–Whitney *U*-tests used to test for differences between genotypes). Uncorrected *p*-values: ns: non significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. *P*-values followed by ° survived a correction for multiple testing over age classes.

all age classes over the 3 days of recording (**Figure 3B**). The total number of USVs recorded for 3 months-old *Shank3*$^{-/-}$ females over the 3 days (5,527.0 ± 3,236.8) was not statistically different from that of USVs recorded for age-matched B6 females (6,589.3 ± 1,979.8; Mann–Whitney *U*-test: $U = 8.0$, $p = 0.228$; **Figure 3C**). *Shank3*$^{-/-}$ mice emitted significantly fewer USVs than B6 mice only during the second night ($U = 2.0$, $p = 0.021$; data not shown).

Contexts of Emission of Ultrasonic Vocalizations

We next examined the contexts in which USVs were emitted. Again, we observed a strong sex-related difference in the contexts of emission. Indeed, in males, most USVs were emitted when the mice were at a distance from each other (i.e., single idle), while in females most USVs were emitted in contact with each other, initiating contact, or approaching in a social range (**Figure 3D**). Interestingly, the Train2 behavior (i.e., following while keeping ano-genital contact) was highly specific to USV emission since up to 70% of these events were accompanied by USVs, in contrast to even closely related social behaviors such as ano-genital contacts or follow without contact (**Figure 3E**). The context-specificity described here was observed in all age classes as well as in 3-months old *Shank3*$^{-/-}$ female pairs (**Figures 3D,E**), despite their perturbed social behavior (see behavioral profiles above).

Variations of the Acoustic Characteristics of USVs With Sex, Age, and Genotype

For clarity of presentation, we selected acoustic variables based on PCA over the whole set of B6 USVs and validated through correlations between acoustic variables (see **Supplementary Methods**—Selection of representative acoustic variables). The whole set of acoustic features is available in the supplementary files. Spectrograms of six USVs for B6 males (**Figure 4A**), B6 females (**Figure 4B**) and *Shank3*$^{-/-}$ females (**Figure 4C**) aged 3 months are given as examples. When all USVs were considered, we observed that mature B6 males emitted significantly shorter (**Figure 4D**), less modulated (**Figure 4E**) and less harsh (i.e., less noisy; **Figure 4F**) USVs compared to females (see also **Supplementary Figure 7**). These differences did not reach significance in juveniles (5 weeks of age) but only when animals were fully mature (i.e., at 3 and 7 months of age; **Supplementary Figure 7**). B6 males also emitted USVs with limited frequency

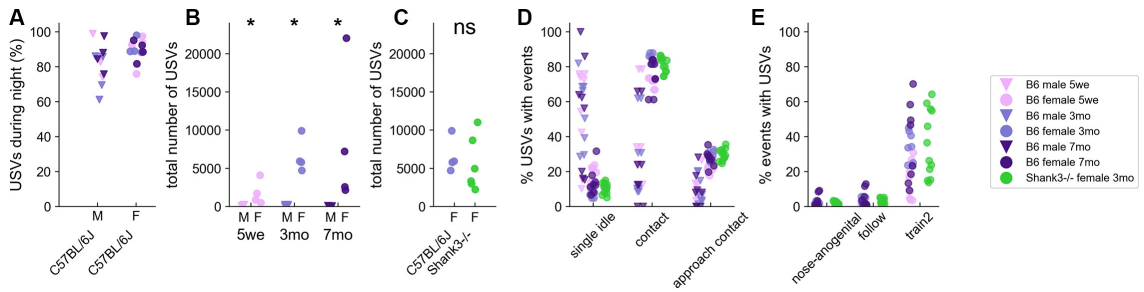


FIGURE 3 | Spontaneous emission of USVs in B6 male and female pairs as well as in *Shank3*$^{-/-}$ female pairs over three nights. **(A)** The proportion of USVs emitted during the night periods of the 3 days of recording in B6 pairs. **(B)** Total number of USVs recorded in male-male and female-female B6 pairs at 5 weeks, 3 months, and 7 months of age over the 71 h of recording (Mann–Whitney *U*-tests; uncorrected *p*-values: $^{*}p < 0.05$; none of the *p*-values survived a correction for multiple testing. **(C)** Total number of USVs recorded in female-female B6 pairs and in female-female *Shank3*$^{-/-}$ pairs aged 3 months (Mann–Whitney *U*-tests; ns: not significant). **(D)** The proportion of USVs occurring simultaneously with single idle and contact events in male-male and female-female B6 pairs at 5 weeks, 3 months, and 7 months of age (age classes represented by colors) as well as in female-female *Shank3*$^{-/-}$ pairs aged 3 months. **(E)** The proportion of nose-anogenital contacts, follow and train2 behaviors occurring simultaneously with USVs in female-female B6 pairs at 5 weeks, 3 months, and 7 months of age (age classes represented by colors) as well as in female-female *Shank3*$^{-/-}$ pairs aged 3 months.

difference between the start and the end points of the signal compared to females at any age (**Supplementary Figure 7**). Overall, USVs emitted by males appeared to be shorter, purer, and flatter compared to females' USVs.

When examining the influence of age on vocal production, we found that, in B6 females, the duration, frequency modulations, harshness, and loudness of USVs increased with increasing age, while the frequency-related traits decreased (**Supplementary Figure 8**). In males, this evolution with increasing age was more restricted, with a significant increase in duration, frequency, and harshness only between 3 and 7 months of age, as if it took more time for males to establish their social signals. Age-related variations in frequency-related traits appeared not to be stable between age classes, with the lowest frequency-related traits at 3 months of age (**Supplementary Figure 8**).

Contrary to our hypothesis of simpler vocal communication in *Shank3*$^{-/-}$ mice, *Shank3*$^{-/-}$ mice emitted USVs with acoustic characteristics that were not significantly different from those of B6 females (**Figures 4G–J**). There was only a trend for USVs to be starting at higher frequency (genotype: $z = 2.31$ [396.0, 4,775.1], $p = 0.021$; intercept: $z = 75.66$, $p < 0.001$ [63,713.9, 67,102.7]; **Figure 4J**) and to be quieter (i.e., less powerful) in *Shank3*$^{-/-}$ mice compared to B6 female mice (genotype: $z = -2.80$ [−0.42, −0.07], $p = 0.005$; intercept: $z = 15.22$, $p < 0.001$ [0.90, 1.17]; **Figure 4K**; see also **Supplementary Figure 9**). The differences in the power characteristics were not related to differences in the proximity to the microphone (**Supplementary Figures 10A,B**). The slightly smaller body volume (previously reported for body size in Schmeisser et al., 2012) and here estimated through body surface on the tracking mask; **Supplementary Figure 10C**) of *Shank3*$^{-/-}$ mice relative to that of B6 mice could partially explain the reduced power of *Shank3*$^{-/-}$ mouse USVs.

We investigated further these variations related to age, sex, and genotype across the different contexts of emission. We aimed at testing whether the differences described above depend on the behavioral contexts.

Acoustic Variations Related to the Context of Emission of USVs

Age- and Sex-Related Variations Across Contexts

In B6 mice, sex-related acoustic variations were significant mostly in USVs emitted during a single idle (**Supplementary Figures 11–13**). The direction of variations differed between 5 weeks of age and mature age classes. Indeed, at 5 weeks of age, during single idle, females emitted USVs that were shorter, steeper, with less frequency modulations than males (**Supplementary Figure 11**). In contrast, at three (and to a lesser extent seven) months of age, in single idle, females emitted longer (**Figure 5A**), steeper, more modulated (**Figure 5B**), higher-pitched (**Figure 5C**), and more frequency-broken (i.e., more jumps) USVs compared to males (**Supplementary Figures 12, 13**).

In males, the variations observed with increasing age remained significant only in single idle (and to a lower extent in approaching before a contact and in nose-nose contacts), where frequency-related traits were minimum at 3 months of age (**Supplementary Figure 14**). In other behavioral contexts, either the number of USVs emitted was too low or the variations were not significant. In females, the variations related to age were consistent across all behavioral contexts, with increasing duration, frequency modulations, power, and harshness as well as decreasing frequency-related traits with increasing age (**Supplementary Figure 15**).

Shank3$^{-/-}$ mice emitted USVs with higher frequency-related traits compared to B6 females in the least social contexts (single idle and break contact; **Supplementary Figure 16**). Interestingly, the reduced power detected in USVs emitted by *Shank3*$^{-/-}$ females compared to B6 females was not significant in the most intense social context (Train2), while it was significant in other types of social contacts and in approach before a contact (**Supplementary Figure 16**). This suggests that the high level of arousal in the Train2 context masks the power deficit displayed by *Shank3*$^{-/-}$ mice.

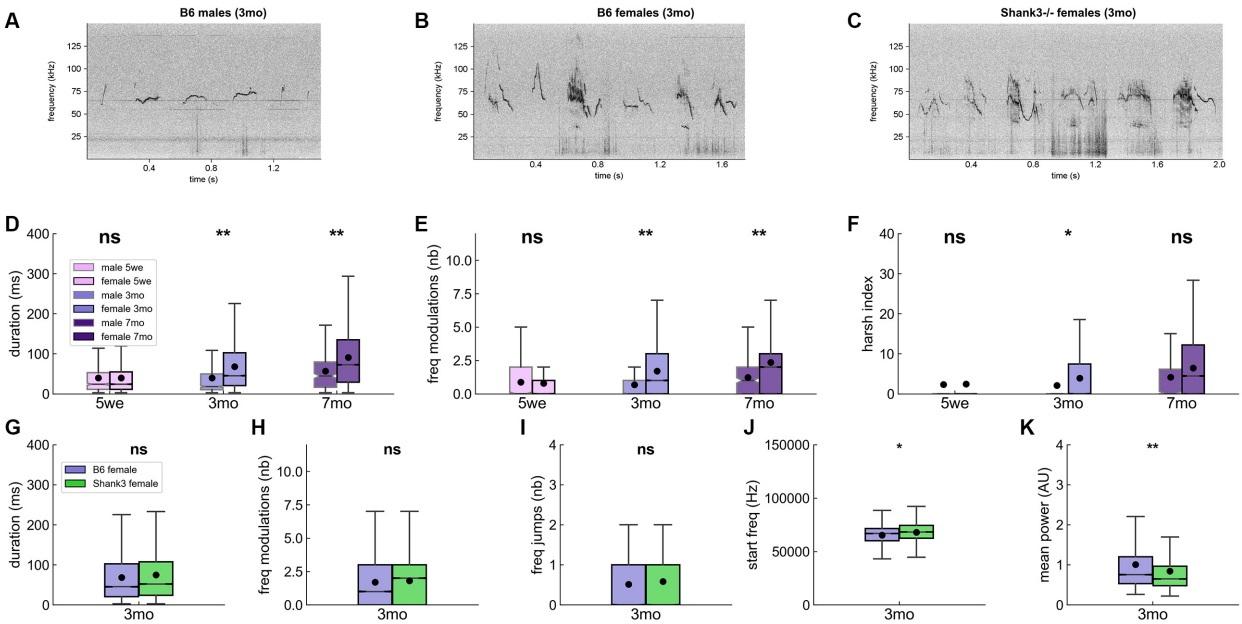


FIGURE 4 | Acoustic characteristics over age and sex classes in B6 male and female pairs of mice and in *Shank3*$^{-/-}$ female pairs. Spectrograms of six examples of USVs in **(A)** B6 males, **(B)** B6 females, and **(C)** *Shank3*$^{-/-}$ females aged 3 months (1,024 FFT points, 1.7 ms time resolution, Hamming window, 50% overlap). Comparisons of the duration of the USVs **(D)**, number of frequency modulations per USV **(E)**, and harsh index of the USVs **(F)** between males and females at 5 weeks (males: 630 USVs, females: 7,167 USVs), 3 months (males 688 USVs, females: 26,357 USVs) and 7 months (males: 241 USVs, females: 33,954 USVs) of age in B6 mice (Linear Mixed Model with sex as fixed factor and pair as random factor). Comparisons of the duration of USVs **(G)**, the number of frequency modulations per USV **(H)**, the number of frequency jumps **(I)**, the frequency at the beginning of the USV **(J)**, and the mean power of USVs **(K)** between B6 females (26,357 USVs, purple) and *Shank3*$^{-/-}$ females (33,162 USVs, green) aged 3 months (Linear Mixed Model with genotype as fixed factor and pair as random factor). ns, not significant; uncorrected *p*-values: $^{*}p < 0.05$, $^{**}p < 0.01$. None of the tests survived a correction for multiple testing.

Variations of Acoustic Features Between Contexts

From this step of the analysis, we restricted our investigations to females, given the low number of USVs emitted by males in other contexts than single idle. In B6 females, the context specificity increased with increasing age. Indeed, at 5 weeks (**Supplementary Figure 17**) and 3 months (**Supplementary Figure 18**) of age, USVs emitted during single idle and break contact displayed significantly shorter duration (**Figure 5D**), lower frequency modulations (**Figure 5E**), lower harsh index compared to USVs emitted in other contexts. USVs emitted in both contexts were similar to each other, except for the mean frequency that was lower in break contact than in any other context (**Figure 5F**). In addition, the USVs emitted during the different types of close contacts (nose-nose, nose-anogenital, side-side) displayed similar frequency characteristics and modulations but differed from other social contexts (see example for the frequency modulations in **Figure 5E**). At 7 months of age (**Supplementary Figure 19**), USVs emitted during the different types of close contacts were again not significantly different. In contrast, USVs emitted during the approach preceding a contact emerged as acoustically different from the USVs emitted during close contacts, follow and train2 behaviors as if the aging animals refined their approach behavior.

In *Shank3*$^{-/-}$ mice recorded at 3 months of age, we observed that USVs emitted during Follow and Train2 behaviors were acoustically similar (**Supplementary Figure 20**), and both types significantly differed compared to USVs emitted in all other contexts, i.e., longer (**Figure 5G**), harsher, larger frequency modulations (**Figure 5H**) and lower mean frequency (**Figure 5I**). Interestingly, USVs emitted by *Shank3*$^{-/-}$ mice during nose-to-nose contacts were significantly different [i.e., shorter duration (**Figure 5G**), lower frequency modulation (**Figure 5H**), higher-pitched (**Figure 5I**)] from USVs emitted in other types of contacts. This observation confirmed the specific impairment of *Shank3*$^{-/-}$ mice to make nose-nose contacts (see above).

Comparisons of Behaviors With and Without USVs

We next hypothesized that behaviors accompanied by USVs were more intense (i.e., longer and with higher mean speed) than behaviors not accompanied by USVs. As a proxy for arousal, we considered the speed of the animals. We thus quantified the mean speed of the animals during behaviors accompanied or not by USVs, as well as the duration of these behaviors.

Our hypothesis was not verified for Train2 events in any age class at the individual level. Indeed, within each individual, the duration of the event and the speed of the mice did not vary significantly, regardless of whether USVs were emitted or not (**Figure 6A**). This might be related to the fact that this behavior was the most specific to USVs (**Figure 3E**). In contrast, follow (**Figure 6B**), approach contact (**Figure 6C**), contact, getaway, and oral-genital contact (data not shown) showed both a significantly

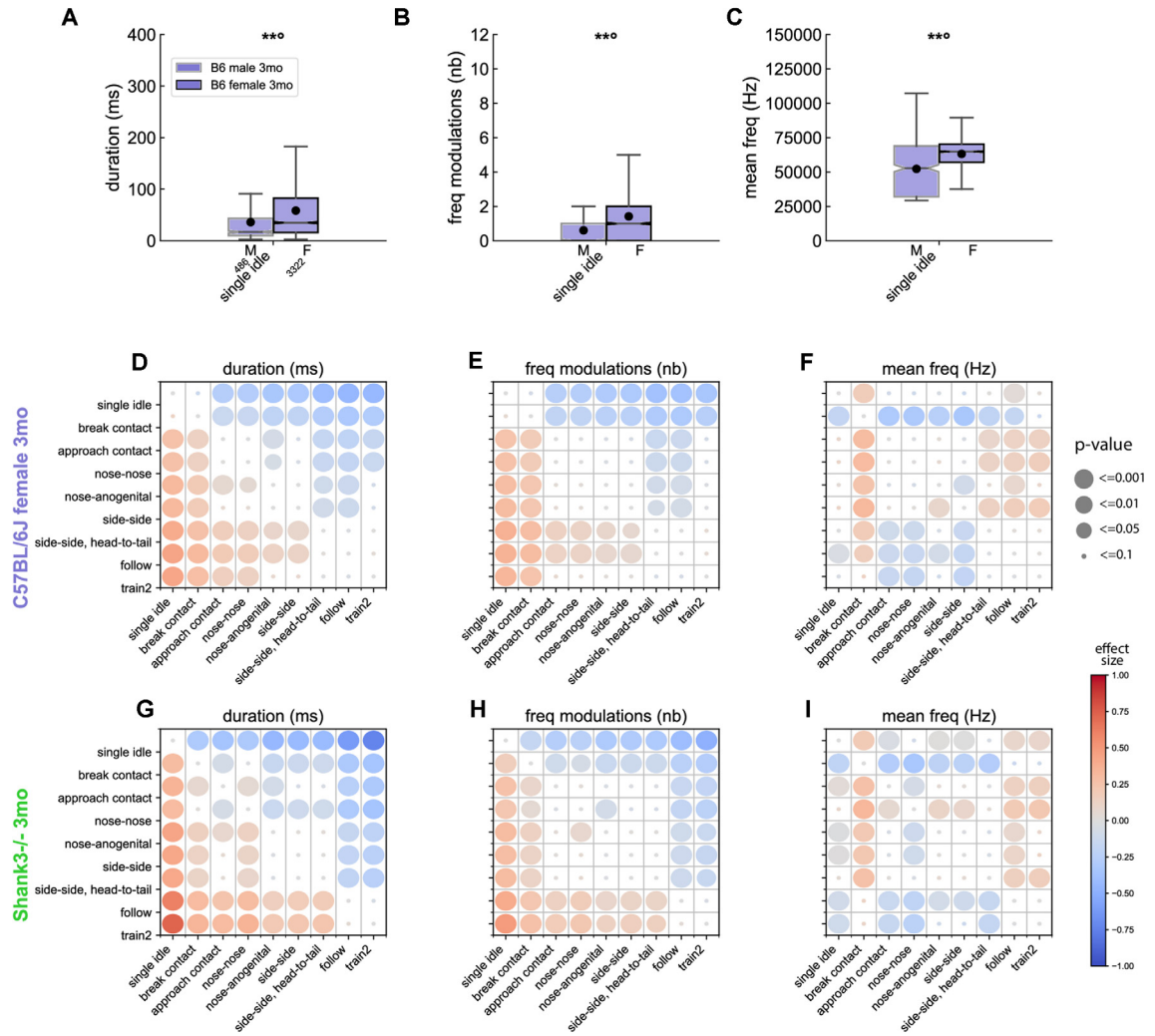


FIGURE 5 | Acoustic variations of USVs between contexts of emission in B6 and *Shank3*$^{-/-}$ mice. Sex-related variations in **(A)** the duration, **(B)** the number of frequency modulations, and **(C)** the mean frequency of USVs emitted during single idle by 3 months old male (486 USVs) and female (3,322 USVs) B6 mice (Linear Mixed Model with sex as fixed factor and pair as random factor; ns: not significant, uncorrected *p*-values: $^{**}p < 0.01$: *p*-values followed by ◦ survived the correction for multiple testing for acoustic variables and events). Comparison of the duration **(D)**, number of frequency modulations **(E)** and mean frequency **(F)** of USVs emitted by 3 months old B6 females between the different contexts of emission (Linear Mixed Models with context as fixed effect and pair as random effect). Blue colors indicate that the acoustic feature of USVs given during y-event is lower than the acoustic feature of USVs given in x-event; red colors indicate that the acoustic feature of USVs given during y-event is higher than the acoustic feature of USVs given in x-event; the effect size is represented by the color scale while the significance levels are represented by the size of the circles. Comparison of the duration **(G)**, number of frequency modulations **(H)** and mean frequency **(I)** of USVs emitted by 3 months old *Shank3*$^{-/-}$ females between the different contexts of emission (Linear Mixed Models with context as fixed effect and pair as random effect; the effect size is represented by the color scale while the significance levels corrected for multiple testing across acoustic variables and contexts combinations are represented by the size of the circles).

higher speed and longer duration when accompanied by USVs, in agreement with our hypothesis at the individual level. For break contact, the presence of USVs was accompanied by higher speed but not necessarily a longer duration (**Figure 6D**). Overall, the relationship between USV emission and behavioral display (speed and duration of the event) suggests USVs can be considered as an expression of higher arousal in simple (e.g., follow, approach contact, break contact, oral-genital contact) social investigation, but not in the very intense and highly specific social investigation (such as Train2). These results obtained in B6 mice were also observed in *Shank3*$^{-/-}$ mice (**Supplementary Figure 21**).

Organization in Bursts

Event-Driven Determination of the Threshold for Burst Definition

We observed that USVs were spontaneously emitted in sequences that we call "USV bursts." The inter-USV interval threshold to classify two successive USVs in two different USV bursts was determined by examining the optimal correlation

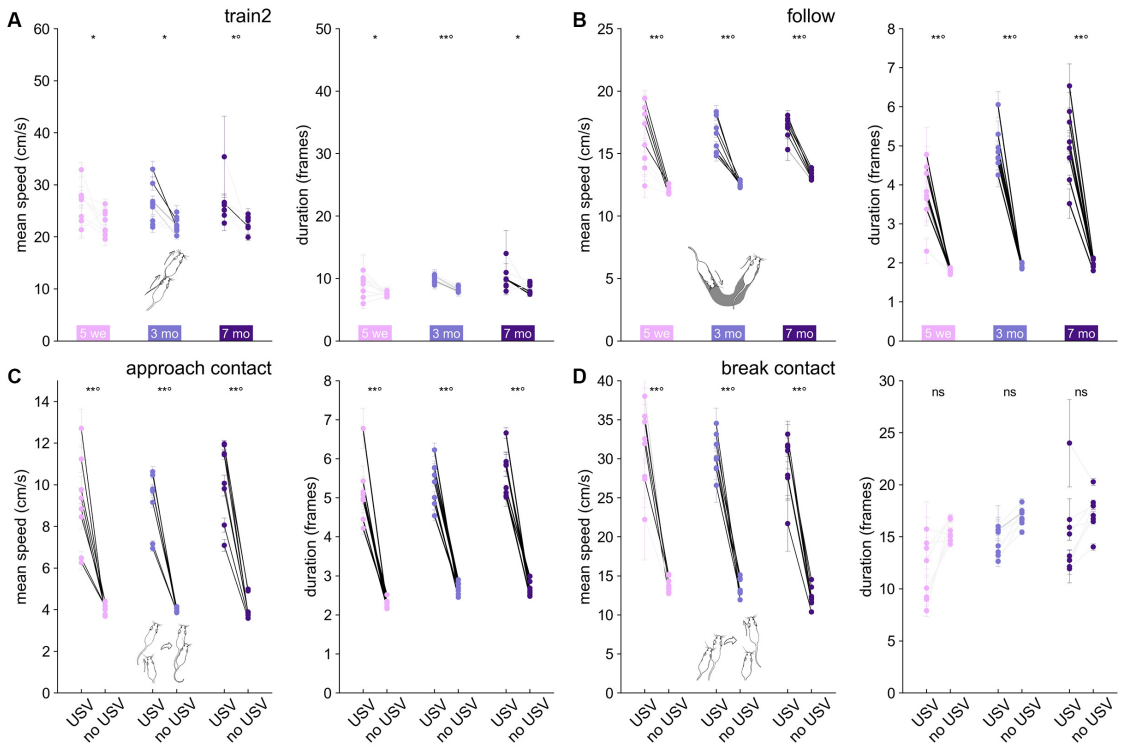


FIGURE 6 | Variations in mean speed and duration of each behavioral event according to the presence or absence of USVs in B6 females. Variations of the mean speed of the animal performing the behavior (left panel) and of the event duration (right panel) for **(A)** Train2, **(B)** follow, **(C)** approach contact, and **(D)** break contact. Paired comparisons between behaviors with and without USVs were conducted at the group level using the paired Wilcoxon test for each age class (*$p < 0.05$, **$p < 0.01$; p-values remaining significant after correction for multiple testing over age classes were followed by °). Significant differences at the individual level using Mann–Whitney U-tests after Bonferroni correction are depicted by the color of the segments linking the mean value, with and without USVs. light gray: not significant, medium gray: $p < 0.05$, dark gray: $p < 0.01$, black: $p < 0.001$.

between USV bursts and behavioral events (see **Supplementary Methods**—Determining intervals between USVs to define USV burst). We defined that USVs separated by intervals longer than 750 ms belong to two different bursts. This 750 ms threshold was valid in pairs of B6 males, in pairs of B6 females in all age classes, as well as in *Shank3*$^{-/-}$ pairs.

Sex, Age, and Genotype Effects on Bursts Characteristics

Using this definition of a USV burst, we again observed a robust sex-related difference in the burst organization. Indeed, males tended to emit USV bursts containing less USVs than females at any age (Linear Mixed Models, hereafter LMM: 5 weeks: sex: $z = -2.22$ [−8.53, −0.52], $p = 0.027$; intercept: $z = 5.83$, $p < 0.001$ [5.20, 10.48]; 3 months: sex: $z = -9.55$, $p < 0.001$ [−12.45, −8.21], intercept: $z = 29.87$, $p < 0.001$ [12.19, 13.90]; 7 months: sex: $z = -3.43$, $p = 0.001$ [−12.12, −3.31], intercept: $z = 14.89$, $p < 0.001$ [10.11, 13.18]; **Figure 7A**). No significant age effect was detected for the number of USVs per USV burst in males. In contrast, females aged 5 weeks emitted USV bursts with fewer USVs than 3-month-old females (LMM: age: $z = -7.77$, $p < 0.001$ [−6.01, −3.64], intercept: $z = 22.96$, $p < 0.001$ [11.82, 14.021]), but the number of USVs per USV burst did not differ significantly between 3 and 7 months of age (age: $z = 1.68$, $p = 0.093$ [0.13, 1.69]; **Figure 7A**). Interestingly, the burst organization was perturbed in *Shank3*$^{-/-}$ females. Indeed, at 3 months of age, *Shank3*$^{-/-}$ mice emitted USV bursts that contained fewer USVs (mean ± standard deviation: 9.6 ± 10.6 USVs per USV burst) compared to age-matched B6 mice (13.1 ± 15.9 USVs per USV burst; LMM: genotype: $z = -4.93$, $p < 0.001$ [−5.29, −2.28], intercept: $z = 21.96$, $p < 0.001$ [11.87, 14.20]; **Figure 7B**). This was verified across all contexts (**Figure 7C**).

Context-Specificity of Burst Structure

Across all age classes, USV bursts emitted during single idle and approach contact were the shortest compared to bursts emitted in all other contexts in B6 pairs and in *Shank3*$^{-/-}$ pairs (**Figure 7D** for B6 females aged 3 months; see **Supplementary Figure 22** for other age classes and for between-contexts comparisons in *Shank3*$^{-/-}$ females). In contrast, USV bursts emitted during Train2 and to a lesser extent in side-side, head-to-tail, and break contact were the longest (**Figure 7D** and **Supplementary Figure 22**). USV bursts emitted during the different types of close contacts were not different from one another.

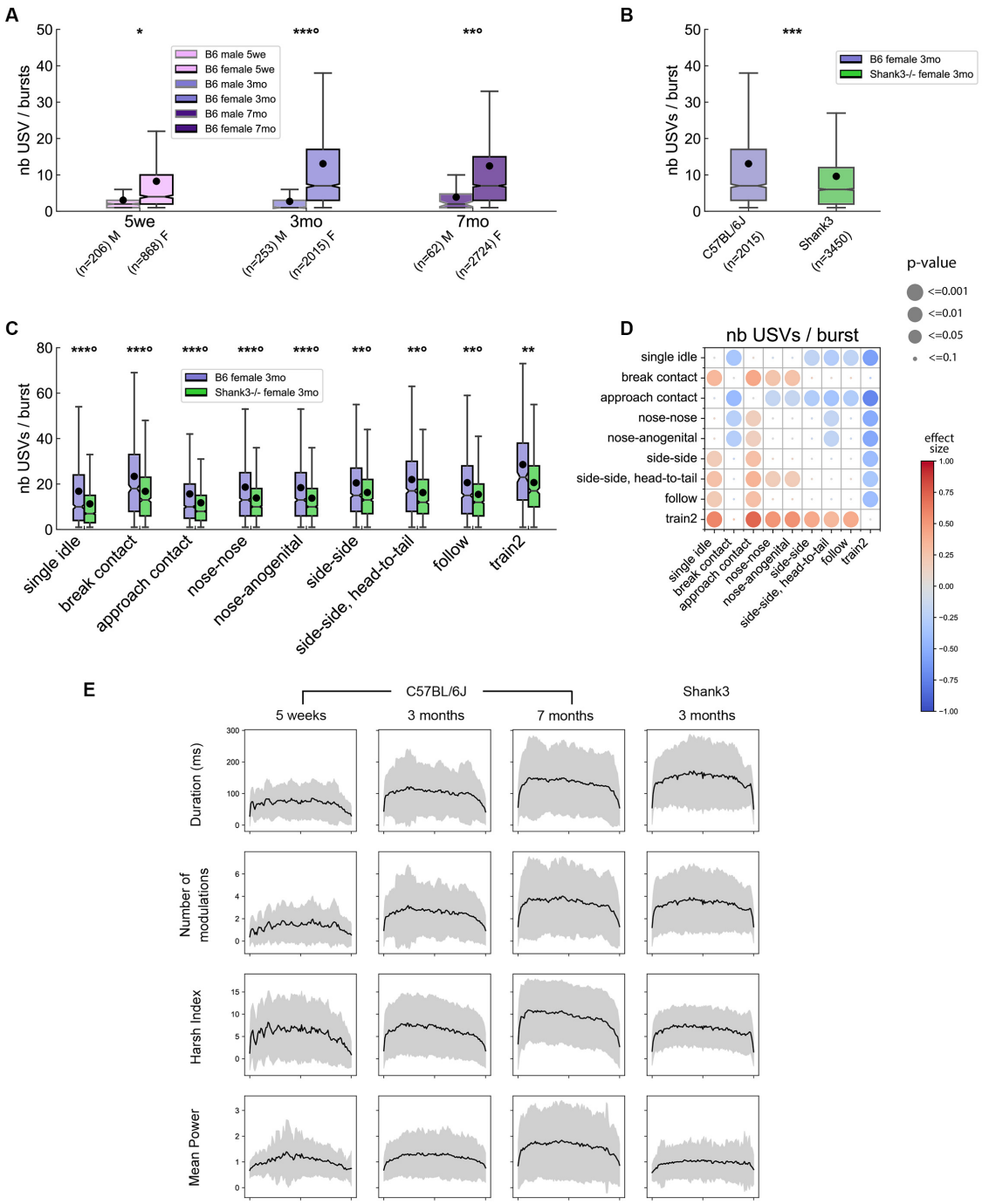


FIGURE 7 | Bursts characteristics in male and female pairs of B6 mice and in *Shank3*$^{-/-}$ female mice. **(A)** Comparison of the number of USVs per burst between B6 males and females at 5 weeks, 3 months and 7 months of age (Linear Mixed Model with sex as fixed factor and pair as random factor). **(B)** Comparison of the number of USVs per burst between B6 females and *Shank3*$^{-/-}$ females aged 3 months (Linear Mixed Model with genotype as fixed factor and pair as random factor). **(C)** Comparison of the number of USVs per burst between B6 females and *Shank3*$^{-/-}$ females aged 3 months across the different contexts of emission (Linear Mixed Model with sex as fixed factor and pair as random factor). ns: not significant, uncorrected *p*-values: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; *p*-values followed by ° survived a correction for multiple testing (**A**: age classes, **B**: no correction, **C**: number of behavioral contexts). **(D)** Comparison of the number of USVs per burst emitted by 3 months old B6 females between the different contexts of emission (Linear Mixed Models with context as fixed factor and pair as random factor). Blue colors indicate that the number of USVs per burst given during y-event is lower than the number of USVs per burst given in x-event; red colors indicate that the number of USVs per burst given during y-event is higher than the number of USVs per burst given in x-event; the effect size is represented by the color scale while the significance levels are represented by the size of the circles. **(E)** Evolution of acoustic characteristics of USVs (duration, frequency modulations, harsh index, mean power) according to their position within a burst in B6 females aged 5 weeks, 3 months, and 7 months as well as in *Shank3*$^{-/-}$ female mice aged 3 months.

Evolution of Variables Over the Bursts

For bursts containing nine or more USVs, we examined the evolution of the acoustic characteristics over the sequences of USVs. We restrained our analyses to females since such long USV bursts were extremely rare in males. In 3 months old B6 females, USVs appeared to be more complex in the middle of the USV bursts than at the extremities. This means that USVs in the middle of a burst displayed longer duration, larger frequency modulations, higher harsh index, and higher power in comparison with those at the beginning or at the end of the burst. These profiles were more subtle at 5 weeks of age and clearer at 7 months of age (**Figure 7E**, three left columns). Interestingly, *Shank3*$^{-/-}$ mice displayed similar variations over the bursts, except for the mean power. Indeed, *Shank3*$^{-/-}$ pairs showed a reduced ability to modulate the power of their USVs over a burst compared to B6 mice (**Figure 7E**, right column).

DISCUSSION

Tackling the Complexity of Mouse Ultrasonic Vocalizations

In the vast majority of previous studies investigating mouse social communication, USVs were triggered by social deprivation (e.g., 2 weeks of isolation) and recorded in the first minutes or hours of interaction (Scattoni et al., 2011; Neunuebel et al., 2015; Warren et al., 2018a, 2020; Sangiamo et al., 2020). The USV types were either classified in a pre-determined repertoire (Scattoni et al., 2011) or simplified for modeling to reduce the complexity of the signals (e.g., ignoring harmonic components or frequency jumps or normalizing over duration; (Neunuebel et al., 2015; Warren et al., 2018a, 2020; Sangiamo et al., 2020; but see Warren et al., 2021). In our study, we provide a complementary approach by determining a large set of acoustic variables to avoid masking the complexity of the signals and leave the door open for any user to design their own classification as in Kikusui et al. (2021). This approach has the limit of inducing a large number of statistical tests over the different acoustic traits. In addition, the reduced number of pairs here recorded allowed to highlight only large effects between age or sex classes, and also between genotypes, despite the large amount of USVs recorded from each pair. We kept this limit in mind in our exploratory study, and we remain as cautious as possible throughout the study in the interpretation of the results. Future studies will elaborate on these results which work as working paths.

We chose not to build an exhaustive repertoire, as the link between meaning and classification can be a pitfall due to the normalization of time and frequency, which provides the same meaning to short, long, or differently dynamic USVs. We chose to examine specific acoustic characteristics and relate them directly to the behavioral context. This approach should help to understand the information carried in USVs (Hertz et al., 2020). We observed that USVs emitted during intense social interactions were longer, harsher, more modulated, and showed more frequency jumps than USVs emitted during non-social behaviors or at the end of social contacts (**Figure 5**). This is reminiscent of USVs emitted by wild house mice during social contact, which are longer than USVs emitted when standing alone (near a food spot, in the nest; Hoier et al., 2016).

Hypotheses About the Functions of USVs

As previously observed in captive wild house mice (von Merten et al., 2014), we confirm over a long period that males in same-sex pairs emit few spontaneous USVs. Over short-term experiments, C57BL/6J male-male pairs also emitted the lowest number of USVs compared to male-female and female-female pairs, with the shortest bouts and simplest USVs (Matsumoto and Okanoya, 2021). In contrast, socially-deprived males have been shown to emit a non-negligible number of USVs (e.g., Chabout et al., 2012; Hammerschmidt et al., 2012; Ferhat et al., 2015). In these previous studies, the vocal repertoire did not differ significantly between males and females (Hammerschmidt et al., 2012; Ey et al., 2013; but see Matsumoto and Okanoya, 2021) and both sexes emitted USVs mostly during ano-genital sniffing and approach behavior (Ferhat et al., 2015, 2016). In the present study, males emitted USVs in short bursts with context-specific rates a 100 times lower than that of females, probably because females were more active (longer distance traveled during the night) and spent more time in social interactions than males (**Supplementary Figure 1**; see also de Chaumont et al., 2019). Overall, because males are slightly less social and active than females, they may encounter fewer situations that trigger USVs. Nevertheless, this incommensurate decrease in call rate relative to females may be related to the natural social structure of mice, with only subordinate males regrouped in same-sex subgroups (Palanza et al., 2005). Other communication modalities, such as body posture and tactile contacts, may be sufficient for males to regulate their social interactions.

We observed that females emitted USVs mostly during intense social interactions, such as Train2, and follow behaviors. The speed of the mice was also significantly higher during behaviors associated with USVs than those without (**Figure 6**). It is thus possible that USVs reflect a high level of arousal, at least in C57BL/6J, the mouse strain tested here. The emission of a large number of spontaneous USVs suggests that female mice are sufficiently aroused in their "home-cage-like" life. In short-term experiments, socially-housed mice never reach this level of excitement during the first minutes of interaction (probably focusing on the exploration of the environment) and therefore emit few USVs (e.g., Hammerschmidt et al., 2012; Ey et al., 2018). The emission of a large quantity of USVs can only be reached by social deprivation over such short periods. Social behaviors, such as follow, or approach contact, were significantly longer when accompanied by USVs than when occurring without (**Figure 6**). This result parallels the reported increased duration of chasing when USVs were emitted by females chased by males (Neunuebel et al., 2015). It is thus possible that some USVs are emitted only when a certain level of excitement is reached, for example after a specific time spent in contact. Whether certain USVs serve to maintain social contacts is currently unknown and would

need to be tested using playback experiments synchronized with behavioral monitoring.

Subtle Social Communication Abnormalities in *Shank3* Mutant Mice

In our setting, *Shank3*$^{-/-}$ mice were less active, took a longer time to approach, and shorter time to escape their conspecific (**Supplementary Figure 3**), while individual contact events tended to be longer than those of WT mice (**Supplementary Figure 4**). These activity and social specificities are consistent with the results of previous studies using the same mutant mouse strain (Vicidomini et al., 2017; de Chaumont et al., 2019). Other genetic models of *Shank3* mutant mice also display reduced social interest (*Shank3*-KO ex4–9: Bozdagi et al., 2010; Wang et al., 2011; Yang et al., 2012; *Shank3*-KO in ex21: Duffney et al., 2015; *Shank3* ex4–9: Jaramillo et al., 2016; *Shank3*-cKI: Mei et al., 2016; *Shank3* exon 4–22 complete KO: Wang et al., 2016; *Shank3B*: Balaan et al., 2019; *Shank3* exon 13–16: Peca et al., 2011; Dhamne et al., 2017; Fourie et al., 2018; *Shank3*B$^{+/-}$: Orefice et al., 2019; Pagani et al., 2019). Nevertheless, other models failed to detect any atypical social interest in *Shank3* mutant mice (*Shank3*-KO ex4–9: Drapeau et al., 2014; *Shank3*-KO ex9: Lee et al., 2015; *Shank3*-KO ex 21: Kouser et al., 2013; *Shank3*-KO and HZ in ex21: Speed et al., 2015; *Shank3*-KO ex11–21 rat model: Song et al., 2019; conditional *Shank3* exon 4–22 knockout in the forebrain, striatum, and striatal D1 and D2 cells: Bey et al., 2018; *Shank3*$^{+/Q321R}$ and *Shank3*$^{Q321R/Q321R}$: Yoo et al., 2019). Such variability may be the consequence of differences in the type of mutation, the behavioral protocols, or housing conditions.

There are few studies that have investigated ultrasonic communication in *Shank3* mutant mice. Pup isolation calls were not affected in *Shank3*-KO ex4–9 mice (Yang et al., 2012; Jaramillo et al., 2016), the *Shank3*-KO ex11–21 rat model (Song et al., 2019), or *Shank3b*-KO mice (Balaan et al., 2019), whereas those of complete *Shank3*-KO ex4–22 mice showed a lower rate, as well as shorter duration, lower frequency, and lower amplitude (Wang et al., 2016). In the context of a male briefly interacting with an estrus female, the rate of USV emission was not affected in *Shank3*-KO ex21 (Kouser et al., 2013) or *Shank3*-KO ex4–9 mice (Yang et al., 2012). Nevertheless, it was lower in *Shank3B*-KO mutant males (Dhamne et al., 2017; Pagani et al., 2019) and higher in *Shank3*-KO ex4–9 males (Wang et al., 2011; Jaramillo et al., 2016) than in WT male mice. In *Shank3*$^{Q321R/Q321R}$ mutant mice, the number of USVs in male-female interactions was not significantly different from that of WT mice, but *Shank3*$^{+/Q321R}$ mutants showed a higher mean duration of USVs than WT mice (Yoo et al., 2019). Male complete *Shank3*-KO ex4–22 mice also showed a lower rate of USVs than WT males when encountering an estrus female, but they were of shorter duration, reduced amplitude, and normal peak frequency (Wang et al., 2016). One strength of our study is that it was designed to directly relate communication deficits with behavioral deficits. For example, the lower call rate of Shank3$^{-/-}$ mice on the second night may be related to a significantly reduced total duration of Train2 and significantly fewer of these Train2 events (data not shown), a behavior that triggers a high rate of USV emission in WT mice.

Concerning USV bursts, *Shank3*$^{-/-}$ mice emitted fewer USVs per burst, but their USVs were weaker than those of B6 mice, with stable power throughout the burst, a trait rarely observed in B6 mice. Variations in USV duration were shown to be present in the *Shank3* mutant carrying the Q321R point mutation (Yoo et al., 2019), while the reduction of amplitude/power was also highlighted in the *Shank3* KO ex4–22 mice (Wang et al., 2016), in parallel with the unstable dominance hierarchy in complete *Shank3* KO triads relative to WT triads (Wang et al., 2016). Overall, in our hands, *Shank3*$^{-/-}$ mice were able to emit USVs with similar acoustic structures and used them in similar contexts as B6 mice but perturbed hierarchical relationships may modify the structure of the USV bursts.

Perspectives

Here, we showed sex- and age-related differences in the spontaneous communication of mice. Major quantitative and qualitative differences emerged between males and females from 3 months of age on. Females emitted more USVs than males, specifically during intense social investigations. With increasing age, mice emitted longer and more complex USVs, with specific differences between USV acoustic structure and behavioral contexts. Female *Shank3*$^{-/-}$ mice emitted USV bursts with fewer but weaker USVs than age-matched WT females. Future developments will include the implementation of the identification of the emitter. Indeed, identifying the sound source will refine the contexts in which USVs are emitted, allowing also to test the interaction between mice of different sexes or genotypes. Our system offers the possibility to characterize spontaneous mouse communication and paves the way for new studies investigating the complex interplay between genetic background, social experience, and hierarchy in the richness of social communication.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Zenodo repository (DOI: 10.5281/zenodo.5060503).

ETHICS STATEMENT

The animal study was reviewed and approved by Ethical committee of the Institut Pasteur (CETEA n°89), followed by the approval of the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation under the reference APAFIS#7706-2016112317252460 v2.

AUTHOR CONTRIBUTIONS

FC created the segmentation and analysis methods and performed the data mining. SC and NL performed the genotyping of *Shank3* mutant mice. EE designed the study, performed and analyzed experiments. EE, FC, and TB conceived

the project and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially funded by the Institut Pasteur, the Bettencourt-Schueller Foundation, the Cognacq–Jay Foundation, the Conny–Maeva Foundation, the ERANET–NEURON ALTRUISM program, the Agence Nationale de la Recherche through grant number ANR-10-LABX-62-IBEID, France-BioImaging infrastructure through grant number ANR-10-INBS-04 and the INCEPTION program through grant number ANR-16-CONV-0005, the Centre National de la Recherche Scientifique, the University Paris Diderot, the BioPsy Labex, and the Foundation for Medical Research (Equipe DEQ20130326488). This study also benefited from the financial support of GIS "Autisme et Troubles du Neuro-Développement". The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Raimund Specht for Avisoft Bioacoustics for upgrading the Avisoft Recorder software, Laetitia Breton, François Rimlinger, Didier Montéan, and Christophe Joubert for animal facility support, Benoît Forget and Nicolas Torquet for critical reading of the manuscript.

SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2021.735920/full#supplementary-material.

REFERENCES

Arriaga, G., Zhou, E. P., and Jarvis, E. D. (2012). Of mice, birds and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS One* 7:e46610. doi: 10.1371/journal.pone.0046610

Balaan, C., Corley, M. J., Eulalio, T., Leite-ahyo, K., Pang, A. P. S., Fang, R., et al. (2019). Juvenile Shank3b deficient mice present with behavioral phenotype relevant to autism spectrum disorder. *Behav. Brain Res.* 356, 137–147. doi: 10.1016/j.bbr.2018.08.005

Bey, A. L., Wang, X., Yan, H., Kim, N., Passman, R. L., Yang, Y., et al. (2018). Brain region-specific disruption of Shank3 in mice reveals a dissociation for cortical and striatal circuits in autism-related behaviors. *Transl. Psychiatry* 8:94. doi: 10.1038/s41398-018-0142-6

Bozdagi, O., Sakurai, T., Papapetrou, D., Wang, X., Dickstein, D. L., Takahashi, N., et al. (2010). Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction and social communication. *Mol. Autism* 1:15. doi: 10.1186/2040-2392-1-15

Bradbury, J. W., and Vehrencamp, S. L. (2011). *Principles of Animal Communication*, Second Edition. Sunderland, MA: Sinauer Associates.

Brennan, P. A., and Kendrick, K. M. (2006). Mammalian social odours: attraction and individual recognition. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2061–2078. doi: 10.1098/rstb.2006.1931

Chabout, J., Sarkar, A., Dunson, D. B., and Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Front. Behav. Neurosci.* 9:76. doi: 10.3389/fnbeh.2015.00076

Chabout, J., Serreau, P., Ey, E., Bellier, L., Aubin, T., Bourgeron, T., et al. (2012). Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS One* 7:e29401. doi: 10.1007/s11356-021-16495-6

Chen, P., and Hong, W. (2018). Neural circuit mechanisms of social behavior. *Neuron* 98, 16–30. doi: 10.1016/j.neuron.2018.02.026

Coll-Tané, M., Krebbers, A., Castells-Nobau, A., Zweier, C., and Schenck, A. (2019). Intellectual disability and autism spectrum disorders 'on the fly': insights from *Drosophila*. *Dis. Model. Mech.* 12:dmm039180. doi: 10.1242/dmm.039180

Crawley, J. N. (2012). Translational animal models of autism and neurodevelopmental disorders. *Dialogues Clin. Neurosci.* 14, 293–305. doi: 10.31887/DCNS.2012.14.3/jcrawley

de Chaumont, F., Ey, E., Torquet, N., Lagache, T., Dallongeville, S., Imbert, A., et al. (2019). Real-time analysis of the behaviour of groups of mice *via* a depth-sensing camera and machine learning. *Nat. Biomed. Eng.* 3, 930–942. doi: 10.1038/s41551-019-0396-1

Dhamne, S. C., Silverman, J. L., Super, C. E., Lammers, S. H. T., Hameed, M. Q., Modi, M. E., et al. (2017). Replicable *in vivo* physiological and behavioral phenotypes of the Shank3B null mutant mouse model of autism. *Mol. Autism* 8:26. doi: 10.1186/s13229-017-0142-z

Drapeau, E., Dorr, N. P., Elder, G. A., and Buxbaum, J. D. (2014). Absence of strong strain effects in behavioral analyses of Shank3-deficient mice. *Dis. Model. Mech.* 7, 667–681. doi: 10.1242/dmm.013821

Duffney, L. J., Zhong, P., Wei, J., Matas, E., Cheng, J., Qin, L., et al. (2015). Autism-like deficits in Shank3-deficient mice are rescued by targeting actin regulators. *Cell Rep.* 11, 1400–1413. doi: 10.1016/j.celrep.2015.04.064

Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* 39, 25–27. doi: 10.1038/ng1933

Egnor, S. E. R., and Seagraves, K. M. (2016). The contribution of ultrasonic vocalizations to mouse courtship. *Curr. Opin. Neurobiol.* 38, 1–5. doi: 10.1016/j.conb.2015.12.009

Ey, E., Leblond, C. S., and Bourgeron, T. (2011). Behavioral profiles of mouse models for autism spectrum disorders. *Autism Res.* 4, 5–16. doi: 10.1002/aur.175

Ey, E., Torquet, N., de Chaumont, F., Lévi-Strauss, J., Ferhat, A.-T., Le Sourd, A.-M., et al. (2018). Shank2 mutant mice display hyperactivity insensitive to methylphenidate and reduced flexibility in social motivation, but normal social recognition. *Front. Mol. Neurosci.* 11:365. doi: 10.3389/fnmol.2018.00365

Ey, E., Torquet, N., Le Sourd, A.-M., Leblond, C. S., Boeckers, T. M., Faure, P., et al. (2013). The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. *Behav. Brain Res.* 256, 677–689. doi: 10.1016/j.bbr.2013.08.031

Ferhat, A.-T., Le Sourd, A.-M., de Chaumont, F., Olivo-Marin, J.-C., Bourgeron, T., and Ey, E. (2015). Social communication in mice - are there optimal cage conditions? *PLoS One* 10:e0121802. doi: 10.1371/journal.pone.0121802

Ferhat, A.-T., Torquet, N., Le Sourd, A.-M., de Chaumont, F., Olivo-Marin, J.-C., Faure, P., et al. (2016). Recording mouse ultrasonic vocalizations to evaluate social communication. *J. Vis. Exp.* 112:e53871. doi: 10.3791/53871

Fischer, J., and Hammerschmidt, K. (2010). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* 10, 17–27. doi: 10.1111/j.1601-183X.2010.00610.x

Fourie, C., Vyas, Y., Lee, K., Jung, Y., Garner, C. C., and Montgomery, J. M. (2018). Dietary zinc supplementation prevents autism related behaviors and striatal synaptic dysfunction in Shank3 exon 13–16 mutant mice. *Front. Cell. Neurosci.* 12:374. doi: 10.3389/fncel.2018.00374

Gerlai, R., and Clayton, N. S. (1999). Analysing hippocampal function in transgenic mice: an ethological perspective. *Trends Neurosci.* 22, 47–51. doi: 10.1016/s0166-2236(98)01346-0

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., and Fischer, J. (2009). Female mice respond to male ultrasonic "songs" with approach behaviour. *Biol. Lett.* 5, 589–592. doi: 10.1098/rsbl.2009.0317

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., and Fischer, J. (2012). The structure and usage of female and male mouse ultrasonic vocalizations reveal only minor differences. *PLoS One* 7:e41133. doi: 10.1371/journal.pone.0041133

Hanson, J. L., and Hurley, L. M. (2012). Female presence and estrous state influence mouse ultrasonic courtship vocalizations. *PLoS One* 7:e40782. doi: 10.1371/journal.pone.0040782

Hertz, S., Weiner, B., Perets, N., and London, M. (2020). Temporal structure of mouse courtship vocalizations facilitates syllable labeling. *Commun. Biol.* 3:333. doi: 10.1038/s42003-020-1053-7

Hoier, S., Pfeifle, C., von Merten, S., and Linnenbrink, M. (2016). Communication at the garden fence - context dependent vocalization in female house mice. *PLoS One* 11:e0152255. doi: 10.1371/journal.pone.0152255

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

Jaramillo, T. C., Speed, H. E., Xuan, Z., Reimers, J. M., Liu, S., and Powell, C. M. (2016). Altered striatal synaptic function and abnormal behaviour in Shank3Exon4–9 deletion mouse model of autism: Shank3Exon4-9 deletion mouse model of autism. *Autism Res.* 9, 350–375. doi: 10.1002/aur.1529

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science* 366, 50–54. doi: 10.1126/science.aax0287

Kelley, D. B., Ballagh, I. H., Barkan, C. L., Bendesky, A., Elliott, T. M., Evans, B. J., et al. (2020). Generation, coordination and evolution of neural circuits for vocal communication. *J. Neurosci.* 40, 22–36. doi: 10.1523/JNEUROSCI.0736-19.2019

Kikusui, T., Sonobe, M., Yoshida, Y., Nagasawa, M., Ey, E., de Chaumont, F., et al. (2021). Testosterone increases the emission of ultrasonic vocalizations with different acoustic characteristics in mice. *Front. Psychol.* 12:680176. doi: 10.3389/fpsyg.2021.680176

Kouser, M., Speed, H. E., Dewey, C. M., Reimers, J. M., Widman, A. J., Gupta, N., et al. (2013). Loss of predominant Shank3 isoforms results in hippocampus-dependent impairments in behavior and synaptic transmission. *J. Neurosci.* 33, 18448–18468. doi: 10.1523/JNEUROSCI.3017-13.2013

Latham, N., and Mason, G. (2004). From house mouse to mouse house: the behavioural biology of free-living Mus musculus and its implications in the laboratory. *Appl. Anim. Behav. Sci.* 86, 261–289. doi: 10.1016/j.applanim.2004.02.006

Leblond, C. S., Nava, C., Polge, A., Gauthier, J., Huguet, G., Lumbroso, S., et al. (2014). Meta-analysis of SHANK mutations in autism spectrum disorders: a gradient of severity in cognitive impairments. *PLoS Genet.* 10:e1004580. doi: 10.1371/journal.pgen.1004580

Lee, J., Chung, C., Ha, S., Lee, D., Kim, D.-Y., Kim, H., et al. (2015). Shank3-mutant mice lacking exon 9 show altered excitation/inhibition balance, enhanced rearing and spatial memory deficit. *Front. Cell. Neurosci.* 9:94. doi: 10.3389/fncel.2015.00094

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289. doi: 10.1038/s41593-018-0209-y

Matsumoto, Y. K., and Okanoya, K. (2021). Mice modulate ultrasonic calling bouts according to sociosexual context. *R. Soc. Open Sci.* 5:180378. doi: 10.1098/rsos.180378

Mei, Y., Monteiro, P., Zhou, Y., Kim, J.-A., Gao, X., Fu, Z., et al. (2016). Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature* 530, 481–484. doi: 10.1038/nature16971

Modi, M. E., Brooks, J. M., Guilmette, E. R., Beyna, M., Graf, R., Reim, D., et al. (2018). Hyperactivity and hypermotivation associated with increased striatal mGluR1 signaling in a Shank2 rat model of autism. *Front. Mol. Neurosci.* 11:107. doi: 10.3389/fnmol.2018.00107

Moles, A., Costantini, F., Garbugino, L., Zanettini, C., and D'Arnato, F. R. (2007). Ultrasonic vocalizations emitted during dyadic interactions in female mice: a possible index of sociability? *Behav. Brain Res.* 182, 223–230. doi: 10.1016/j.bbr.2007.01.020

Naguib, M., Janik, V., Clayton, N., and Zuberbuhler, K. (2009). ''Vocal communication in birds and mammals,'' in *Advances in the Study of Behaviour*, (Oxford: Academic Press), pp. 350.

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., and Egnor, S. R. (2015). Female mice ultrasonically interact with males during courtship displays. *eLife* 4:e06203. doi: 10.7554/eLife.06203

Nyby, J. (1983). Ultrasonic vocalizations during sex behavior of male house mice (Mus-musculus) - a description. *Behav. Neural Biol.* 39, 128–134. doi: 10.1016/s0163-1047(83)90722-7

Orefice, L. L., Mosko, J. R., Morency, D. T., Wells, M. F., Tasnim, A., Mozeika, S. M., et al. (2019). Targeting peripheral somatosensory neurons to improve tactile-related phenotypes in ASD models. *Cell* 178, 867–886. doi: 10.1016/j.cell.2019.07.024

Pagani, M., Bertero, A., Liska, A., Galbusera, A., Sabbioni, M., Barsotti, N., et al. (2019). Deletion of autism risk gene Shank3 disrupts prefrontal connectivity. *J. Neurosci.* 39, 5299–5310. doi: 10.1523/JNEUROSCI.2529-18.2019

Palanza, P., Della Seta, D., Ferrari, P. F., and Parmigiani, S. (2005). Female competition in wild house mice depends upon timing of female/male settlement and kinship between females. *Anim. Behav.* 69, 1259–1271. doi: 10.1016/j.anbehav.2004.09.014

Peca, J., Feliciano, C., Ting, J. T., Wang, W., Wells, M. F., Venkatraman, T. N., et al. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442. doi: 10.1038/nature09965

Pomerantz, S. M., Nunez, A. A., and Bean, N. J. (1983). Female behavior is affected by male ultrasonic vocalizations in house mice. *Physiol. Behav.* 31, 91–96. doi: 10.1016/0031-9384(83)90101-4

Portfors, C. V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J. Am. Assoc. Lab. Anim. Sci.* 46, 28–34.

Sangiamo, D. T., Warren, M. R., and Neunuebel, J. P. (2020). Ultrasonic signals associated with different types of social behavior of mice. *Nat. Neurosci.* 23, 411–422. doi: 10.1038/s41593-020-0584-z

Scattoni, M. L., Ricceri, L., and Crawley, J. N. (2011). Unusual repertoire of vocalizations in adult BTBR T plus tf/J mice during three types of social encounters. *Genes Brain Behav.* 10, 44–56. doi: 10.1111/j.1601-183X.2010.00623.x

Schmeisser, M. J., Ey, E., Wegener, S., Bockmann, J., Stempel, A. V., Kuebler, A., et al. (2012). Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature* 486, 256–260. doi: 10.1038/nature11015

Sewell, G. D. (1970). Ultrasonic communication in rodents. *Nature* 227:410. doi: 10.1038/227410a0

Song, T.-J., Lan, X.-Y., Wei, M.-P., Zhai, F.-J., Boeckers, T. M., Wang, J.-N., et al. (2019). Altered behaviors and impaired synaptic function in a novel rat model with a complete Shank3 deletion. *Front. Cell. Neurosci.* 13:111. doi: 10.3389/fncel.2019.00111

Speed, H. E., Kouser, M., Xuan, Z., Reimers, J. M., Ochoa, C. F., Gupta, N., et al. (2015). Autism-associated insertion mutation (InsG) of Shank3 Exon 21 causes impaired synaptic transmission and behavioral deficits. *J. Neurosci.* 35, 9648–9665. doi: 10.1523/JNEUROSCI.3125-14.2015

Tu, Z., Zhao, H., Li, B., Yan, S., Wang, L., Tang, Y., et al. (2019). CRISPR/Cas9-mediated disruption of SHANK3 in monkey leads to drug-treatable autism-like symptoms. *Hum. Mol. Genet.* 28, 561–571. doi: 10.1093/hmg/ddy367

Vicidomini, C., Ponzoni, L., Lim, D., Schmeisser, M. J., Reim, D., Morello, N., et al. (2017). Pharmacological enhancement of mGlu5 receptors rescues behavioral deficits in SHANK3 knock-out mice. *Mol. Psychiatry* 22, 689–702. doi: 10.1038/mp.2016.30

von Merten, S., Hoier, S., Pfeifle, C., and Tautz, D. (2014). A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (Mus musculus domesticus). *PLoS One* 9:e97244. doi: 10.1371/journal.pone.0097244

Wang, X., Bey, A. L., Katz, B. M., Badea, A., Kim, N., David, L. K., et al. (2016). Altered mGluR5-Homer scaffolds and corticostriatal connectivity in a Shank3 complete knockout model of autism. *Nat. Commun.* 7:11459. doi: 10.1038/ncomms11459

Wang, X., McCoy, P. A., Rodriguiz, R. M., Pan, Y., Je, H. S., Roberts, A. C., et al. (2011). Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. *Hum. Mol. Genet.* 20, 3093–3108. doi: 10.1093/hmg/ddr212

Warburton, V. L., Stoughton, R., Demaine, C., Sales, G. D., and Milligan, S. R. (1988). Long-term monitoring of mouse ultrasonic vocalizations. *Physiol. Behav.* 44, 829–831. doi: 10.1016/0031-9384(88)90071-6

Warren, M. R., Clein, R. S., Spurrier, M. S., Roth, E. D., and Neunuebel, J. P. (2020). Ultrashort-range, high-frequency communication by female mice shapes social interactions. *Sci. Rep.* 10:2637. doi: 10.1038/s41598-020-59418-0

Warren, M. R., Sangiamo, D. T., and Neunuebel, J. P. (2018a). High channel count microphone array accurately and precisely localizes ultrasonic signals from freely-moving mice. *J. Neurosci. Methods* 297, 44–60. doi: 10.1016/j.jneumeth.2017.12.013

Warren, M. R., Spurrier, M. S., Roth, E. D., and Neunuebel, J. P. (2018b). Sex differences in vocal communication of freely interacting adult mice depend upon behavioral context. *PLoS One* 13:e0204527. doi: 10.1371/journal.pone.0204527

Warren, M. R., Spurrier, M. S., Sangiamo, D. T., Clein, R. S., and Neunuebel, J. P. (2021). Mouse vocal emission and acoustic complexity do not scale linearly with the size of a social group. *J. Exp. Biol.* 224:jeb239814. doi: 10.1242/jeb.239814

Yang, M., Bozdagi, O., Scattoni, M. L., Woehr, M., Roullet, F. I., Katz, A. M., et al. (2012). Reduced excitatory neurotransmission and mild autism-relevant phenotypes in adolescent Shank3 null mutant mice. *J. Neurosci.* 32, 6525–6541. doi: 10.1523/JNEUROSCI.6107-11.2012

Yoo, Y.-E., Yoo, T., Lee, S., Lee, J., Kim, D., Han, H.-M., et al. (2019). Shank3 mice carrying the human Q321R mutation display enhanced self-grooming, abnormal electroencephalogram patterns and suppressed neuronal excitability and seizure susceptibility. *Front. Mol. Neurosci.* 12:155. doi: 10.3389/fnmol.2019.00155

Zippelius, H.-M., and Schleidt, W. M. (1956). Ultraschall-Laute bei jungen Mäusen. *Naturwissenschaften* 43:502. doi: 10.1007/BF00632534

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Behavioral Neuroscience

REVIEW
published: 20 December 2021
doi: 10.3389/fnbeh.2021.811737

Toward a Computational Neuroethology of Vocal Communication: From Bioacoustics to Neurophysiology, Emerging Tools and Future Directions

Tim Sainburg[1,2*] *and Timothy Q. Gentner*[1,3,4,5]

[1] *Department of Psychology, University of California, San Diego, La Jolla, CA, United States,* [2] *Center for Academic Research & Training in Anthropogeny, University of California, San Diego, La Jolla, CA, United States,* [3] *Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA, United States,* [4] *Neurobiology Section, Division of Biological Sciences, University of California, San Diego, La Jolla, CA, United States,* [5] *Kavli Institute for Brain and Mind, University of California, San Diego, La Jolla, CA, United States*

Recently developed methods in computational neuroethology have enabled increasingly detailed and comprehensive quantification of animal movements and behavioral kinematics. Vocal communication behavior is well poised for application of similar large-scale quantification methods in the service of physiological and ethological studies. This review describes emerging techniques that can be applied to acoustic and vocal communication signals with the goal of enabling study beyond a small number of model species. We review a range of modern computational methods for bioacoustics, signal processing, and brain-behavior mapping. Along with a discussion of recent advances and techniques, we include challenges and broader goals in establishing a framework for the computational neuroethology of vocal communication.

Keywords: neuroethology, computational neuroethology, denoising, vocalization, UMAP, sequence model, morph

OPEN ACCESS

Edited by:
Susanna Pietropaolo, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Andrea M. Simmons, Brown University, United States
Andrew Bass, Cornell University, United States

***Correspondence:**
Tim Sainburg
tsainburg@ucsd.edu

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *09 November 2021*
Accepted: *29 November 2021*
Published: *20 December 2021*

Citation:
Sainburg T and Gentner TQ (2021) Toward a Computational Neuroethology of Vocal Communication: From Bioacoustics to Neurophysiology, Emerging Tools and Future Directions. Front. Behav. Neurosci. 15:811737. doi: 10.3389/fnbeh.2021.811737

1. INTRODUCTION

Over the past several years emerging methods have enabled biologists to capture and quantify ethological data in ways that yield new insights into the structure and organization of behavior. These methods capitalize on two advances: the ability to record and annotate very-large behavioral datasets, and the use of new computational tools to reveal structure within and between these datasets. The ethological and neuro-ethological study of animal communication has a long history, and its future stands to benefit greatly from these new methods. Here, we discuss this emerging set of tools available to the animal communication researcher. We contextualize these computational methods within the emerging field of computational ethology more broadly and discuss how these tools can be applied in behavior and neurophysiology.

Many of the challenges that exist in the computational neuroethology of vocal behavior are neither new nor unique and parallel those in other areas of human and animal behavior. For example, the algorithmic discovery of vocal units and sequential organization in animal communication parallels the zero-speech challenge in language acquisition: given limited sensory information, can we build a system that discovers subwords, words, and sequential and syntactic organization present in speech (Versteegh et al., 2015). In animal communication the challenge

is similar: can we infer vocal segment boundaries, categories, and temporal organization from the physical and temporal characteristics of the signal. The computational neuroethology of vocal communication also parallels the emerging field of motion sequencing and the mapping behavioral kinematics, where new technologies allowing researchers to map postures and behavioral kinematics have facilitated new understandings of behavioral dynamics across scales (Anderson and Perona, 2014; Berman, 2018; Brown and De Bivort, 2018; Christin et al., 2019; Datta et al., 2019; Pereira et al., 2019). It is the goal of computational neuroethology to not only develop an understanding of the organization of behaviors, but also the neural and cognitive mechanisms that facilitate behavior. This review synthesizes work from several fields including bioacoustics, systems neuroscience, and computational neuroethology to discuss emerging methodologies and frameworks which span these fields and are available to vocal communication researchers.

The review begins with considerations in bioacoustics and signal processing and then shifts to a consideration of acoustic structure, sequential organization, and eventually to mapping the acoustic and sequential structure of vocal communication to neurophysiology correlates of behavior and perception. Throughout our review of current approaches, we relay ongoing challenges, discuss future directions, and attempt to give practical advice on vocal analyses.

2. SIGNAL PROCESSING AND DENOISING

Recorded sounds typically contain a mixture of both relevant and irrelevant components. Computational ethology often relies on modeling structure in data without making assumptions about the relevant features. Thus it is often important to remove irrelevant features (i.e., background noise) prior to analysis. Ones operationalization of noise can vary based upon the end goal of the analysis. A simple example is band-pass filtering: because vocalizations typically occur in a confined frequency range, it is reasonable to consider signal outside of that range noise and filter it away. When a recording contains vocalizations from two animals, a songbird with song in a high-frequency range, and heterospecific calls in a low-frequency range, if the subject of interest is the songbird, a simple high-pass filter can be applied to attenuate the non-target calls. When frequency ranges overlap between signal and noise, however, the problem of noise reduction becomes more difficult.

2.1. Noise Reduction

Determining what constitutes noise in recordings is non-trivial and impacts what type of noise reduction algorithm can and should be used. In a systematic review of noise reduction methods in bio-acoustics, Xie et al. (2020) outline six classes of noise reduction algorithms used for bio-acoustics: (1) Optimal FIR filter (e.g., Kim et al., 2000), (2) spectral subtraction (e.g., Boll, 1979; Kiapuchinski et al., 2012; Sainburg et al., 2020b), (3) minimum-mean square error short-time spectral amplitude estimator (MMSE-STSA; e.g., Ephraim and Malah, 1984; Alonso et al., 2017; Brown et al., 2017) (4) wavelet based denoising (e.g., Ren et al., 2008; Priyadarshani et al., 2016) (5) image processing based noise reduction, and (6) deep learning based noised reduction. These noise reduction algorithms can be broadly divided into two categories: stationary and non-stationary noise reduction (**Figure 1A**). Stationary noise reduction acts on noise that is stationary in intensity and spectral shape over time, such as the constant hum of electronics. Non-stationary noise reduction targets background noise that is non-stationary and can fluctuate in time, like the on-and-off presence of a plane flying overhead (**Figure 1B**). Stationary noise reduction algorithms operationalize noise as stationary signals, for example, the constant hum from a nearby electronic device in a laboratory setting, or insect noise in a field setting.

One approach to stationary noise reduction is spectral gating, a spectral-subtraction algorithm (e.g., Kiapuchinski et al., 2012; Sainburg et al., 2020b). The general notion is that for each frequency component of the signal, any time-frequency component below a threshold is discarded as noise. Spectral gating computes the mean and standard deviation of each frequency channel of a Short-Time Fourier Transform (STFT) of a signal (e.g., a spectrogram) and optionally a noise clip. A threshold, or gate, for each frequency component is then set at some level above the mean (e.g., three standard deviations). This threshold determines whether a time-frequency component in the spectrogram is considered signal or noise. The spectrogram is then masked based upon this threshold and inverted (with an inverse STFT) back into the time domain.

2.2. Non-stationary Noise Reduction

While stationary noise reduction algorithms can operationalize noise as any stationary acoustic signal, non-stationary algorithms vary in how they determine what is signal and what is noise. Non-stationary noise can be more challenging to remove because it can be difficult to algorithmically define the difference between signal and noise. Because the hum of a computer in the background of a lab-recording is stationary, it can be defined as noise and can be readily removed. A bird hopping around its cage can produce time-varying sounds in the same frequency range as song, making it especially pernicious.

One approach for determining the boundary between signal and non-stationary noise is to determine the timescale on which the signal acts and treat anything outside of that timescale as noise. For example, zebra finch motifs are generally between 0.5 and 1.5 s long repeated one to four times (Bruno and Tchernichovski, 2019). Any acoustic event that is outside of that time range could be considered noise. Spectral gating can be extended to non-stationary noise reduction by computing a variable gate based upon the current estimate of background noise. In the Python package noisereduce (Sainburg, 2019), this background estimate is computed using a time-smoothed spectrogram (using a forward and backward IIR filter) on a timescale parameterized by the expected signal length, an approach motivated by the Per-Channel Energy Normalization algorithm (outlined in Section 3). An example of this is given in **Figure 1**, where stationary and non-stationary spectral gating noise reduction is applied to birdsong with an airplane noise occurring in the background of the middle of the recording. Because the airplane noise is non-stationary, The stationary

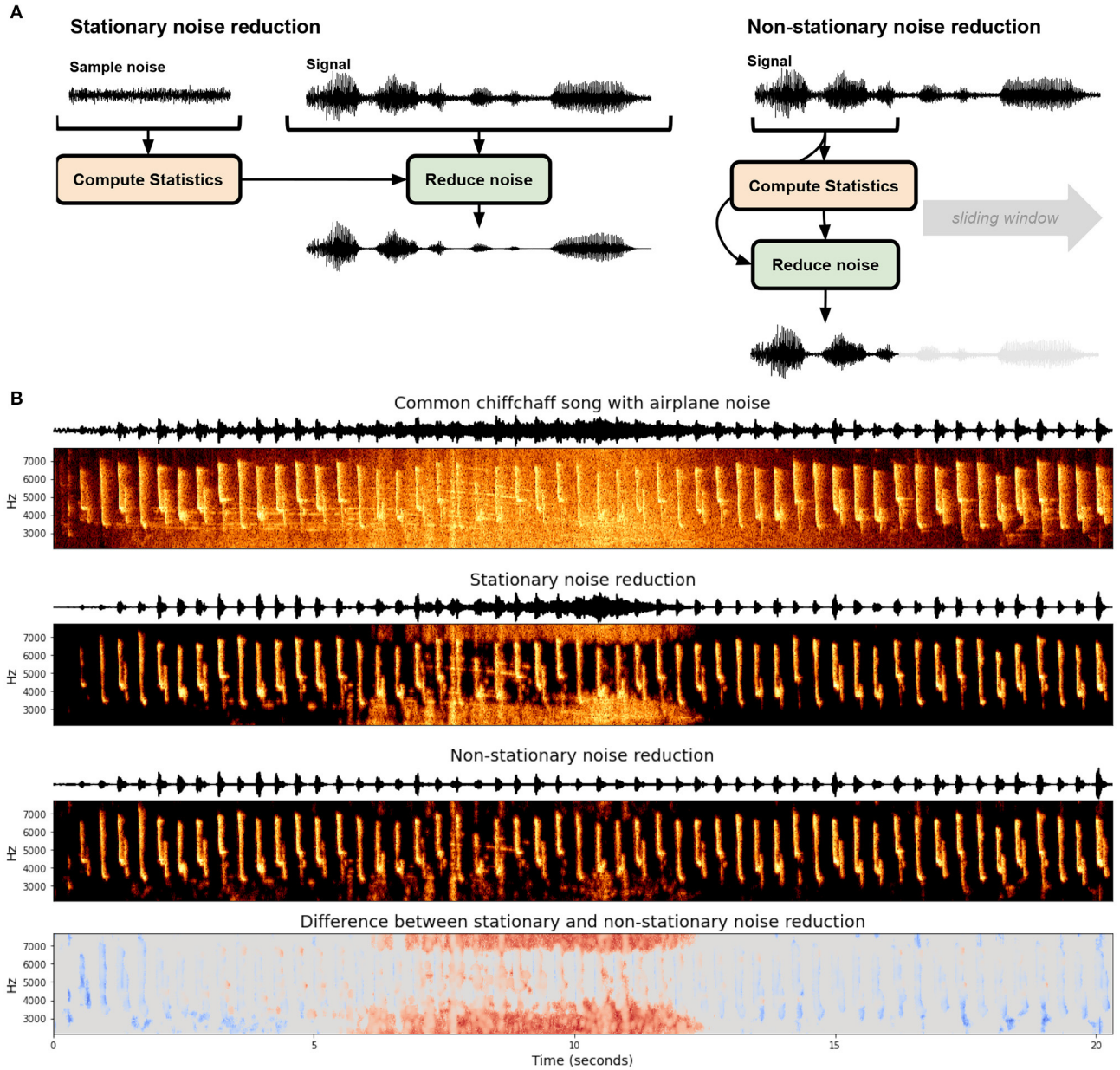


FIGURE 1 | Stationary and non-stationary spectral gating noise reduction. **(A)** An overview of each algorithm. Stationary noise reduction typically takes in an explicit noise signal to calculate statistics and performs noise reduction over the entire signal uniformly. Non-stationary noise reduction dynamically estimates and reduces noise concurrently. **(B)** Stationary and non-stationary spectral gating noise reduction using the noisereduce Python package (Sainburg, 2019) applied to a Common chiffchaff (*Phylloscopus collybita*) song (Stowell et al., 2019) with an airplane noise in the background. The bottom frame depicts the difference between the two algorithms.

approach fails in two ways relative to the non-stationary approach: the airplane noise is not fully successfully gated at its peak in the middle of the recording (shown as red in the bottom panel) and weaker syllables of song are treated as noise and reduced in the beginning and end of the clip (shown in blue in the bottom panel). Advantages of non-stationary noise reduction are not unique to acoustic noise: when we know the timescale of a signal we can use the same non-stationary principles to remove noise occurring at different timescales. For example in the continuous recording of neural data, action potentials occur within the range of 1 ms. Events occurring over tens or hundreds of milliseconds can therefore be treated as noise.

2.3. Reducing Noise With Deep Learning

A promising future avenue for noise reduction is in explicitly training machine learning algorithms to mask or remove

noise, as is done in speech enhancement and segregation (Wang and Chen, 2018). At present, however, deep learning based noise reduction has not been utilized directly in bio-acoustics (Xie et al., 2020). Xie et al. (2020) attribute this to a lack of utility when using denoising in some applications of deep learning-based bio-acoustics detection (Kong et al., 2017). The utility of noise reduction exists beyond classification tasks, however. For example, computing spectral features and acoustic similarity between vocalizations can be susceptible to background noise. Recent work by Stowell et al. (2019) shows that manipulating datasets by superimposing background environment noise on vocal datasets can reduce confounds and improve identification across recording conditions. Similar approaches could be used to remove noise. For example, spectral gating could be extended with neural networks by training a neural network to learn a mask to gate away background noise and recover the lower-noise spectrogram, as has been done in speech enhancement applications (Wang and Chen, 2018; Lee and Kim, 2020).

It is also important to consider what information is being removed by pre-processing techniques such as denoising. Pre-processing methods throw away potentially valuable information that will influence downstream analyses. De-noising vocal data without careful consideration can remove lower amplitude syllables of birdsong or infrequent vocalizations outside of the expected frequency range.

3. SIGNAL REPRESENTATION

An important consideration in any analysis pipeline is how to represent the data that goes in. Animal vocalizations are typically recorded using one or more microphones at a sampling rate that can capture the full spectral range of the vocalization. Performing analyses directly upon recorded waveforms is not always optimal for capturing informative structure in vocal data, however. Waveforms are high-dimensional representations of audio that can make it difficult for learning algorithms to capture time-frequency structure in vocalizations. Spectro-temporal representations can be both lower-dimensional, and more explicitly capture complex time-frequency relationships in vocalizations.

Spectrograms are, at present, the most common form of vocalization representation, both for visualization and as input to learning algorithms, both in bio-acoustics and speech. When representing vocal data with a spectrogram, the parameters used to compute the spectrogram can have an important influence on the performance of the algorithm (Elie and Theunissen, 2016; Knight et al., 2020). The most important parameterization of spectrograms is the trade-off between temporal and frequency resolution when computing a spectrogram, a result of the Heisenburg Uncertainty Principle (Gardner and Magnasco, 2006; Moca et al., 2021). For example, three spectrograms are shown in **Figures 2A–C** with different windows used to compute the Short-Time Fourier Transform. The first has an intermediate-sized window with intermediate time and frequency resolution (**Figure 2A**), the second uses a short window with high time-resolution and low frequency-resolution (**Figure 2B**), and the third uses a long window with high frequency-resolution but low time-resolution (**Figure 2C**).

A number of approaches exist to improve time and frequency resolution. Time-frequency reassigned spectrograms attempt to improve time-frequency resolution using additional information from the phase spectrum (**Figure 2D**) (Fulop and Fitz, 2006; Gardner and Magnasco, 2006; Xiao and Flandrin, 2007). Wavelet transforms (**Figure 2E**) have more recently been used in representing animal vocalizations (Main and Thornton, 2015; Priyadarshani et al., 2016, 2020; Hsu et al., 2018), and allows multi-scaled emphasis on time vs. frequency, for example emphasizing frequency resolution at lower frequencies and time-resolution at higher frequencies, intuitively because an uncertainty of 50 Hz is more relevant at 500 Hz than at 5,000 Hz. Most recently, the superlet (**Figure 2F**) enables time-frequency super-resolution by geometrically combining sets of wavelets with increasing constrained bandwidths (Moca et al., 2021).

There are also several variants of spectrograms and time-frequency representations that differentially emphasize time-frequency information. For example, log-scaling spectrograms in frequency emphasizes lower frequency ranges over higher frequency ranges, which parallels both the cochlea and perception (Eldredge et al., 1981). Mel-scaling (**Figure 2G**), is a form of log-scaling fit to fit human perception (Stevens et al., 1937), though the specific scaling range relative to human perception are imperfect (Greenwood, 1997). Mel-Frequency Cepstral Coefficients (MFCCs; **Figure 2H**) additionally compute the Discrete Cosine Transform on the Mel-spectrogram, and were, until recently, commonly used for speech recognition because they are generally robust to noise and emphasize the frequency range of speech (**Figure 2H**) (Muda et al., 2010). Another model, directly relevant to physiology, is the Cochleagram (Brown and Cooke, 1994; Feather et al., 2019; Rahman et al., 2020). Cochleagrams mimic the cochlea by using a filter bank associated with points on the basilar membrane to mimic an impulse response **Figure 2I**).

A new approach that has shown much promise in bio-acoustics is Per-Channel Energy Normalization (PCEN; **Figure 2J**; Wang et al., 2017; Lostanlen et al., 2018). Lostanlen et al. (2018) identify three advantages of PCEN: (1) temporal integration, (2) adaptive gain control, and (3) dynamic gain compression. Temporal integration estimates the background noise at each frequency band. Adaptive gain control then adapts the gain of the spectral representation. Finally, dynamic range compression adaptively shifts the range of low and high amplitude components of the signal. Adaptive gain control is ubiquitous to mammalian auditory processing and is also often used in cochleagrams (Rahman et al., 2020). PCEN has been shown to aid in enhancing animal vocalizations relative to background noise across distances from the microphone (Lostanlen et al., 2019a) and reduce biases in bio-acoustics background settings such as dawn vs. dusk (Lostanlen et al., 2019b).

Descriptive basis-features features can also be used to represent vocalizations for downstream analyses. One challenge with using basis-features for vocal analysis is in determining

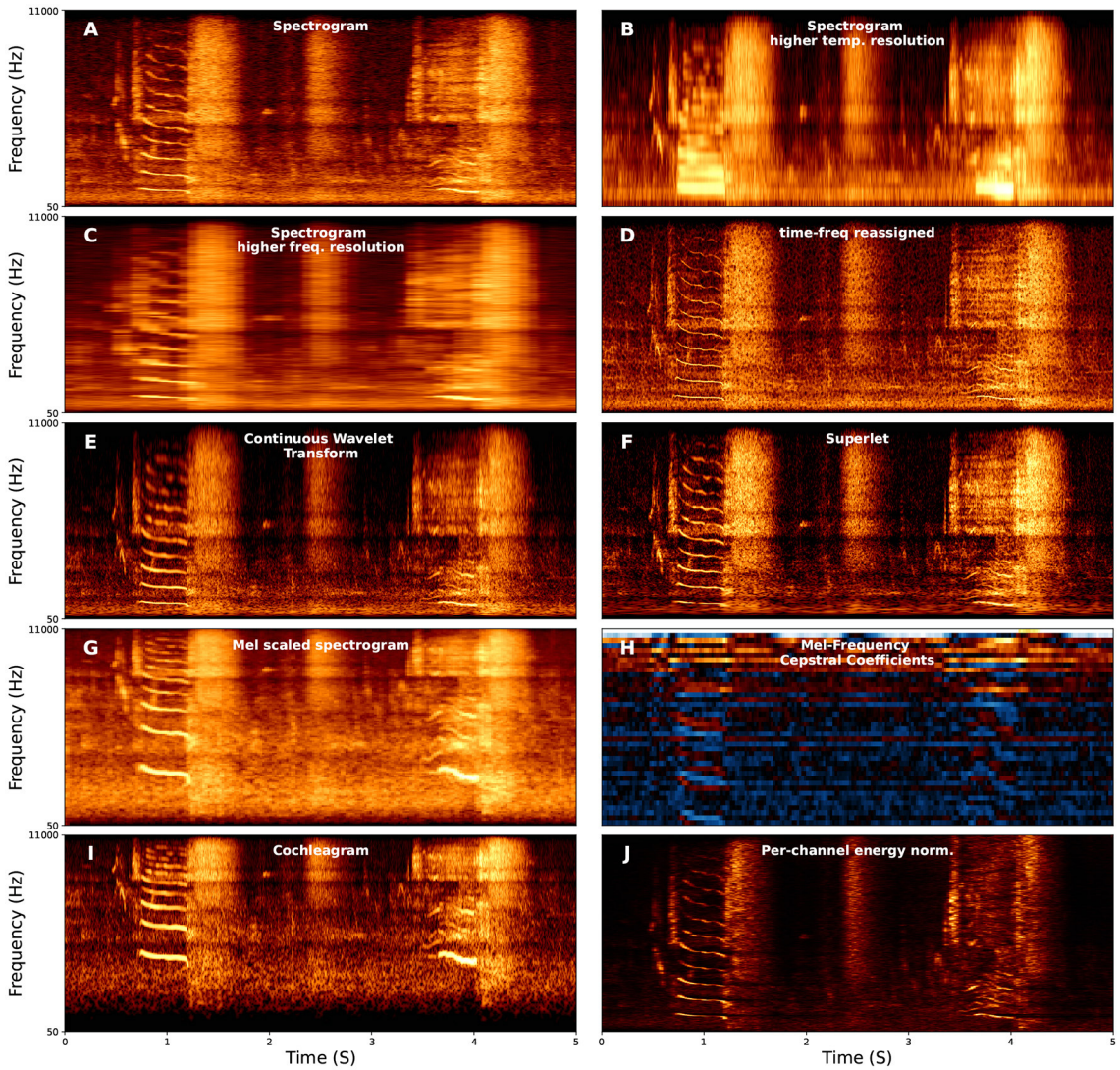


FIGURE 2 | Examples of several different spectral representations of a five-second red deer (*Cervus elaphus*) vocalization. For each axis, the *x*-axis corresponds to time, and the *y*-axis corresponds to frequency. The y-axis corresponds to frequency and is linearly spaced in **(A–F,J)** between 50 and 11,000 Hz and log-spaced in the same range for **(G–I)**. **(F)** Continuous Wavelet Transform using the Morlet (i.e., Gabor) wavelet.

what basis-features are relevant (Tchernichovski et al., 2000; Elie and Theunissen, 2016). Very few species have been rigorously examined to determine what acoustic features distinguish vocal units (Elie and Theunissen, 2016; Kershenbaum et al., 2016). Swamp sparrow notes, for example, are relatively simple vocalizations and can be well-described using just the length of the note, the peak frequency at the start of the note, and the peak frequency at the end of the note (Clark et al., 1987). One approach to determining what features are relevant in a vocal signal is to train classifiers to predict behaviorally-relevant information, such as individual identity, age, or the activity the animal is engaged on a full set to basis features, and retain those features which are highly informative (Elie and Theunissen, 2016, 2018).

4. IDENTIFYING, SEGMENTING, AND LABELING VOCALIZATIONS

Vocalization data can be recorded in a number of different settings, ranging from single individuals in well-controlled and acoustically isolated lab settings to multi-individual and multi-species recordings taken next to a busy highway. When vocalizations are produced by isolated, single individuals, segmenting out vocalizations can often be performed simply by thresholding the vocal envelope and assuming any detected noise events that match the statistics of the vocalizing animal (e.g., frequency and length of vocalization) are vocalizations (Tchernichovski et al., 2000). More complex environments and

species with more complex vocal structure require more complex solutions (Priyadarshani et al., 2018).

Experimental paradigms in neuroethology differ from bio-acoustics in that environmental sounds can usually be controlled, but are still faced with the challenge of often being made in colony settings with multiple vocalizing individuals or individuals who make non-vocal sounds such as interaction with a living space. Regardless of context, recent advances in machine learning algorithms for passive monitoring of acoustic environments allow for real-time labeling of species and individuals in noisy environments.

Automatic vocalization annotation can be broken down into three related tasks: identification, segmentation, labeling. Identifying refers to what animal is vocalizing and at what times and frequency channels. Segmentation refers to the segmentation of vocalizations into their constituent units, labeling then refers to grouping units into discrete element categories. A spectrogram outlining all three tasks is given in **Figure 3A**. Two individuals in the target species, Australian pied butcherbird *Cracticus nigrogularis* are vocalizing over top of background noise from another, unidentified, species of songbird, as well as an unidentified species of insect. Each bird's song can be divided into segmental units (notes) which can be further categorized into discrete element categories ("A," "B," "C," ...). In such a dataset, labeling challenges occur over multiple levels: identifying the species, identifying the individual, segmenting vocal units, and labeling vocal units into discrete categories. Some algorithms perform only one of these steps at a time, while others perform all three.

4.1. Detecting Species and Individuals

To detect species in continuous bio-acoustic data, several open-source tools and datasets have recently been made available for passive acoustic monitoring. A summary of many of these software and their features are given in Priyadarshani et al. (2018, Table 4). Over the past few years machine learning competitions challenging researchers to produce species recognition algorithms have motivated an increasing number of open-source approaches to bioacoustic sound recognition (e.g., Lasseck, 2013; Murcia and Paniagua, 2013; Goëau et al., 2014; Stowell et al., 2019). The same tools can be applied to differentiating between individuals in the same recording environment (e.g., Adi et al., 2010; Mielke and Zuberbühler, 2013). Most recent approaches rely on deep neural networks to detect vocalizations in noisy environments (e.g., Stowell et al., 2019; Cohen et al., 2020a). Current neural networks generally rely on some combination of convolutional filters in the temporal-frequency space of spectrograms (Convolutional Neural Networks or CNNs, **Figure 3B**) and temporal-recurrence (Recurrent Neural Networks, or RNNs, **Figure 3C**). Convolutional filters in the time-frequency space of spectrograms allow neural networks to learn complex spectro-temporal features used to classify sounds (**Figure 3B**). Temporal recurrence allows neural networks to learn sequential and temporal relationships that unfold over long time delays (**Figure 3C**). In combination, recurrent and convolutional architectures allow complex, non-linear spectrotemporal features that occur over arbitrary timescales to be captured by neural network architectures.

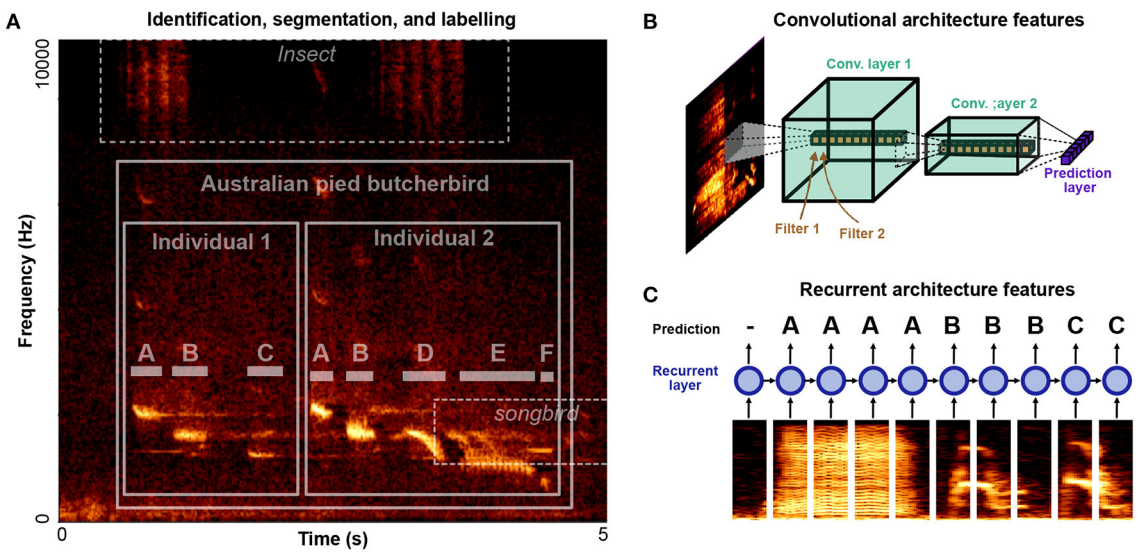


FIGURE 3 | Levels of organization and architectural design features for identifying, segmenting, and labeling vocalizations. **(A)** An example clip of Australian pied butcherbird song (Janney et al., 2016) is shown containing two male butcherbirds, alongside background noise containing insect noise and another bird's song. A classification task exists at several levels: identifying the target species, differentiating individuals, detecting note boundaries, and classifying notes. **(B)** A cartoon diagram of a convolutional neural network architecture applied to a spectrogram. Convolutional filters are applied in time-frequency space. Deeper layers have larger spectrotemporal receptive fields and learn more complex filters. **(C)** A cartoon diagram of a recurrent neural network applied to a song spectrogram. Spectral slices are input to recurrent layers in the network (depicted as a circle) which are recurrent in time, allowing information to be integrated over time.

4.2. Segmenting and Labeling Vocal Units

Beyond identifying individuals and species, many analyses of vocal communication rely on the temporal segmentation and categorization of vocalizations into discrete units. Unlike identifying species or individuals, where an objective measure exists of what animal produced a vocalization, the segmental units that comprise animal vocalizations are less well-defined. In comparison to human language, where linguistic units are determined based on their functional role, substantially less is known about the function each vocal unit plays in most species' communication, or even what should define the beginning and ending of a vocal unit (Kershenbaum et al., 2016; Mizuhara and Okanoya, 2020). Analyses of most animals, therefore, rely on easily discernible physical features of vocalizations. For example in songbirds, songs are typically segmented at different hierarchical levels, though no strict definition of these levels of organization are agreed upon by all researchers. Common units of birdsong are notes, corresponding to abrupt changes in frequency, syllables, defined by periods of silence surrounding continuous vocalizations, motifs, stereotyped repetitive combinations of acoustic elements, and phrases, series of stereotyped or commonly associated syllables. Despite the ubiquity with which these terms are used, most vocal units have not been validated in terms of the species' own perceptual system, and those that do, like the Bengalese finch 'syllable' (Mizuhara and Okanoya, 2020) call into question the commonly assumed role they play in communication. It is therefore ideal, but not always feasible, to validate decisions about vocal units based upon perceptual, physiological, or functional roles those vocal units play in the animal's communication (Suzuki et al., 2006; Kershenbaum et al., 2016). Still, most analyses of animal communications rely on human perceptual decisions at some level, whether it is to label discrete classes of birdsong phrases, or determine the representational space upon which an "unsupervised" learning algorithm will discretize units (discussed in Section 5).

When vocal units are defined and vocal classes are chosen, machine learning algorithms can be used to systematize and vastly speed up the classification and segmentation of vocal units. Most commonly, supervised recognition algorithms are used, where the algorithm explicitly learns to algorithmically map acoustic data to the researcher's labeling scheme. Over the past decades, vocalization labeling algorithms have paralleled those used in other acoustic domains, such as speech and music recognition. At present, tools rely on deep neural networks. The field of deep learning has changed rapidly over the past decade, with different architectures of neural networks quickly outperforming the previous architectures (Nassif et al., 2019). Prior to deep learning, automated birdsong element recognition relied on algorithms such as Hidden Markov Models (Kogan and Margoliash, 1998), support vector machines (Tachibana et al., 2014), template matching (Anderson et al., 1996), or k-Nearest-Neighbors labeling (Nicholson, 2016), following alongside contemporary speech recognition algorithms. Like sound event detection, current approaches tend to rely on recurrent and convolutional neural network architectures. TweetyNet (Cohen et al., 2020a), for example, uses a recurrent and convolutional architecture to capture complex spectro-temporal patterns over long timescales. Future advances in neural network architectures will likely continue to follow those in speech recognition, for example, using transformer architectures (Karita et al., 2019) as well as semi-supervised and unsupervised pre-training methods such as wav2vec (Schneider et al., 2019). One important divergence between speech recognition and animal vocalization classification is the reliance upon data availability, however. An ideal animal vocalization classifier works well on very small amounts of labeled data, requiring less experimenter time, whereas speech recognition systems tend to have an abundance of data available (though speech recognition for low-resource languages may be an area to watch).

A second approach to labeling vocalizations is to actively involve the experimenter in the algorithm via human-in-the-loop labeling (e.g., Wimmer et al., 2010; Kim and Pardo, 2018). Human-in-the-loop algorithms rely on a combination of supervised and unsupervised learning. Supervised learning comprises learning algorithms that are trained with labeled data, such as classification tasks. Unsupervised learning refers to algorithms that do not require supervised labels, such as dimensionality reduction. Human-in-the-loop algorithms leverage both, by proposing an initial coarse segmentation and/or labeling of the dataset through unsupervised learning, which the human then partially revises (e.g., merging or splitting putative classes of vocalizations) via a graphical user interface (GUI). The revised data is then re-processed by the algorithm and sent back to the user to revise, until the experimenter is content with the resulting labeled dataset. Using a combination of human expertise and machine processing enables quicker labeling of large bio-acoustics data with minimal human effort. A further discussion of unsupervised algorithms is discussed below in Section 5.

5. EXTRACTING RELATIONAL STRUCTURE AND CLUSTERING

Classifying vocal elements into discrete categories (e.g., "A," "B," "C," ...) is for many analyses a necessary abstraction that enables the analysis of recurring events. At the same time, this symbolic abstraction ignores acoustic relationships both within discrete element categories and between them. For example, in **Figure 3**, are the syllables of birdsong **Figure 3A** more similar to the syllables **Figure 3B** or the syllables **Figure 3C**? Determining the relatedness (or distance) between vocalizations can enable the quantification of how vocalizations change over time (Mets and Brainard, 2018; Kollmorgen et al., 2020), how vocal repertoires differ across individuals and species (Miller, 1979; Sainburg et al., 2020b), and map and visualize broad structure present in vocal repertoires (Sainburg et al., 2020b; Goffinet et al., 2021).

5.1. Operationalizing Relatedness

Given a dataset of vocalizations segmented into discrete units, relatedness is a measure quantifying the similarity of vocalizations relative to one another. The basis for operationalizing relatedness can utilize physical properties

of signals, perceptual judgments, or behavioral and physiological responses to the signal. Most commonly, the relationships between vocal elements are computed on either spectrotemporal representations or on the basis of descriptive features of the vocalization, such as frequency modulation, fundamental frequency, and vocal envelope (Miller, 1979; Sainburg et al., 2020b; Goffinet et al., 2021).

How different aspects of the vocalization should weigh into a measure of similarity is non-trivial. No metric for similarity is objectively correct, even when metrics are derived purely from objective physical features. For example, what is the relative importance of a vocalization's duration vs. fundamental frequency in determining similarity? One ground truth metric for an algorithms judgement of similarity is its relationship with human's perceptual judgment of similarity (Tchernichovski et al., 2000), though there is no guarantee that these measures reflect the animal's own perception and physiology (Dooling and Prior, 2017). An ideal measure of similarity could be derived through careful experimentation gleaning the animal's own judgment of similarity (Kershenbaum et al., 2016), but in most cases, this task would be unfeasible and time-consuming. Even when performed carefully, perception varies from animal to animal, based upon experience (Lachlan et al., 2010).

In addition, when vocal features are continuous, accounting for differences in duration and temporal alignment requires consideration. Approaches vary from averaging over time (Elie and Theunissen, 2016), pooling using attention mechanisms (Morfi et al., 2021), using dynamic time warping (Kogan and Margoliash, 1998), and zero-padding (Sainburg et al., 2020b). Similarly, at least some animals rely on spectral shape rather than absolute pitch when recognizing acoustic objects (Bregman et al., 2016). A recent approach accounting for variability in frequency is dynamic frequency warping (Somervuo, 2019). Striking a balance between spectrotemporal tolerance and absolutely discounting spectrotemporal alignment can have substantial impact on the final measure of similarity.

5.2. Learning a Similarity Space

Once a metric for similarity is determined, that distance can be used to infer a structured representation of the relationships in a repertoire as a whole, providing a new representational space with which to quantify vocalizations.

Perhaps the most intuitive and pervading example of a learned embedding space for vocal similarity is multi-dimensional scaling [MDS, e.g., (Miller, 1979; Dooling et al., 1987; Morfi et al., 2021)]. Multi-dimensional scaling takes a graph of pairwise similarity measures between each vocalization in the dataset and attempts to find an embedding that best preserves the similarity structure of that graph. As the number of vocal elements in a dataset gets larger, however, the number of pairwise distances between vocal elements increases exponentially. This is computationally an issue because computing 10,000 pairwise distances between 100 elements is computationally feasible, but 10,000,000,000 pairwise distances between 100,000 elements is not.

Trying to preserve the pairwise relationships between every element in a dataset can also over-emphasize irrelevant relationships in vocal data. For example, if a bird's vocal repertoire comprises 10 motifs classes all produced with the same frequency, the vast majority of pairwise distance relationships computed (90%) will be between class, while only 10% of pairwise relationships computed will be within class. In many cases, both in animal communication and in dimensionality reduction more broadly, there is utility in weighing relationships between similar vocal elements more highly than relationships between less similar vocalizations. This contrast is defined in the dimensionality reduction literature as the emphasis of local vs. global structure (De Silva and Tenenbaum, 2002). Algorithms that attempt to preserve every pairwise relationship are called global dimensionality reduction algorithms, while algorithms that emphasize capturing relationships only to nearby points in dataspace (more similar vocalizations) are called local dimensionality reduction algorithms. In many vocalization datasets, emphasizing local over global structure better preserves categorical structure such as individual and call identity (Sainburg et al., 2020c; Goffinet et al., 2021; Morfi et al., 2021). A visual demonstration contrasting local and global structure preservation is given in **Figure 4A**. While global embedding algorithms like MDS attempt to preserve every pairwise relationship, local algorithms preserve only local (e.g., nearest-neighbor) relationships, capturing more within-cluster structure. In **Figures 4B–G** an example is given with macaque coo calls, in which a local structure-preserving algorithm (UMAP, described below) more clearly pulls apart clusters corresponding to individual identity than MDS.

At present, the two dominant local dimensionality reduction algorithms are UMAP and t-SNE. UMAP and t-SNE differ in several important ways beyond the scope of this paper, but their key intuition and the steps underlying the algorithms remain similar: first, compute a (nearest-neighbors) graph of pairwise relationships between nearest neighbors in the original dataset (e.g., using Euclidean distance or an arbitrary similarity metric) then, embed that graph into an embedding space via gradient descent (Sainburg et al., 2021). UMAP, in particular, has been shown to capture complex structure in vocal repertoires such as differences in vocal dialect, vocal stereotypy, vocal element categories, inter-species similarity, and individual identity, in contrast to classic methods like MDS and PCA (Goffinet et al., 2021; Morfi et al., 2021; Sainburg et al., 2021).

One challenge with graph-based dimensionality reduction algorithms like MDS, UMAP and t-SNE is that they are non-parametric dimensionality reduction algorithms, meaning they do not learn the relationship between input data (e.g., a spectrogram of the vocalization) and their embeddings. Learning a parametric relationship between vocalizations and their embeddings allows a fast mapping between data and embedding, i.e., for applications that necessitate real-time feedback such as brain-machine interfacing.

The most common parametric dimensionality reduction algorithm is PCA, where a linear transform is learned between data and an embedding space. Similarly, neural networks such as autoencoders can be used to learn a set of basis features which can be complex and non-linear (Kohlsdorf et al., 2020; Sainburg et al., 2020c; Goffinet et al., 2021; Singh Alvarado et al., 2021). For example, an autoencoder trained on images of

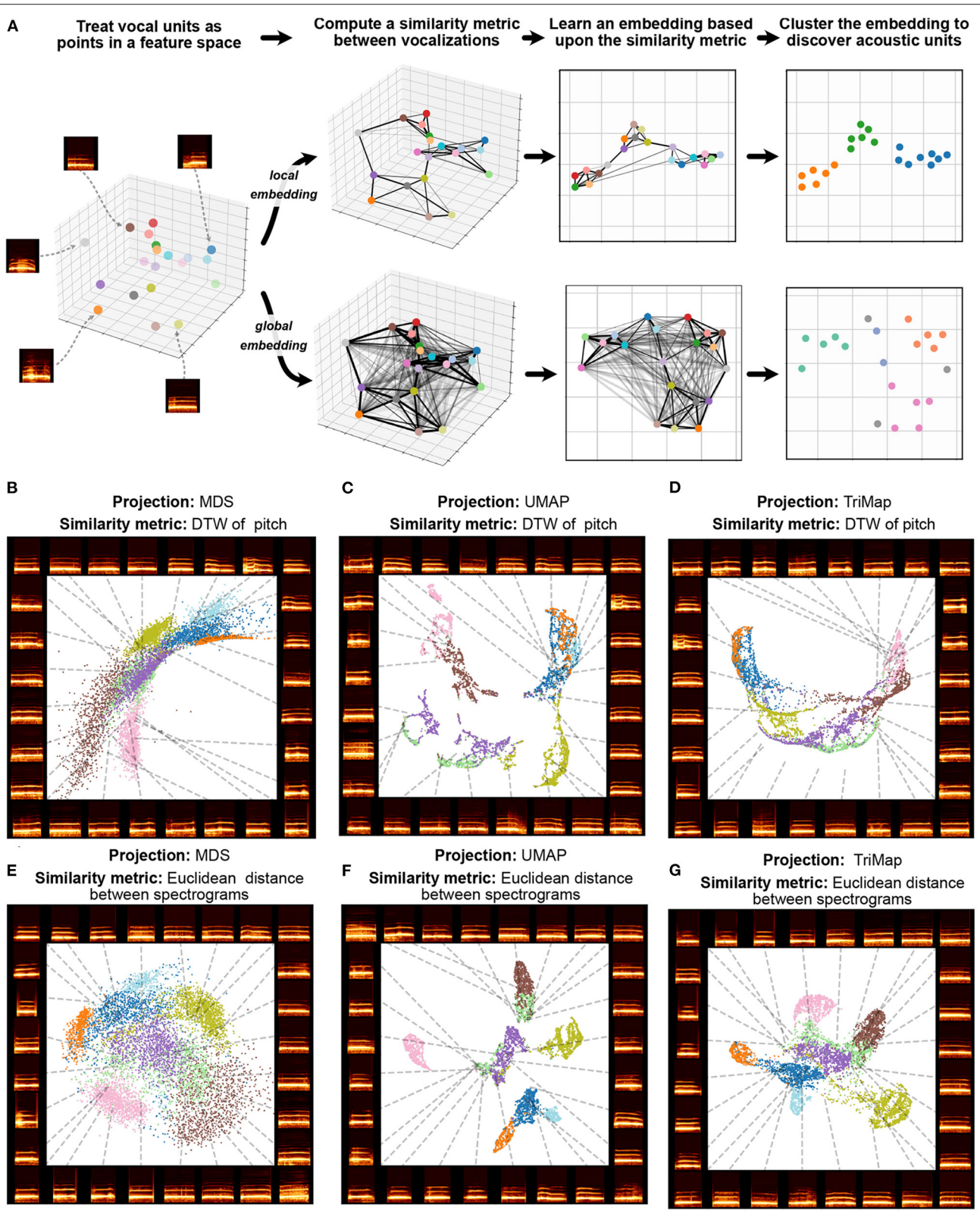


FIGURE 4 | Local and global embeddings. **(A)** The steps outlined in Section 5 exhibit the differences between the relationships preserved in local and global embeddings. **(B–D)** Projections of a dataset of macaque coo calls (Fukushima et al., 2015) using two similarity metrics (Dynamic Time Warping over frequency, and Euclidean distance between spectrograms) and three projection algorithms (Multidimensional Scaling, UMAP, and TriMap) **(E–G)**. Colors represent individual identity.

faces can learn to linearize the presence of glasses or a beard (Radford et al., 2015; Sainburg et al., 2018b, 2021). Autoencoders trained on animal vocalization data can similarly learn complex non-linear relationships in vocal data. In Section 8 we discuss how these complex learned features could be utilized in animal vocalizations to learn acoustic features such animal age, sex, and attractiveness, which can, in principle, be utilized for playback experiments.

A recent extension to UMAP, Parametric UMAP weds the advantages of UMAP with the parametric embedding of neural networks (Sainburg et al., 2021). Parametric UMAP acts by optimizing the UMAP loss function over arbitrary neural networks (e.g., convolutional recurrent networks were used with Cassin's vireo song in Sainburg et al., 2021) which can be balanced with additional losses such as MDS and autoencoding, to preserve additional global structure in UMAP projections. Parametric neural network-based approaches such as Parametric UMAP can also embed data on a similar timescale as PCA, enabling real-time applications, as opposed to non-parametric methods such as UMAP, t-SNE, and MDS.

Another class of neural network based dimensionality reduction algorithms rely on triplet-loss-based similarity preservation. Triplet-based embeddings have been used for birdsong for classification and embedding (Morfi et al., 2021; Renteria et al., 2021). Triplet networks learn an embedding space by sampling three types of vocal units: an anchor, a positive sample that is perceptually similar to the anchor point, and a negative sample that is perceptually distant from a vocal unit. The loss then encourages the positive sample to be pulled to the anchor, and the negative sample to be pushed further away. For example, Morfi et al. (2021) describe a triplet-loss-based network trained to produce vocal embeddings based upon a metric of perceptual distances. Like graph-based dimensionality reduction algorithms, triplet-loss-based embeddings rely on a pre-defined experimenter-determined notion of distance. Morfi et al. suggest a forthcoming animal-defined metric but in-lieu use a descriptive feature-based metric in the software Luscinia (Lachlan, 2007) which is correlated with human perceptual judgments of zebra finch song (Holveck et al., 2008).

5.3. Finding Latent Units Through Clustering

Learned embedding spaces enable the inference of broad structure acoustic structure from the statistics of vocalizations, enabling further downstream discovery of vocal units based upon distributional properties in embedding spaces (Kershenbaum et al., 2016; Sainburg et al., 2020b; Keen et al., 2021). Unsupervised clustering of vocal elements lies in contrast with supervised learning, where class labels are determined by experimenters, as in Section 4. Sainburg et al. (2020b) observe that labels obtained by clustering UMAP embeddings of Cassin's vireo and Bengalese finch syllables are more similar to experimenter labels than clustering PCA projections or spectrograms. Further, these latent projections capture additional acoustic and syntactic structure than the ground truth experimenter labels. In addition to acoustic structure, vocal elements can be clustered on the basis of syntactic organization. For example, incorporating transition information through Partially observable Markov Models (POMMs; Jin and Kozhevnikov, 2011) and Hidden Markov Models (HMMs; Katahira et al., 2011; Sainburg et al., 2020b) into a labeling scheme for birdsong better explains sequential structure than hand-labels or clustering without reference to temporal sequencing. An alternative approach is to perform clustering prior to embedding, directly upon the inferred relational graph (Frasier et al., 2017).

One challenge in unsupervised vocal unit discovery through methods such as UMAP embeddings is their reliance upon pre-defined vocal unit temporal boundaries. Although clustering on latent projections enables an unsupervised extraction of vocal categories from segmental units, latent projections rely on a pre-defined temporal segmentation of acoustic units from the vocal stream. In some species, atomic vocal units can be determined by clearly defined physical features of the signal, like long pauses between syllables, however, even in the case of clearly defined physical features, those units are not necessarily the base units of perception (Mizuhara and Okanoya, 2020). An open issue in vocal analysis is the unsupervised temporal segmentation of vocalizations into elements when clear physical boundaries are not available. This problem parallels both unsupervised speech discovery (i.e., ZeroSpeech), and the challenge of discovering behavior units in other areas of computational neuroethology (e.g., Motion Sequencing). In speech, phonemes are not clearly defined by physical characteristics, thus approaches for segmentation rely upon a combination of temporal and distributional information alongside imposed priors. Ongoing efforts in unsupervised speech segmentation, syllabic unit discovery, and word discovery can motivate parallel approaches in animal communication. In addition, physiological and kinematic measures such as articulation and breathing rate can aid in determining vocal boundaries. In computational neuroethology, new methods in tracking behavioral kinematics provide similar continuous behavioral datasets to those discussed in this paper (e.g., Wiltschko et al., 2015, 2020; Berman et al., 2016; Mathis et al., 2018; Pereira et al., 2019; Dunn et al., 2021; Marshall et al., 2021). For example, MoSeq (Wiltschko et al., 2015, 2020) discovers animal behavioral states using depth camera recordings of animals by fitting the behavioral data to an Autoregressive Hiden Markov Model. They find stereotyped sub-second mouse behavioral states, dubbed syllables, that underlie a syntax of behavior, much like birdsong. Communicative behavior is also not produced solely in the auditory domain. Improving methods for uncovering structure in animal behavior more broadly will facilitate research on the interaction between multi-sensory and multi-modal vocal behavior, like the dances that accompany many bird songs (Williams, 2001).

5.4. Data Augmentation

Another approach that is largely underutilized in bio-acoustic vocal recognition algorithms is data augmentation, an approach that is currently used in most state-of-the-art machine perception applications. In automatic speech recognition, for example, several current state-of-the-art approaches (e.g., Baevski et al.,

2020; Gulati et al., 2020) use SpecAugment (Park et al., 2019) in which the classifier learns a policy of various augmentations such as warping and masking frequency channels in time. Lostanlen et al. (2019b) demonstrate the utility of augmenting bio-acoustics datasets with diverse background acoustics to facilitate better generalization across environments and conditions. Augmentation in settings where little labeled data are available has also proven successful on several semi-supervised learning benchmarks (e.g., Berthelot et al., 2019). One difficulty with performing data augmentation with bio-acoustics data, however, is the extent to which slight manipulations can affect the perceptual class that vocalizations fall into (Morfi et al., 2021).

6. INFERRING TEMPORAL AND SEQUENTIAL STRUCTURE

Identifying sequential organization typically relies upon the abstraction of vocalizations into discrete sequences of elements, effectively treating vocal data as corpora from which to perform symbolic analyses. Kershenbaum et al. (2016) identify six classes of models and analyses for analyzing temporal sequences: Markov chains, Hidden Markov Models, Network-based analyses, Formal grammars analyses, and temporal models. Analyses of temporal organization in animal communication has traditionally been largely influenced by Chomsky's hierarchy of formal grammars, with a focus on trying to understand what class of the Chomsky hierarchy animal's behaviors belong within (Hauser et al., 2002; Rohrmeier et al., 2015; Jiang et al., 2018; Morita and Koda, 2019). For example, Markov models, Hidden Markov Models, and Network models are all finite-state models in the Chomsky hierarchy.

6.1. Short-Timescale Organization and Graphical Analysis

Broadly, analyses over vocal organization can be broken down into two classes: analyses over short- and long-distance (i.e., short- and long-timescale) sequential organization. Short-timescale analyses are concerned with relationships between adjacent, or near adjacent elements in a sequence. Markov models, for example, capture short-timescale dynamics of vocal communication. A typical Markov model of birdsong is simply a

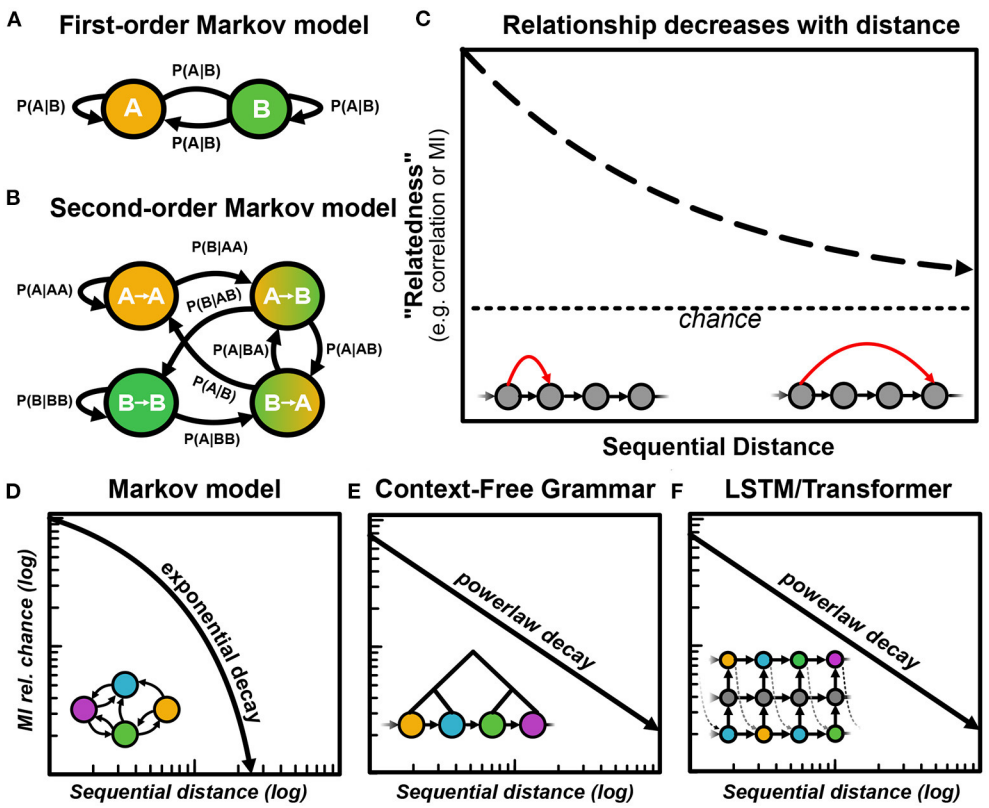


FIGURE 5 | Capturing long and short-range sequential organization with different models. **(A)** An example of a 2-state Markov model, capturing $2^2 = 4$ transitional probabilities between states. **(B)** An example second-order Markov model, capturing $2^3 = 8$ transition probabilities between states. **(C)** A visualization of the general principle that as sequential distances increase, the relatedness between elements (measured through mutual information or correlation functions) decays toward chance. **(D)** Sequences generated by Markov models decay exponentially toward chance. **(E)** Context-free grammars produce sequences that decay following a power law. **(F)** Certain neural network models such as LSTM RNNs and Transformer models produce sequences that also decay following a power law.

transition matrix representing the probability of transitions from each element to each other elements [e.g., $P(B|A)$ **Figure 5A**]. As Markov models increase in order, they become increasingly capable of capturing long-distance relationships, though high-order Markov models are rarely used in practice because of the number of parameters and amount of data needed to compute them (**Figure 5B**). Approaches such as Hidden Markov Models (Katahira et al., 2011) and Probabilistic Suffix Trees (Markowitz et al., 2013; Cohen et al., 2020b) can compute more succinct high-order Markov relationships, though the amount of data needed to capture these deeply contextualized relationships (e.g., $P(F|A, B, C, D)$) is still a limiting factor in capturing long-range organization with Markov models. Short-range relationships are also often captured graphically, treating any transition probability above zero as an edge in the graph. Graphical representations and metrics for vocal sequencing can explain general sequencing characteristics of vocalizations such as network motifs, communities, and clusters (Sasahara et al., 2012; Weiss et al., 2014; Hedley, 2016; Kershenbaum et al., 2016; Patricelli and Hebets, 2016).

6.2. Mutual Information and Long-Timescale Organization

Relationships that extend beyond adjacencies and over longer timescales are called long-range orlong-timescale relationships. For example, how related are two notes within a phrase, two phrases within a bout of song, or two bouts of song sung within a day?

Broadly, elements that are further displaced in a vocalization from one another tend to be less related. When two elements in a sequence are further apart, the relatedness between those two elements tends to be lower. For example, in birdsong, notes within a phrase are more likely to be related than notes separated by multiple phrases. The same is true of most sequential and temporal data: we can better predict what a stock price will look like tomorrow, than in 10 years. As we look further and further out into a sequence, the relatedness between elements will decrease alongside our ability to predict the future, until the relatedness approaches chance (**Figure 5C**). We can capture this relatedness over symbolic sequences using information theory. For example, given a sequence of discrete elements $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \rightarrow f$, we can estimate the mutual information between pairs of elements at e.g., a distance of 2 elements ($a - c$, $b - d$, $c - e$, and $d - f$) or 3 elements ($a - d$, $b - e$, and $c - f$). As the distance increases between pairs of elements, we expect the relatedness (mutual information) to decay toward chance as a function of sequential distance.

We can estimate the extent to which a signal exhibits long-range relationships by computing how long the mutual information between pairs of elements remains above chance. Such approaches have been used variously across animal vocalization datasets in birds and whales (Suzuki et al., 2006; Sainburg et al., 2019). Similar approaches have also been used to observe long-range structure in animal motion ethology data, such as the long-range structure in *Drosophila* (Berman et al., 2016) motility.

6.3. Inferring Structure From Sequential Relationships

The shape of the decay in relatedness as a function of sequential distance can not only tell us about the timescales that vocal sequences are operating over but can also give indications about the structure underlying sequential organization. For example, sequences generated by Markov processes, such as finite-state grammars decay exponentially (Li, 1990; Lin and Tegmark, 2017) (**Figure 5D**). Intuitively, Markov models are memoryless; each state is dictated only by the set of transition probabilities associated with the previous state. As a result, the relatedness between states decays very quickly. When there are deep latent relationships present in the structure underlying the sequence, relatedness between sequentially disparate elements decays more slowly. For example, Probabilistic Context-Free Grammars can produce power-law relationships in mutual information as a function of sequential distance (Lin and Tegmark, 2017) (**Figure 5E**).

Characterizations of statistical relationships over abstracted discrete units enables comparative analyses across species because these measures make no assumptions about units or temporal organization underlying the signal. Characterizing correlations and information decay has an especially rich history in uncovering long-range structure dating back to Claude Shannon's original work (Shannon, 1951; Li, 1990; Lin and Tegmark, 2017). Language corpora such as speech and written text decay in information following the combination of a power-law over longer distances, and exponential decay over shorter distances, attributed to the finite-state processes underlying phonological organization (Sainburg et al., 2019) and the hierarchical organization underlying language at higher levels of organization such as syntax and discourse (Alvarez-Lacalle et al., 2006; Altmann et al., 2012; Lin and Tegmark, 2017; Sainburg et al., 2019). At the same time, however, young children's speech contains the same long-range information context before complex syntax is present in speech, indicating possible extra-linguistic mechanisms at play dictating these long-range statistical relationships (Sainburg et al., 2020a). Long-range mutual information decay and correlations have also been demonstrated that in animals such as songbirds (Sainburg et al., 2019) and humpback whales (Suzuki et al., 2006), extending over minute- and hour-long timescales. In particular, birdsong exhibits similar exponential short-range and power-law long-range mutual information decay to human speech, indicating potential parallels in the mechanisms governing how patterns of vocalizations are temporally sequenced. Similar observations in non-vocal behavioral sequences (Berman et al., 2016; Sainburg et al., 2020a) also exhibit these long-range sequential organizations, suggesting similarities in latent dynamics that facilitate long-range statistical relationships.

It is tempting to suggest that these parallels suggest shared underlying structure generating mechanisms, such as universals in the hierarchical organization of behavior (e.g., Lashley, 1951; Dawkins, 1976), though we should be wary of making any extended inferences based upon the observation of long-range information decay. For example, we can infer

that power-law sequential relationships are produced by non-Markovian mechanisms because the decay is not exponential. However, the set of generative mechanisms that can produce power-law relationships in signals is not understood well enough to attribute the origins of these relationships to, for example, any specific class of formal grammar. Power-law mutual information decay in signals can also be drawn simply from coupling vocal or behavioral 1/f noise found in exogenous environmental signals.

While it is well-acknowledged that many animal vocalizations are organized hierarchically (Dawkins, 1976; Rohrmeier et al., 2015), the implications of that hierarchy in terms of underlying cognitive and physiological mechanisms are not well understood. For example, on very short timescales, birdsong motor sequencing is dictated by a hierarchical cascade of motor programs running originating in the premotor region HVC eventually ending in motor output (Doupe and Kuhl, 1999). Recent physiological evidence shows that these high-level nuclei also contain information about future states displaced from current vocalizations as well (Cohen et al., 2020b), though the mechanisms by which those relationships are learned, maintained, and ultimately dictate behavior are not yet clear.

Although we do not have access to the mechanisms underlying the observed long-distance relationships in vocal and non-vocal behavioral sequences, we do know that many vocal and behavioral sequences cannot be well-captured by Markovian models, thus alternative methods for modeling, characterizing, and forming hypotheses about the long-range organization in behavioral sequences are crucial to furthering our understanding of long-range structure in behavioral sequences. One promising approach is the use of deep neural network models such as RNNs and transformer networks (Tran et al., 2018; Morita et al., 2020). Unlike Markov models, recent neural network models like RNNs and transformer models do capture power-law relationships in sequential data (**Figure 5F**) (Lin and Tegmark, 2017; Shen, 2019). In language, transformer networks, in particular, have changed the landscape of natural language processing by capturing deeply contextual and complex implicit relationships in linguistic sequences. In birdsong, the same approaches show promise (Morita et al., 2020). For example, Morita et al. (2020) train a transformer network on Bengalese finch song and find that it captures long-range dependencies extending well beyond that of a Markov model. Like modeling language sequences, however, neural-network-based approaches suffer from the same issues of being black box and providing little explanatory power over the sequential structure they learn. In addition, the amount of data required to train a model to capture complex sequential dependencies is vast. Although the number of parameters does not increase exponentially with the amount of context the model captures (as in Markov models) state-of-the-art transformer models have billions of parameters requiring training data comprised of billions to trillions of characters (Brown et al., 2020). In language, the dataset size needed to train transformer models scales with the number of parameters in the model to prevent overfitting (Kaplan et al., 2020). When dataset sizes are smaller, LSTM RNNs perform better than more state-of-the-art Transformer language models (Ezen-Can, 2020), though transformers allow you to explicitly specify the length of temporal context allowed in the model, making them an attractive model for controlling context when generating vocal sequences (Morita et al., 2020). Although non-human animal vocalization repertoires are smaller and syntactic organization is less complex than language, birdsong analyses relying on language models will need to address the same challenges.

Neural network-based models also provide the ability to capture temporal dependencies that mutual information and correlation functions do not. Mutual-information-function-based and correlation-based analyses compute relationships between vocal elements as a function of sequential distance, ignoring any temporal relationships between disparate elements. This is both a benefit and a shortcoming of correlation methods. Ignoring intermediary temporal relationships enables the characterization of structure at temporal distances without having to additionally model higher-order combinatorial relationships [e.g., $P(F|A)$) vs. $P(F|A, B, C, D)$]. For the same reason, mutual-information-function-based and correlation-based analyses are coarse descriptions of temporal structure and miss the full temporal dynamics of the signal that neural-network-based models can capture (Morita et al., 2020).

7. SYNTHESIZING VOCALIZATIONS

Although the methods discussed in Section 5 allow us to learn representational spaces of animal vocalizations, providing new ways to infer structure in vocal repertoires, analyses on vocal signals alone lack grounding in animal behavior, perception, and physiology. In this section, we give an overview of methods for synthesizing animal vocalizations as a means to systematically control vocalization stimuli and relate vocal representations to physiology and behavior.

An ideal model for vocal synthesis exhibits several features: (1) it can model the entire vocal repertoire of a species or multiple species, (2) the parameters of the model can be related to physiological properties of the vocalizing species, and (3) the parameters of the model can be explained in terms of understandable features (i.e., it is not a black box algorithm). Throughout this section, we find that current synthesis algorithms have tradeoffs in how they balance aspects of these ideals.

One reason to systematically synthesize animal vocalizations is to probe their perceptual and physiological representations of vocal space, for example, determining how animals categorically perceive the difference between two categories of vocal units (Nelson and Marler, 1989). Traditionally, categorical perception in animals has been studied on the basis of human speech sound stimuli (Sinnott et al., 1976; Kuhl and Miller, 1978; Kuhl and Padden, 1983). Even with speech, however, the features that can be manipulated are limited. Recently methods in machine learning have furthered our ability to manipulate complex non-linear speech features substantially. These same approaches

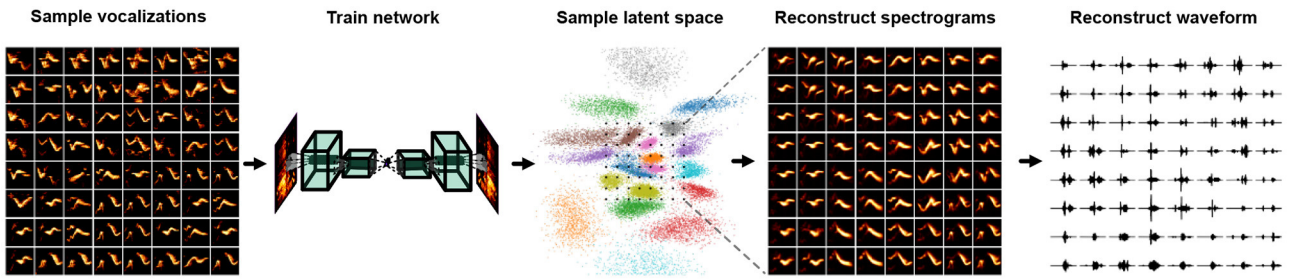


FIGURE 6 | Steps involved in synthesizing vocalization from a Variational Autoencoder (VAE) trained on spectrograms.

can often be applied to animal communication (Anikin, 2019; Sainburg et al., 2020c).

7.1. Source-Filter Models

Source-filter models have their origins in vocoding speech (Dudley, 1939), but have been used in numerous animal vocalization synthesis paradigms (DiMattina and Wang, 2006; Chakladar et al., 2008; Arneodo and Mindlin, 2009; Furuyama et al., 2017). Source-filter models decompose vocalizations into the source of the voice and filters (Kawahara, 2006). For example, the STRAIGHT algorithm (Kawahara et al., 1999; Kawahara, 2006) has been used to morph between macaque monkey vocalizations for investigations of monkey and human perception and physiology related to categorization (Chakladar et al., 2008; Furuyama et al., 2017). STRAIGHT breaks down the macaque vocalization into the fundamental frequency (the source) and its harmonics from higher-resonant or formant frequencies (the filter) (Chakladar et al., 2008). It then uses landmarks based upon these estimated parameters from the two sounds being morphed and interpolates between them to generate the morph stimuli. Furuyama et al. (2017), for example, used this method to parametrically vary generated morphs based on source and filter properties to determine the features macaques use to distinguish between conspecifics. Soundgen (Anikin, 2019) is a recent open-source GUI-based web app for R that is designed to synthesize nonverbal vocalizations using a source-filter model, including animal vocal signals such as birdsong and primate vocalizations. Related source-filter models have been developed to synthesize birdsong based upon underlying physiological mechanisms (Fee et al., 1998; Sitt et al., 2008, 2010; Arneodo and Mindlin, 2009; Arneodo et al., 2012). Recently, Arneodo et al. (2021) demonstrated that synthetic source-filter models can be coupled with neural recordings accurately reconstruct vocalizations from neural data alone. One drawback of source-filter models is the difficulty with which they can be fitted to the diversity of non-human vocalizations that exist. For example, the source-filter models of birdsong described above can well describe the dynamics of zebra finch song, but not the dual-syringeal dynamics of European starling song. Without reference to explicit hypotheses about underlying production mechanism, HMM based source-filter approaches provide one potential solution to this problem birdsong (Bonada et al., 2016).

7.2. Neural Network Models

An alternative approach to synthesizing animal vocalizations is the use of neural-network-based synthesis algorithms. These neural-network-based algorithms can be used to sample directly from the learned representational spaces described in Section 3. A simple example is autoencoder-based synthesis (**Figure 6**) (Sainburg et al., 2018a; Zuidema et al., 2020). Autoencoders can be trained on spectral representations of vocal data, and systematically sampled in the learned latent space to produce new vocalizations. Insofar as the neural network or latent projection can learn to represent the entire vocal repertoire, the entire vocal repertoire can be sampled from. In addition to sampling vocalizations from a latent distribution, vocal features can be manipulated in latent space. Well-defined latent spaces and higher-dimensional latent projections can learn to linearize complex non-linear reltionships in data. For example, in pictures of faces, the presence of a glasses, hair color, and the shape of a person's face can all be manipulated as linear features (Radford et al., 2015; Sainburg et al., 2018b, 2021). With more complex features, such as the attractiveness of a call or the age of the vocalizer, a promising avenue for future research would be to synthesize vocalizations, varying these complex non-linear features for playback studies.

Like most areas of deep learning, substantial progress has been made on the task of audio synthesis in the past few years. Basic methods comprise autoencoders (Engel et al., 2017; Kohlsdorf et al., 2020; Sainburg et al., 2020c), Generative Adversarial Networks (GANd) (Donahue et al., 2018; Engel et al., 2019; Sainburg et al., 2020c; Tjandra et al., 2020; Pagliarini et al., 2021) and autoregressive approaches (Mehri et al., 2016; Oord et al., 2016; Kalchbrenner et al., 2018; Prenger et al., 2019). One advantage of GAN-based models is that their loss is not defined directly by reconstruction loss, resulting in higher-fidelity syntheses (Larsen et al., 2016). Typically, approaches for synthesizing vocalizations based on neural networks rely on treating magnitude spectrogram like an image, training a neural network architecture in the same manner as one would an image, and finally inverting the sampled spectrogram into a waveform (Sainburg et al., 2020b; Zuidema et al., 2020; Pagliarini et al., 2021). When synthesizing vocalizations from neural networks trained on the magnitude spectrogram, the estimation of phase is necessary to invert the spectrogram into a waveform signal

for playback. The de-facto algorithm for spectral inversion has been Griffin and Lim (Griffin and Lim, 1984), though several recent approaches have been shown to improve over the Griffen-Lim algorithm recently (Prša and Rajmic, 2017; Masuyama et al., 2019). An alternative to Griffen-Lim inversion is to train neural networks to invert spectrograms either directly in the neural network architecture (Kumar et al., 2019), or perform inversion in a second network (Masuyama et al., 2019). Spectrogram-based audio synthesis can also be sidestepped entirely, training the network directly on waveform (Mehri et al., 2016; Oord et al., 2016; Engel et al., 2017).

7.3. Sound Texture Synthesis

Another approach to sound synthesis is the synthesis of sound texture (Saint-Arnaud and Popat, 1995; McDermott et al., 2009). For example, McDermott et al., (McDermott et al., 2009) propose an approach that relies on computing a set of statistics over stationary elements of sounds, and synthesizing new sounds based upon the computed statistics. By manipulating or interpolating between sound statistics, they synthesize new sound textures. One application, for example, is to manipulate components of sound textures for stimulus playback to determine what sound texture statistics listeners rely upon for recognition (McDermott and Simoncelli, 2011).

7.4. Generating Sequences

A parallel approach to synthesizing vocalizations is to generate vocal sequences from symbolically labeled vocal elements. Synthetic song sequences can be used to understand how animals process and represent temporal and sequential organization. For example, can a songbird differentiate between sounds generated using different underlying models of song syntax? Traditional approaches to song sequence generation rely upon the explicit, hand-crafted, generation of artificial grammars for playback studies. By crafting artificial grammars that differ in underlying structure, such as belonging to different classes of the Chomsky hierarchy (Gentner et al., 2006; Fitch and Friederici, 2012; Kershenbaum et al., 2014), playback studies can be used to determine whether animals can learn these grammars. A number of challenges exist with artificial grammar learning studies, however. One such challenge is the difficulty in crafting sequences that can exclusively be learned by inferring the structure that generated them, for example, making it impossible for the animal to learn by brute-force memorizing every sequence (Fitch and Friederici, 2012). When using artificial grammars, computational and modeling considerations aid in forming hypotheses about how generated grammars can be used. In the context of the neuroethology of vocal communication, these cognitive models can be related to physiological measures (Zuidema et al., 2020). An additional challenge with artificial grammar learning is constructing sequences that are structured in a similar way to natural and behaviorally relevant signals to the animal. For example, artificial grammar studies usually rely on short sequences modeled after human language syntax, rather than the animal's own communication systems. Because the task of generating vocal sequences is performed over symbolic representations of syllables, generating vocal sequences can be performed using the same methods as in text or musical note generation. These approaches can range from generating sequences using Markov models of various orders, to explicitly modeling hierarchical organization in the signal generation algorithm (Roberts et al., 2018).

8. MAPPING VOCAL COMMUNICATION TO PERCEPTION, BEHAVIOR, AND PHYSIOLOGY

The methods discussed here provide a framework to develop a set of constrained spaces from which to understand and model vocal behavior in relation to perception, production, and physiology. Perceptual or relational vocal spaces, such as UMAP projections of spectrograms, provide a low-dimensional space that can be used to infer structure in vocal repertoires. Likewise, symbolic abstractions of vocal behavior to large corpora provides a categorical representation in which vocal behavior is seen as sequential actions on those category sets. In both cases, the methods provide a constrained behavioral representation for physiological analyses.

8.1. Brain-Computer Interfacing

One of the primary challenges facing the field of brain-computer interfacing is scaling up from simple behavioral spaces like moving a cursor on a screen to complex behaviors (Gao and Ganguli, 2015). A clear advantage to the approaches discussed in Section 5 is that we can learn to bring complex vocal behavioral spaces into a compressive low-dimensional behavior spaces, even without a prior model of the structure in that space. For example, Arneodo et al. (2021) find that directly predicting the acoustic structure of zebra finch song from neural data does not perform as well as predicting the parameters of a low-dimensional biophysical model of song production. In the many species in which we do not have access to a biophysical model of vocal production, learned acoustic spaces may be a viable alternative. In contrast, the current state-of-the-art vocal prostheses for speech bypass biophysical models, directly predicting sentences (i.e., symbolic sequences) with the aid of language models (i.e., a sequence model) (Moses et al., 2021). Such methods do not capture important extra-linguistic information such as emotional tone and stress. In future work, a clear pathway forward is to develop BCI models that can both capture symbolic organization aided by sequential models, as well as within-symbol variability in the acoustic signal.

8.2. Vocal Production

Songbirds as a model for systems neuroscience are perhaps best known for the role they play in our understanding of vocal learning (Doupe and Kuhl, 1999). In addition to songbirds, rodent and non-human primate vocal behavior are becoming increasingly prominant models in the neuroscience of vocal production. In non-human primates, recent evidence has suggested some degree of constrained vocal learning in some species (Fischer and Hammerschmidt, 2020). In rodents, recent focus has been placed upon variability and structure

mouse in ultrasonic vocalizations (USVs) (Holy and Guo, 2005; Arriaga et al., 2012; Petkov and Jarvis, 2012), singing mice have emerged as a physiological model of turn-taking (Okobi et al., 2019), and the cultural transmission of vocal dialect has been observed in the naked mole rat (Barker et al., 2021). In each of these cases, quantification of how vocalizations vary as well as relationships between vocalizations (either within individual, between conspecifics, or from tutor to pupil) is integral to understanding how we learn to navigate vocal space. For example, Kollmorgen et al. (2020) use nearest-neighbor graphs and t-SNE projections to quantify and visualize the developmental trajectory of zebra finch song during vocal learning. For each syllable, they compute a nearest-neighbors graph based metric termed the "neighborhood production time," which quantifies the developmental time point at which similar (neighboring) syllables were sung. For example, a syllable song on day 45 might have 10 neighbors, sung on days between day 40 and 50, comprising its neighborhood production times. Syllable renditions that are neighbors with predominantly future syllables are deemed anticipations, while syllable repetitions that are neighbors with predominantly past syllables are deemed regressions. They observe that day-by-day, zebra finch songs gradually moves along a constant vocal learning trajectory, but anticipations and regressions differ in how they are consolidated overnight.

The number of neurons we can simultaneously record from physiologically has increased the dimensionality of neural datasets substantially over the past decade, making methods for dimensionality reduction on neural signals such as spike trains increasingly necessary for neural data analysis and opening the door to computational methods that directly link the latent representations of behavioral and neural datasets. Population modeling approaches such as LFADS (Latent Factor Analysis via Dynamical Systems; Pandarinath et al., 2018) reduce large population spiking datasets into low-dimensional trajectories, similar to the approaches discussed here with vocal signals. In the case of LFADS, these embeddings are performed over single trials using a recurrent autoencoder. One promising direction for computational neuroethology is learning the relationship between latent behavioral states and latent physiological states. By developing tools that allow us to learn the relationship between physiological and behavioral representations, we hope to untangle how, for example, movements in behavioral space reflect changes in physiology, and vice-versa. Singh Alvarado et al. (2021) developed a joint encoding model in which they used variational autoencoders to learn a joint latent representation of spectrograms of zebra finch song, and corresponding ensemble neural activity of spiny neurons in songbird basal ganglia (average calcium fluorescence of around 60 ROIs, or putative neurons, per bird). In a series of experiments leading up to this joint mapping, Singh Alvarado et al. demonstrated that Area X spiny neurons are involved in the regulation of vocal variability; exhibiting suppressed activity during female-directed song and enhanced activity during practice. Using the joint vocal-neural latent mapping, they were able to uncover the mapping between specific features of song and variants present in neural ensemble activity. In **Figure 7** we outline several similar maps between behavior, perception, and neural dynamics. Singh Alvarado et al.'s work exhibit that one such latent map, a vocal-motor mapping between motor physiology and vocal behavior, can uncover complex and detailed relationships that traditional methodology cannot. Similar mappings between the physiology, perception, and behavior of sender-receiver dynamics (e.g., **Figure 7**) are also well poised to benefit from emerging latent approaches.

The physiology of vocal syntax is another area poised to benefit from computational ethology. One example is the role of

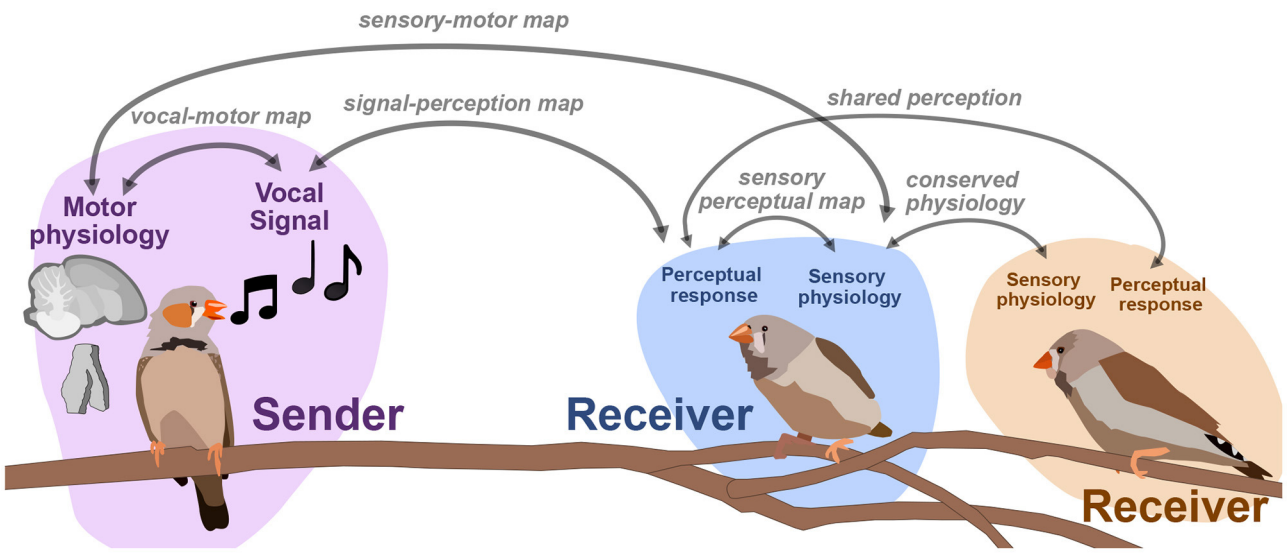


FIGURE 7 | An outline of mappings between perceptual, acoustic, and physiological signals. One focus in sensory and motor neuroscience is to learn the relationships between signals, perception, and physiology.

the songbird premotor nucleus, HVC, in encoding song syntax. Birdsong has a long history of being described sequentially in terms of low-order Markovian transitions between song elements. HVC's role in song syntax, until recently has been described exclusively in terms of these low-order transition statistics (Fujimoto et al., 2011). In a recent example, however, (Cohen et al., 2020b) made use of an automated birdsong labeling paradigm and high-order sequence model to observe 'hidden neural states' encoding sequentially displaced (i.e., high-order) transitions in the premotor nucleus HVC of canaries. To identify non-adjacent dependencies in the song, they used a Prediction Suffix Tree (Markowitz et al., 2013), which can capture high-order Markovian relationships in the song syntax. Prediction Suffix Trees have previously been used to observe long-range dependencies up to the 7th order in canaries (Markowitz et al., 2013). While birds were singing, Cohen et al., used a miniature microscope to image neurons from HVC, a region involved in the songbird vocal motor circuit. They observed that HVC ROIs were locked to individual song-phrases and transitions, and that this phrase locking is modified by non-adjacent context, displaced by several phrases and seconds. As more recent approaches give access to larger datasets enabling the identification of longer-range dependencies in birdsong, it is currently not clear whether we have yet found an upper bound on the sequential displacement of long-range representations of vocal syntax in physiology.

Outside of songbirds and mammals, male *Drosophila* song, although not strictly vocal, is temporally patterned and driven by both environment and internal states (Coen et al., 2014; Calhoun et al., 2019). Calhoun et al. (2019) jointly model and uncover relationships between temporal song structure and interactions with a potential mate. Using a sequential model (an HMM-GLM hybrid) they demonstrate that song patterning is underlined by three hidden sensorimotor states, under which male's song productions differ in their relationship to female behavior. Using optogenetic activation, they were then able to identify neurons involved in switching between these sensorimotor states.

8.3. Vocal Perception

Similar to vocal production, latent and sequential models are promising avenues for better understanding cognitive and physiological underpinnings of vocal perception. In songbirds, primates, and rodents, many foundational studies of auditory categorical perception, perceptual decision making, and their underlying physiology rely upon either relatively simple stimuli such as tones or complex stimuli like human speech (Kuhl and Miller, 1978; Kuhl and Padden, 1983; Russ et al., 2007; ten Cate, 2014; Xin et al., 2019). Categorization in these stimulus spaces are attractive because they are well-characterized and understood. Across species, however, neural responses are often tied to complex and more behaviorally-relevant acoustic phenomena such as recognizing and discriminating between conspecific vocalizations (Bailey et al., 2002; Liu et al., 2019). When the acoustic features underlying vocal repertoires are simple and known, categorical stimuli can be selected directly based upon those features. For example, Lachlan and Nowicki (2015) manipulate a single dimension, the duration of swamp sparrow notes, to determine how notes are categorically perceived in different sequential contexts. In speech, voice onset time (VOT) is a similar single-dimension commonly used for categorical perception paradigms (Liberman et al., 1957). However, it is rarely the case that categorical perception is driven by a single dimension. Thus, building stimuli in more complex feature spaces will be necessary to untangle the relationship between vocal features, perception, and physiology. When biophysical models of vocal structure exist, species relevant stimuli can be generated using biophysical parameters (Arneodo and Mindlin, 2009). When the underlying acoustical structure of a vocal repertoire is more complex and biophysical models of vocalizations have not been defined, neural-network synthesized vocalizations are an attractive alternative. For example, as discussed above, birdsong can be synthesized with neural networks for physiological and perceptual playback studies to determine perceptual similarity between syllables or learn categorical boundaries between song-morphs (Sainburg et al., 2018a; Thielk et al., 2018; Zuidema et al., 2020). By systematically controlling the signal space of a vocal repertoire, we can systematically explore how changes in that space relate to changes in physiology.

Algorithmic approaches are similarly well poised to aid in our understanding of how vocal sequences are maintained and represented. Sequence learning research in human and non-human primates is largely dominated by artificial grammar learning (AGL) research, an umbrella category that comprises several different forms of sequence learning ranging from hierarchically nested tree structures to transitional probabilities (Dehaene et al., 2015). Artificial grammar learning studies aim to determine what structures animals (and humans) are capable of learning, what cognitive mechanisms underlie grammar induction, and what physiological systems underlie those cognitive mechanisms. In the domain of primate sequence learning, neural pathways are generally conserved between humans and non-human primates and involve the ventral regions of cortex (Wilson et al., 2017). Determining an appropriate stimuli set is requisite for developing an AGL paradigm. Latent representations of vocalizations can aid in choosing stimuli from a well-defined stimulus space. For example, when choosing a stimulus set for an A^nB^n grammar, it is desirable depending on the goal of the task to ensure that the constituent vocalizations comprising A and B belong to equidistant or separate clusters in acoustic or perceptual spaces (Zuidema et al., 2020).

While artificial grammar learning has also played a prominent role in birdsong sequence learning (ten Cate and Okanoya, 2012), the structure underlying an animal's own vocal syntax provides an opportunity to study the neural and cognitive underpinnings of a more ethologically-relevant complex sequential structure. Despite the important role vocal syntax production has played in establishing birdsong as a model in systems neuroscience, a surprising gap exists in our knowledge of the physiological circuits underlying how syntactic information is recognized and sequentially integrated when listening to song. Songbird vocal communication contains often very complex syntax that can be structured over long timescales comprising often tens to

hundreds of unique, stereotyped vocal units (Cody et al., 2016). Conspecifics pay attention to the structure of that song. Abe and Watanabe (2011) developed a habituation/dishabituation paradigm with Bengalese finches alongside immediate early gene (IEG) expression and lesioning experiments to explore the role of song nuclei on the recognition of grammatical sequences. They found that IEG expression increased when presented with non-conforming/nonpredictive strings in the nuclei LMAN, a basal ganglia output nuclei characterized by recurrent loops that is also involved in vocal learning (Bottjer and Altenau, 2010). Abe and Wantenable then lesioned LMAN and measured song discrimination with their habituation paradigm. They found that discrimination was disturbed in birds where LMAN was lesioned, implicating LMAN in the ability to discriminate syntactic song. How syntactic information is learned, integrated, and maintained in LMAN and associated striatal regions of songbird brain are still open questions.

In contrast to the auditory domain where little is known about syntactic integration, a neural correlate for complex and abstract information integration, NCL, has been well established and characterized in songbird vision with pigeons and corvids (Kröner and Güntürkün, 1999; Güntürkün, 2005). Strong parallels exist between NCL and the primate prefrontal cortex, which is involved in sequence learning. NCL has variously been associated with rule learning (Veit and Nieder, 2013), numerosity (Ditz and Nieder, 2015; Wagener et al., 2018), directed forgetting (Rose and Colombo, 2005; Milmine et al., 2008; Helduser et al., 2013), choice behavior (Kalenscher et al., 2003), working memory (Diekamp et al., 2002; Rinnert et al., 2019), sequence learning (Helduser and Güntürkün, 2012), and reward learning. Anatomically and neurochemically NCL also exhibits strong parallels to the primate prefrontal cortex. NCL is characterized by similar circuitry from auditory and dopaminergic afferents, as well as multi-sensory projections (Kröner and Güntürkün, 1999; von Eugen et al., 2020). Surprisingly, however, an auditory equivalent to the visual working-memory region in NCL has not been found, though they have been observed in the multi-sensory audio-visual integration and association (Moll and Nieder, 2015, 2017). Birdsong is well poised as a signal to be a model of vocal syntax perception, To establish this model, however, it will be imperative to uncover the systems in songbird brain related to working memory and temporal context integration in song. NCL appears to be a likely candidate for processing syntactic vocal signals, though, as yet, this has not been found to be the case.

Although mouse USVs do not appear to contain temporal structure to the same extent as songbirds, mouse USVs are temporally organized (Castellucci et al., 2018) and female mice also show preference for more complex syllables and sequences (Holy and Guo, 2005), making mouse USVs another potential target for the study of syntactic and sequential integration in vocal perception.

9. DISCUSSION

This review covers emerging approaches in the computational neuroethology of vocal communication enabling researchers to engage with large and diverse datasets of vocal signals and to represent them in computationally tractable frameworks.

We started by discussing techniques to process and represent acoustic signals. We then discussed how to parse complex vocal datasets into species, individuals, and discrete vocal elements. Next, we discussed how relational structure can be extracted from vocal signals, how these signals can be clustered in learned latent spaces, and how these latent spaces capture different aspects of the information contained within the underlying signals. We then discussed how temporal structure can be inferred from vocal units, including emerging work on the non-Markovian dynamics underlying vocal behavior. In the next section, we discussed how vocalizations can be synthesized for use in playback experiments that allow an unprecedented degree of control over non-linear and complex vocal feature spaces. Finally, we discussed how these approaches are being applied to the field of neuroethology and emerging frameworks for understanding vocal signals and their underlying physiology.

The methods discussed here provide a promising avenue for a broader, more diverse, and larger-scale neuroethology of vocal communication, than the research practices that have dominated the past several decades, and hold the promise of expanding both the breadth and depth of our understanding. Instead of focusing on a small number of model species, new computational techniques provide a framework for studying vocal behavior across a wide range of animals. While much of vocal neuroethology has recently focused on songbirds and mammals the techniques discussed here are equally applicable to the abundance of other species studied in bioacoustics and behavioral ecology including fish, amphibians, and insects. Even within songbirds, research on vocal learning in songbirds has ignored the majority of species, female birdsong, and most call types (Loo and Cain, 2021). Likewise, because these new computational methodologies can often deal with unstructured data, they enable us to expand beyond simplified, isolated behaviors in controlled environments to more natural or naturalistic behavioral contexts where dynamics involving multi-modal integration and multi-animal social interactions arise. As we capture increasing levels of detail in behavior, our understanding of its sophistication naturally follows. Already, these new computational framework have revealed deep structure in the sequential organization of communication, where large-scale datasets of both symbolic sequences, and latent projections that capture rendition-to-rendition variability, have enabled quantitative analyses of rare (but perhaps meaningful) events, such as long-range syntactic organization. Together, all of these approaches point toward a new framework, in which complex and non-linear behavioral and physiological signals can be represented in compressive and tractable spaces that can capture the complex dynamics and relationships in the increasingly rich datasets available to researchers.

As with any powerful tool, these techniques require careful consideration when put into practice. Broadly, automation and machine learning in data analysis can be fraught with unexpected complications and confounds that may be hard to spot. For example, automating the labeling of large datasets of birdsong syllables can speed up the task of labeling by days, weeks, or

months, but can also leave the experimenter with less intuitive knowledge of the animal's vocal repertoire, resulting in a loss of domain knowledge. As we have noted elsewhere (Sainburg et al., 2020b), when domain knowledge is available it should be integrated with one computational approach. Another potential pitfall (and a source of much needed research effort) is in understanding the structure of the latent manifolds that are yielded in many of the described methods. In particular, non-linear latent modeling techniques like UMAP or neural networks can capture complex relationships in vocal data, but interpreting these projections requires an understanding of how data are represented within the geometry of the latent space. For example, UMAP captures primarily local structure in datasets that are present in nearest neighbor graphs, meaning that the relative distances of vocal elements have no explicit relation to the data, as is the case in PCA for example.

Attending to the cautions of computational abstraction, the approaches discussed in this manuscript provide a framework from which to quantify vocal signals that promises to yield important new insights into vocal behavior and neurobiology. These approaches enable neuroethologists to project vocalizations onto low dimensionaland latent manifolds, visualize and quantify the transitional structure and information decay of vocal syntax, and map vocal and neural repertoires into shared neural spaces for functional representation action. As the richness of datasets grow to capture more of the complexities of behavior and physiology, methods and frameworks for modeling and inferring structure in ethological data are increasingly necessary for hypothesis formulation and testing. The methods and frameworks discussed in this review parallel and supplement those in the broader field of computational neuroethology.

AUTHOR CONTRIBUTIONS

This manuscript was written by TS and TG. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by a CARTA Fellowship to TS, NIH 5T32MH020002-20 to TS, and 5R01DC018055-02 to TG.

REFERENCES

Abe, K., and Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat. Neurosci.* 14, 1067–1074. doi: 10.1038/nn.2869

Adi, K., Johnson, M. T., and Osiejuk, T. S. (2010). Acoustic censusing using automatic vocalization classification and identity recognition. *J. Acoust. Soc. Am.* 127, 874–883. doi: 10.1121/1.3273887

Alonso, J. B., Cabrera, J., Shyamnani, R., Travieso, C. M., Bolaños, F., García, A., et al. (2017). Automatic anuran identification using noise removal and audio activity detection. *Expert. Syst. Appl.* 72, 83–92. doi: 10.1016/j.eswa.2016.12.019

Altmann, E. G., Cristadoro, G., and Degli Esposti, M. (2012). On the origin of long-range correlations in texts. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11582–11587. doi: 10.1073/pnas.1117723109

Alvarez-Lacalle, E., Dorow, B., Eckmann, J.-P., and Moses, E. (2006). Hierarchical structures induce long-range dynamical correlations in written texts. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7956–7961. doi: 10.1073/pnas.0510673103

Anderson, D. J., and Perona, P. (2014). Toward a science of computational ethology. *Neuron* 84, 18–31. doi: 10.1016/j.neuron.2014.09.005

Anderson, S. E., Dave, A. S., and Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *J. Acoust. Soc. Am.* 100, 1209–1219. doi: 10.1121/1.415968

Anikin, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behav. Res. Methods* 51, 778–792. doi: 10.3758/s13428-018-1095-7

Arneodo, E. M., Chen, S., Brown, I. I., D. E., Gilja, V., and Gentner, T. Q. (2021). Neurally driven synthesis of learned, complex vocalizations. *Curr. Biol.* 31, 3419.e5–3425.e5. doi: 10.1016/j.cub.2021.05.035

Arneodo, E. M., and Mindlin, G. B. (2009). Source-tract coupling in birdsong production. *Phys. Rev. E* 79, 061921. doi: 10.1103/PhysRevE.79.061921

Arneodo, E. M., Perl, Y. S., Goller, F., and Mindlin, G. B. (2012). Prosthetic avian vocal organ controlled by a freely behaving bird based on a low dimensional model of the biomechanical periphery. *PLoS Comput. Biol.* 8:e1002546. doi: 10.1371/journal.pcbi.1002546

Arriaga, G., Zhou, E. P., and Jarvis, E. D. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE* 7:e46610. doi: 10.1371/journal.pone.0046610

Baevski, A., Zhou, H., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: a framework for self-supervised learning of speech representations. *arXiv preprint* arXiv:2006.11477.

Bailey, D. J., Rosebush, J. C., and Wade, J. (2002). The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch. *J. Neurobiol.* 52, 43–51. doi: 10.1002/neu.10070

Barker, A. J., Veviurko, G., Bennett, N. C., Hart, D. W., Mograby, L., and Lewin, G. R. (2021). Cultural transmission of vocal dialect in the naked mole-rat. *Science* 371, 503–507. doi: 10.1126/science.abc6588

Berman, G. J. (2018). Measuring behavior across scales. *BMC Biol.* 16:23. doi: 10.1186/s12915-018-0494-7

Berman, G. J., Bialek, W., and Shaevitz, J. W. (2016). Predictability and hierarchy in drosophila behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11943–11948. doi: 10.1073/pnas.1607601113

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., and Raffel, C. (2019). Mixmatch: a holistic approach to semi-supervised learning. *arXiv preprint* arXiv:1905.02249.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust.* 27, 113–120. doi: 10.1109/TASSP.1979.1163209

Bonada, J., Lachlan, R., and Blaauw, M. (2016). "Bird song synthesis based on hidden markov models," in *Interspeech 2016; 2016 Sep 08-12* (San Francisco, CA: [Baixas]: ISCA), 2582–2586.

Bottjer, S. W., and Altenau, B. (2010). Parallel pathways for vocal learning in basal ganglia of songbirds. *Nat. Neurosci.* 13, 153–155. doi: 10.1038/nn.2472

Bregman, M. R., Patel, A. D., and Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1666–1671. doi: 10.1073/pnas.1515380113

Brown, A., Garg, S., and Montgomery, J. (2017). Automatic and efficient denoising of bioacoustics recordings using mmse stsa. *IEEE Access* 6, 5010–5022. doi: 10.1109/ACCESS.2017.2782778

Brown, A. E., and De Bivort, B. (2018). Ethology as a physical science. *Nat. Phys.* 14, 653–657. doi: 10.1038/s41567-018-0093-0

Brown, G. J., and Cooke, M. (1994). Computational auditory scene analysis. *Comput. Speech Lang.* 8, 297–336. doi: 10.1006/csla.1994.1016

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *arXiv preprint* arXiv:2005.14165.

Bruno, J. H., and Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behav. Proc.* 163, 53–59. doi: 10.1016/j.beproc.2017.11.001

Calhoun, A. J., Pillow, J. W., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nat. Neurosci.* 22, 2040–2049. doi: 10.1038/s41593-019-0533-x

Castellucci, G. A., Calbick, D., and McCormick, D. (2018). The temporal organization of mouse ultrasonic vocalizations. *PLoS ONE* 13:e0199929. doi: 10.1371/journal.pone.0199929

Chakladar, S., Logothetis, N. K., and Petkov, C. I. (2008). Morphing rhesus monkey vocalizations. *J. Neurosci. Methods* 170, 45–55. doi: 10.1016/j.jneumeth.2007.12.023

Christin, S., Hervet, É., and Lecomte, N. (2019). Applications for deep learning in ecology. *Methods Ecol. Evolut.* 10, 1632–1644. doi: 10.1111/2041-210X.13256

Clark, C. W., Marler, P., and Beeman, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology* 76, 101–115. doi: 10.1111/j.1439-0310.1987.tb00676.x

Cody, M. L., Stabler, E., Sanchez Castellanos, H. M., and Taylor, C. E. (2016). Structure, syntax and "small-world" organization in the complex songs of california thrashers (toxostoma redivivum). *Bioacoustics* 25, 41–54. doi: 10.1080/09524622.2015.1089418

Coen, P., Clemens, J., Weinstein, A. J., Pacheco, D. A., Deng, Y., and Murthy, M. (2014). Dynamic sensory cues shape song structure in drosophila. *Nature* 507, 233–237. doi: 10.1038/nature13131

Cohen, Y., Nicholson, D. A., Sanchioni, A., Mallaber, E. K., Skidanova, V., and Gardner, T. J. (2020a). Tweetynet: a neural network that enables high-throughput, automated annotation of birdsong. *bioRxiv*. doi: 10.1101/2020.08.28.272088

Cohen, Y., Shen, J., Semu, D., Leman, D. P., Liberti, W. A., Perkins, L. N., et al. (2020b). Hidden neural states underlie canary song syntax. *Nature* 582, 539–544. doi: 10.1038/s41586-020-2397-3

Datta, S. R., Anderson, D. J., Branson, K., Perona, P., and Leifer, A. (2019). Computational neuroethology: a call to action. *Neuron* 104, 11–24. doi: 10.1016/j.neuron.2019.09.038

Dawkins, R. (1976). Hierarchical organisation: a candidate principle for ethology. *Growing Points Ethol.* 7:54.

De Silva, V., and Tenenbaum, J. B. (2002). "Global versus local methods in nonlinear dimensionality reduction," in *NIPS'02: Proceedings of the 15th International Conference on Neural Information Processing Systems Vol. 15*. (Vancouver, BC: Association for Computing Machinery), 705–712.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88, 2–19. doi: 10.1016/j.neuron.2015.09.019

Diekamp, B., Kalt, T., and Güntürkün, O. (2002). Working memory neurons in pigeons. *J. Neurosci.* 22, RC210–RC210. doi: 10.1523/JNEUROSCI.22-04-j0002.2002

DiMattina, C., and Wang, X. (2006). Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. *J. Neurophysiol.* 95, 1244–1262. doi: 10.1152/jn.00818.2005

Ditz, H. M., and Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7827–7832. doi: 10.1073/pnas.1504245112

Donahue, C., McAuley, J., and Puckette, M. (2018). Adversarial audio synthesis. *arXiv preprint* arXiv:1802.04208.

Dooling, R. J., Park, T. J., Brown, S. D., Okanoya, K., and Soli, S. D. (1987). Perceptual organization of acoustic stimuli by budgerigars (melopsittacus undulatus): Ii. vocal signals. *J. Compar. Psychol.* 101, 367. doi: 10.1037/0735-7036.101.4.367

Dooling, R. J., and Prior, N. H. (2017). Do we hear what birds hear in birdsong? *Anim. Behav.* 124, 283–289. doi: 10.1016/j.anbehav.2016.10.012

Doupe, A. J., and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631. doi: 10.1146/annurev.neuro.22.1.567

Dudley, H. (1939). Remaking speech. *J. Acoust. Soc. Am.* 11, 169–177. doi: 10.1121/1.1916020

Dunn, T. W., Marshall, J. D., Severson, K. S., Aldarondo, D. E., Hildebrand, D. G., Chettih, S. N., et al. (2021). Geometric deep learning enables 3d kinematic profiling across species and environments. *Nat. Methods* 18, 564–573. doi: 10.1038/s41592-021-01106-6

Eldredge, D. H., Miller, J. D., and Bohne, B. A. (1981). A frequency-position map for the chinchilla cochlea. *J. Acoust. Soc. Am.* 69, 1091–1095. doi: 10.1121/1.385688

Elie, J. E., and Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* 19, 285–315. doi: 10.1007/s10071-015-0933-6

Elie, J. E., and Theunissen, F. E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nat. Commun.* 9, 1–11. doi: 10.1038/s41467-018-06394-9

Engel, J., Agrawal, K. K., Chen, S., Gulrajani, I., Donahue, C., and Roberts, A. (2019). Gansynth: adversarial neural audio synthesis. *arXiv preprint* arXiv:1902.08710.

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., et al. (2017). "Neural audio synthesis of musical notes with wavenet autoencoders," in *International Conference on Machine Learning* (Sydney, NSW: PMLR), 1068–1077.

Ephraim, Y., and Malah, D. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans. Acoust.* 32, 1109–1121. doi: 10.1109/TASSP.1984.1164453

Ezen-Can, A. (2020). A comparison of lstm and bert for small corpus. *arXiv preprint* arXiv:2009.05451.

Feather, J., Durango, A., Gonzalez, R., and McDermott, J. (2019). "Metamers of neural networks reveal divergence from human perceptual systems," in *Advances in Neural Information Processing Systems*, (Vancouver, BC: Association for Computing Machinery), 10078–10089.

Fee, M. S., Shraiman, B., Pesaran, B., and Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* 395, 67–71. doi: 10.1038/25725

Fischer, J., and Hammerschmidt, K. (2020). Towards a new taxonomy of primate vocal production learning. *Philos. Trans. R. Soc. B* 375, 20190045. doi: 10.1098/rstb.2019.0045

Fitch, W. T., and Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: an overview. *Philos. Trans. R. Soc. B* 367, 1933–1955. doi: 10.1098/rstb.2012.0103

Frasier, K. E., Roch, M. A., Soldevilla, M. S., Wiggins, S. M., Garrison, L. P., and Hildebrand, J. A. (2017). Automated classification of dolphin echolocation click types from the gulf of mexico. *PLoS Comput. Biol.* 13:e1005823. doi: 10.1371/journal.pcbi.1005823

Fujimoto, H., Hasegawa, T., and Watanabe, D. (2011). Neural coding of syntactic structure in learned vocalizations in the songbird. *J. Neurosci.* 31, 10023–10033. doi: 10.1523/JNEUROSCI.1606-11.2011

Fukushima, M., Doyle, A. M., Mullarkey, M. P., Mishkin, M., and Averbeck, B. B. (2015). Distributed acoustic cues for caller identity in macaque vocalization. *R. Soc. Open Sci.* 2, 150432. doi: 10.1098/rsos.150432

Fulop, S. A., and Fitz, K. (2006). Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. *J. Acoust. Soc. Am.* 119, 360–371. doi: 10.1121/1.2133000

Furuyama, T., Kobayasi, K. I., and Riquimaroux, H. (2017). Acoustic characteristics used by japanese macaques for individual discrimination. *J. Exp. Biol.* 220, 3571–3578. doi: 10.1242/jeb.154765

Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* 32, 148–155. doi: 10.1016/j.conb.2015.04.003

Gardner, T. J., and Magnasco, M. O. (2006). Sparse time-frequency representations. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6094–6099. doi: 10.1073/pnas.0601707103

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207. doi: 10.1038/nature04675

Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Rauber, A., and Joly, A. (2014). "Lifeclef bird identification task 2014," in *CLEF: Conference and Labs of the Evaluation Forum, 1180.* (Sheffield: Association for Computing Machinery), 585-597.

Goffinet, J., Brudner, S., Mooney, R., and Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *Elife* 10:e67855. doi: 10.7554/eLife.67855

Greenwood, D. D. (1997). The mel scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios. *Hear. Res*. 103, 199–224. doi: 10.1016/S0378-5955(96)00175-X

Griffin, D., and Lim, J. (1984). Signal estimation from modified short-time fourier transform. *IEEE Trans. Acoust*. 32, 236–243. doi: 10.1109/TASSP.1984.1164317

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., et al. (2020). Conformer: convolution-augmented transformer for speech recognition. *arXiv preprint* arXiv:2005.08100. doi: 10.21437/Interspeech.2020-3015

Güntürkün, O. (2005). The avian 'prefrontal cortex' and cognition. *Curr. Opin. Neurobiol*. 15, 686–693. doi: 10.1016/j.conb.2005.10.003

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). 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

Hedley, R. W. (2016). Composition and sequential organization of song repertoires in cassin's vireo (vireo cassinii). *J. Ornithol*. 157, 13–22. doi: 10.1007/s10336-015-1238-x

Helduser, S., Cheng, S., and Güntürkün, O. (2013). Identification of two forebrain structures that mediate execution of memorized sequences in the pigeon. *J. Neurophysiol*. 109, 958–968. doi: 10.1152/jn.00763.2012

Helduser, S., and Güntürkün, O. (2012). Neural substrates for serial reaction time tasks in pigeons. *Behav. Brain Res*. 230, 132–143. doi: 10.1016/j.bbr.2012.02.013

Holveck, M.-J., Vieira de Castro, A. C., Lachlan, R. F., ten Cate, C., and Riebel, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behav. Ecol*. 19, 1267–1281. doi: 10.1093/beheco/arn078

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

Hsu, S.-B., Lee, C.-H., Chang, P.-C., Han, C.-C., and Fan, K.-C. (2018). Local wavelet acoustic pattern: a novel time-frequency descriptor for birdsong recognition. *IEEE Trans. Multimedia* 20, 3187–3199. doi: 10.1109/TMM.2018.2834866

Janney, E., Taylor, H., Scharff, C., Rothenberg, D., Parra, L. C., and Tchernichovski, O. (2016). Temporal regularity increases with repertoire complexity in the australian pied butcherbird's song. *R. Soc. Open Sci*. 3, 160357. doi: 10.1098/rsos.160357

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., and Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. *Curr. Biol*. 28, 1851–1859. doi: 10.1016/j.cub.2018.04.047

Jin, D. Z., and Kozhevnikov, A. A. (2011). A compact statistical model of the song syntax in bengalese finch. *PLoS Comput. Biol*. 7:e1001108. doi: 10.1371/journal.pcbi.1001108

Kalchbrenner, N., Elsen, E., Simonyan, K., Noury, S., Casagrande, N., Lockhart, E., et al. (2018). "Efficient neural audio synthesis," in *International Conference on Machine Learning* (Stockholm: PMLR), 2410–2419.

Kalenscher, T., Diekamp, B., and Güntürkün, O. (2003). Neural architecture of choice behaviour in a concurrent interval schedule. *Eur. J. Neurosci*. 18, 2627–2637. doi: 10.1046/j.1460-9568.2003.03006.x

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., et al. (2020). Scaling laws for neural language models. *arXiv preprint* arXiv:2001.08361.

Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., et al. (2019). "A comparative study on transformer vs rnn in speech applications," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (Singapore: IEEE), 449–456.

Katahira, K., Suzuki, K., Okanoya, K., and Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden markov processes. *PLoS ONE* 6:e24516. doi: 10.1371/journal.pone.0024516

Kawahara, H. (2006). Straight, exploitation of the other aspect of vocoder: perceptually isomorphic decomposition of speech sounds. *Acoust. Sci. Technol*. 27, 349–353. doi: 10.1250/ast.27.349

Kawahara, H., Masuda-Katsuse, I., and De Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: possible role of a repetitive structure in sounds. *Speech Commun*. 27, 187–207. doi: 10.1016/S0167-6393(98)00085-5

Keen, S. C., Odom, K. J., Webster, M. S., Kohn, G. M., Wright, T. F., and Araya-Salas, M. (2021). A machine learning approach for classifying and quantifying acoustic diversity. *Methods Ecol. Evolut*. 12, 1213–1225. doi: 10.1111/2041-210X.13599

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., et al. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol. Rev*. 91, 13–52. doi: 10.1111/brv.12160

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., and Bohn, K. (2014). Animal vocal sequences: not the markov chains we thought they were. *Proc. R. Soc. B Biol. Sci*. 281, 20141370. doi: 10.1098/rspb.2014.1370

Kiapuchinski, D. M., Lima, C. R. E., and Kaestner, C. A. A. (2012). "Spectral noise gate technique applied to birdsong preprocessing on embedded unit," in *2012 IEEE International Symposium on Multimedia* (Irvine, CA: IEEE), 24–27.

Kim, B., and Pardo, B. (2018). A human-in-the-loop system for sound event detection and annotation. *ACM Trans. Interact. Intell. Syst*. 8, 1–23. doi: 10.1145/3214366

Kim, H.-G., Obermayer, K., Bode, M., and Ruwisch, D. (2000). Real-time noise canceling based on spectral minimum detection and diffusive gain factors. *J. Acoust. Soc. Am*. 108, 2484–2484. doi: 10.1121/1.4743167

Knight, E. C., Poo Hernandez, S., Bayne, E. M., Bulitko, V., and Tucker, B. V. (2020). Pre-processing spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. *Bioacoustics* 29, 337–355. doi: 10.1080/09524622.2019.1606734

Kogan, J. A., and Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: a comparative study. *J. Acoust. Soc. Am*. 103, 2185–2196. doi: 10.1121/1.421364

Kohlsdorf, D., Herzing, D., and Starner, T. (2020). "An auto encoder for audio dolphin communication," in *2020 International Joint Conference on Neural Networks (IJCNN)*, (Glasgow: IEEE), 1–7.

Kollmorgen, S., Hahnloser, R. H., and Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature* 577, 526–530. doi: 10.1038/s41586-019-1892-x

Kong, Q., Xu, Y., and Plumbley, M. D. (2017). "Joint detection and classification convolutional neural network on weakly labelled bird audio detection," in *2017 25th European Signal Processing Conference (EUSIPCO)* (Kos: IEEE), 1749–1753.

Kröner, S., and Güntürkün, O. (1999). Afferent and efferent connections of the caudolateral neostriatum in the pigeon (columba livia): a retro- and anterograde pathway tracing study. *J. Compar. Neurol*. 407, 228–260. doi: 10.1002/(SICI)1096-9861(19990503)407:2<228::AID-CNE6>3.0.CO;2-2

Kuhl, P. K., and Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic vot stimuli. *J. Acoust. Soc. Am*. 63, 905–917. doi: 10.1121/1.381770

Kuhl, P. K., and Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J. Acoust. Soc. Am*. 73, 1003–1010. doi: 10.1121/1.389148

Kumar, K., Kumar, R., de Boissiere, T., Gestin, L., Teoh, W. Z., Sotelo, J., et al. (2019). Melgan: Generative adversarial networks for conditional waveform synthesis. *arXiv preprint* arXiv:1910.06711.

Lachlan, R. (2007). *Luscinia: A Bioacoustics Analysis Computer Program*. Available online at: http://luscinia.sourceforge.net/page3/page21/page21.html

Lachlan, R., Verhagen, L., Peters, S., and Cate, C., t. (2010). Are there species-universal categories in bird song phonology and syntax? a comparative study of chaffinches (fringilla coelebs), zebra finches (taenopygia guttata), and swamp sparrows (melospiza georgiana). *J. Compar. Psychol*. 124, 92. doi: 10.1037/a0016996

Lachlan, R. F., and Nowicki, S. (2015). Context-dependent categorical perception in a songbird. *Proc. Natl. Acad. Sci. U.S.A*. 112, 1892–1897. doi: 10.1073/pnas.1410844112

Larsen, A. B. L., Sønderby, S. K., Larochelle, H., and Winther, O. (2016). "Autoencoding beyond pixels using a learned similarity metric," in *International Conference on Machine Learning* (PMLR), 1558–1566.

Lashley, K. S. (1951). *The Problem of Serial Order in Behavior, Vol. 21*. Oxford: Bobbs-Merrill.

Lasseck, M. (2013). "Bird song classification in field recordings: winning solution for nips4b 2013 competition," in *Proceeding of International Symposium Neural Information Scaled for Bioacoustics* (Lake Tahoe, NV: sabiod. org/nips4b, joint to NIPS), 176–181.

Lee, G. W., and Kim, H. K. (2020). Multi-task learning u-net for single-channel speech enhancement and mask-based voice activity detection. *Appl. Sci.* 10, 3230. doi: 10.3390/app10093230

Li, W. (1990). Mutual information functions versus correlation functions. *J. Stat. Phys.* 60, 823–837. doi: 10.1007/BF01025996

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358. doi: 10.1037/h0044417

Lin, H. W., and Tegmark, M. (2017). Critical behavior in physics and probabilistic formal languages. *Entropy* 19, 299. doi: 10.3390/e19070299

Liu, S. T., Montes-Lourido, P., Wang, X., and Sadagopan, S. (2019). Optimal features for auditory categorization. *Nat. Commun.* 10, 1–14. doi: 10.1038/s41467-019-09115-y

Loo, Y. Y., and Cain, K. E. (2021). A call to expand avian vocal development research. *Front. Ecol. Evolut.* 9:757972. doi: 10.3389/fevo.2021.757972

Lostanlen, V., Palmer, K., Knight, E., Clark, C., Klinck, H., Farnsworth, A., et al. (2019a). Long-distance detection of bioacoustic events with per-channel energy normalization. *arXiv preprint* arXiv:1911.00417. doi: 10.33682/ts6e-sn53

Lostanlen, V., Salamon, J., Cartwright, M., McFee, B., Farnsworth, A., Kelling, S., et al. (2018). Per-channel energy normalization: why and how. *IEEE Signal Process Lett.* 26, 39–43. doi: 10.1109/LSP.2018.2878620

Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., and Bello, J. P. (2019b). Robust sound event detection in bioacoustic sensor networks. *PLoS ONE* 14:e0214168. doi: 10.1371/journal.pone.0214168

Main, L., and Thornton, J. (2015). "A cortically-inspired model for bioacoustics recognition," in *International Conference on Neural Information Processing* (Istanbul: Springer), 348–355.

Markowitz, J. E., Ivie, E., Kligler, L., and Gardner, T. J. (2013). Long-range order in canary song. *PLoS Comput. Biol.* 9:e1003052. doi: 10.1371/journal.pcbi.1003052

Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., and Ölveczky, B. P. (2021). Continuous whole-body 3d kinematic recordings across the rodent behavioral repertoire. *Neuron* 109, 420–437. doi: 10.1016/j.neuron.2020.11.016

Masuyama, Y., Yatabe, K., Koizumi, Y., Oikawa, Y., and Harada, N. (2019). "Deep griffin-lim iteration," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) (Glasgow: IEEE), 61–65.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289. doi: 10.1038/s41593-018-0209-y

McDermott, J. H., Oxenham, A. J., and Simoncelli, E. P. (2009). "Sound texture synthesis via filter statistics," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New Paltz, NY: IEEE), 297–300.

McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* 71, 926–940. doi: 10.1016/j.neuron.2011.06.032

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., et al. (2016). Samplernn: An unconditional end-to-end neural audio generation model. *arXiv preprint* arXiv:1612.07837.

Mets, D. G., and Brainard, M. S. (2018). An automated approach to the quantitation of vocalizations and vocal learning in the songbird. *PLoS Comput. Biol.* 14:e1006437. doi: 10.1371/journal.pcbi.1006437

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

Miller, E. H. (1979). An approach to the analysis of graded vocalizations of birds. *Behav. Neural Biol.* 27, 25–38. doi: 10.1016/S0163-1047(79)92732-8

Milmine, M., Watanabe, A., and Colombo, M. (2008). Neural correlates of directed forgetting in the avian prefrontal cortex. *Behav. Neurosci.* 122, 199. doi: 10.1037/0735-7044.122.1.199

Mizuhara, T., and Okanoya, K. (2020). Do songbirds hear songs syllable by syllable? *Behav. Processes* 174:104089. doi: 10.1016/j.beproc.2020.104089

Moca, V. V., Bârzan, H., Nagy-Dăbâcan, A., and Mureşan, R. C. (2021). Time-frequency super-resolution with superlets. *Nat. Commun.* 12, 1–18. doi: 10.1038/s41467-020-20539-9

Moll, F. W., and Nieder, A. (2015). Cross-modal associative mnemonic signals in crow endbrain neurons. *Curr. Biol.* 25, 2196–2201. doi: 10.1016/j.cub.2015.07.013

Moll, F. W., and Nieder, A. (2017). Modality-invariant audio-visual association coding in crow endbrain neurons. *Neurobiol. Learn. Mem.* 137, 65–76. doi: 10.1016/j.nlm.2016.11.011

Morfi, V., Lachlan, R. F., and Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. *J. Acoust. Soc. Am.* 150, 2–11. doi: 10.1121/10.0005475

Morita, T., and Koda, H. (2019). Superregular grammars do not provide additional explanatory power but allow for a compact analysis of animal song. *R. Soc. Open Sci.* 6, 190139. doi: 10.1098/rsos.190139

Morita, T., Koda, H., Okanoya, K., and Tachibana, R. O. (2020). Birdsong sequence exhibits long context dependency comparable to human language syntax. *bioRxiv*.

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., et al. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* 385, 217–227. doi: 10.1056/NEJMoa2027540

Muda, L., Begam, M., and Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint* arXiv:1003.4083.

Murcia, R. H., and Paniagua, V. S. (2013). "The ICML 2013 Bird Challenge: Bird identification from continuous audio recordings," in *International Conference on Machine Learning 2013* (Sweden: Citeseer), 26–96.

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., and Shaalan, K. (2019). Speech recognition using deep neural networks: a systematic review. *IEEE Access* 7, 19143–19165. doi: 10.1109/ACCESS.2019.2896880

Nelson, D. A., and Marler, P. (1989). Categorical perception of a natural stimulus continuum: birdsong. *Science* 244, 976–978. doi: 10.1126/science.2727689

Nicholson, D. (2016). "Comparison of machine learning methods applied to birdsong element classification," in *Proceedings of the 15th Python in Science Conference* (Austin, TX: Scipy.org), 57–61.

Okobi, D. E., 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

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., et al. (2016). Wavenet: A generative model for raw audio. *arXiv preprint* arXiv:160 9.03499.

Pagliarini, S., Trouvain, N., Lebois, A., and Hinaut, X. (2021). *What does the Canary Say? Low-Dimensional GAN Applied to Birdsong*. [Epub ahead of print].

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., et al. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* 15, 805–815. doi: 10.1038/s41592-018-0109-9

Park, D. S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E. D., et al. (2019). Specaugment: a simple data augmentation method for automatic speech recognition. *arXiv preprint* arXiv:1904.08779. doi: 10.21437/Interspeech.2019-2680

Patricelli, G. L., and Hebets, E. A. (2016). New dimensions in animal communication: the case for complexity. *Curr. Opin. Behav. Sci.* 12, 80–89. doi: 10.1016/j.cobeha.2016.09.011

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., et al. (2019). Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117–125. doi: 10.1038/s41592-018-0234-5

Petkov, C. I., and Jarvis, E. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4:12. doi: 10.3389/fnevo.2012.00012

Prenger, R., Valle, R., and Catanzaro, B. (2019). "Waveglow: a flow-based generative network for speech synthesis," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton: IEEE), 3617–3621.

Priyadarshani, N., Marsland, S., and Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *J. Avian Biol.* 49, jav-01447. doi: 10.1111/jav.01447

Priyadarshani, N., Marsland, S., Castro, I., and Punchihewa, A. (2016). Birdsong denoising using wavelets. *PLoS ONE* 11:e0146790. doi: 10.1371/journal.pone.0146790

Priyadarshani, N., Marsland, S., Juodakis, J., Castro, I., and Listanti, V. (2020). Wavelet filters for automated recognition of birdsong in long-time field recordings. *Methods Ecol. Evolut*. 11, 403–417. doi: 10.1111/2041-210X.13357

Prša, Z., and Rajmic, P. (2017). Toward high-quality real-time signal reconstruction from stft magnitude. *IEEE Signal Process Lett*. 24, 892–896. doi: 10.1109/LSP.2017.2696970

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint* arXiv:1511.06434.

Rahman, M., Willmore, B. D., King, A. J., and Harper, N. S. (2020). Simple transformations capture auditory input to cortex. *Proc. Natl. Acad. Sci. U.S.A*. 117, 28442–28451. doi: 10.1073/pnas.1922033117

Ren, Y., Johnson, M. T., and Tao, J. (2008). Perceptually motivated wavelet packet transform for bioacoustic signal enhancement. *J. Acoust. Soc. Am*. 124, 316–327. doi: 10.1121/1.2932070

Renteria, S., Vallejo, E., and Taylor, C. (2021). Birdsong phrase verification and classification using siamese neural networks. *bioRxiv*. doi: 10.1101/2021.03.16.435625

Rinnert, P., Kirschhock, M. E., and Nieder, A. (2019). Neuronal correlates of spatial working memory in the endbrain of crows. *Curr. Biol*. 29, 2616–2624. doi: 10.1016/j.cub.2019.06.060

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., and Eck, D. (2018). "A hierarchical latent vector model for learning long-term structure in music," in *International Conference on Machine Learning* (Stockholm: PMLR), 4364–4373.

Rohrmeier, M., Zuidema, W., Wiggins, G. A., and Scharff, C. (2015). Principles of structure building in music, language and animal song. *Philos. Trans. R. Soc. B* 370, 20140097. doi: 10.1098/rstb.2014.0097

Rose, J., and Colombo, M. (2005). Neural correlates of executive control in the avian brain. *PLoS Biol*. 3:e190. doi: 10.1371/journal.pbio.0030190

Russ, B. E., Lee, Y.-S., and Cohen, Y. E. (2007). Neural and behavioral correlates of auditory categorization. *Hear. Res*. 229, 204–212. doi: 10.1016/j.heares.2006.10.010

Sainburg, T. (2019). timsainb/noisereduce: v1.0 (db94fe2). *Zenodo*. doi: 10.5281/zenodo.3243139

Sainburg, T., Mai, A., and Gentner, T. Q. (2020a). Long-range sequential dependencies precede complex syntactic production in language acquisition. *bioRxiv*. doi: 10.1101/2020.08.19.256792

Sainburg, T., McInnes, L., and Gentner, T. Q. (2021). Parametric umap: learning embeddings with deep neural networks for representation and semi-supervised learning. *Neural Comput*. doi: 10.1162/neco_a_01434

Sainburg, T., Theilman, B., Thielk, M., and Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nat. Commun*. 10, 1–11. doi: 10.1038/s41467-019-11605-y

Sainburg, T., Thielk, M., and Gentner, T. (2018a). "Learned context dependent categorical perception in a songbird," in *Conference on Cognitive Computational Neuroscience* (Philadelphia, PA: ccneuro.org), 1-4.

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

Sainburg, T., Thielk, M., and Gentner, T. Q. (2020c). Latent space visualization, characterization, and generation of diverse vocal communication signals. *bioRxiv* 870311. doi: 10.1101/870311

Sainburg, T., Thielk, M., Theilman, B., Migliori, B., and Gentner, T. (2018b). Generative adversarial interpolative autoencoding: adversarial training on latent space interpolations encourage convex latent distributions. *arXiv preprint* arXiv:1807.06650.

Saint-Arnaud, N., and Popat, K. (1995). "Analysis and synthesis of sound textures," in *Readings in Computational Auditory Scene Analysis* (CiteSeer). Available online at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.586

Sasahara, K., Cody, M. L., Cohen, D., and Taylor, C. E. (2012). Structural design principles of complex bird songs: a network-based approach. *PLoS ONE* 7:e44436. doi: 10.1371/journal.pone.0044436

Schneider, S., Baevski, A., Collobert, R., and Auli, M. (2019). wav2vec: Unsupervised pre-training for speech recognition. *arXiv preprint* arXiv:1904.05862. doi: 10.21437/Interspeech.2019-1873

Shannon, C. E. (1951). Prediction and entropy of printed english. *Bell Syst. Techn. J*. 30, 50–64. doi: 10.1002/j.1538-7305.1951.tb01366.x

Shen, H. (2019). Mutual information scaling and expressive power of sequence models. *arXiv preprint* arXiv:1905.04271.

Singh Alvarado, J., Goffinet, J., Michael, V., Liberti, W., Hatfield, J., Gardner, T., et al. (2021). Neural dynamics underlying birdsong practice and performance. *Nature* 599, 635–639. doi: 10.1038/s41586-021-04004-1

Sinnott, J., Beecher, M., Moody, D., and Stebbins, W. (1976). Speech sound discrimination by monkeys and humans. *J. Acoust. Soc. Am*. 60, 687–695. doi: 10.1121/1.381140

Sitt, J., Amador, A., Goller, F., and Mindlin, G. (2008). Dynamical origin of spectrally rich vocalizations in birdsong. *Phys. Rev. E* 78, 011905. doi: 10.1103/PhysRevE.78.011905

Sitt, J. D., Arneodo, E. M., Goller, F., and Mindlin, G. B. (2010). Physiologically driven avian vocal synthesizer. *Phys. Rev. E* 81, 031927. doi: 10.1103/PhysRevE.81.031927

Somervuo, P. (2019). Time-frequency warping of spectrograms applied to bird sound analyses. *Bioacoustics* 28, 257–268. doi: 10.1080/09524622.2018.1431958

Stevens, S. S., Volkmann, J., and Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Am*. 8, 185–190. doi: 10.1121/1.1915893

Stowell, D., Petrusková, T., Šálek, M., and Linhart, P. (2019). Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. *J. R. Soc. Interface* 16, 20180940. doi: 10.1098/rsif.2018.0940

Suzuki, R., Buck, J. R., and Tyack, P. L. (2006). Information entropy of humpback whale songs. *J. Acoust. Soc. Am*. 119, 1849–1866. doi: 10.1121/1.2161827

Tachibana, R. O., Oosugi, N., and Okanoya, K. (2014). Semi-automatic classification of birdsong elements using a linear support vector machine. *PLoS ONE* 9:e92584. doi: 10.1371/journal.pone.0092584

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., and Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Anim. Behav*. 59, 1167–1176. doi: 10.1006/anbe.1999.1416

ten Cate, C. (2014). On the phonetic and syntactic processing abilities of birds: from songs to speech and artificial grammars. *Curr. Opin. Neurobiol*. 28, 157–164. doi: 10.1016/j.conb.2014.07.019

ten Cate, C., and Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. R. Soc. B* 367, 1984–1994. doi: 10.1098/rstb.2012.0055

Thielk, M., Sainburg, T., Sharpee, T., and Gentner, T. (2018). "Combining biological and artificial approaches to understand perceptual spaces for categorizing natural acoustic signals," in *Conference on Cognitive Computational Neuroscience* (Philadelphia, PA: ccneuro.org), 1-4.

Tjandra, A., Sakti, S., and Nakamura, S. (2020). Transformer vq-vae for unsupervised unit discovery and speech synthesis: Zerospeech 2020 challenge. *arXiv preprint* arXiv:2005.11676. doi: 10.21437/Interspeech.2020-3033

Tran, K., Bisazza, A., and Monz, C. (2018). The importance of being recurrent for modeling hierarchical structure. *arXiv preprint* arXiv:1803.03585. doi: 10.18653/v1/D18-1503

Veit, L., and Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nat. Commun*. 4, 1–11. doi: 10.1038/ncomms3878

Versteegh, M., Thiolliere, R., Schatz, T., Cao, X. N., Anguera, X., Jansen, A.,et al. (2015). "The zero resource speech challenge 2015," in *Sixteenth Annual Conference of the International Speech Communication Association* (Dresden, IEEE), 1-12.

von Eugen, K., Tabrik, S., Güntürkün, O., and Ströckens, F. (2020). A comparative analysis of the dopaminergic innervation of the executive caudal nidopallium in pigeon, chicken, zebra finch, and carrion crow. *J. Compar. Neurol*. 528, 2929–2955. doi: 10.1002/cne.24878

Wagener, L., Loconsole, M., Ditz, H. M., and Nieder, A. (2018). Neurons in the endbrain of numerically naive crows spontaneously encode visual numerosity. *Curr. Biol*. 28, 1090–1094. doi: 10.1016/j.cub.2018.02.023

Wang, D., and Chen, J. (2018). Supervised speech separation based on deep learning: an overview. *IEEE/ACM Trans. Audio Speech Lang. Proc.* 26, 1702–1726. doi: 10.1109/TASLP.2018.2842159

Wang, Y., Getreuer, P., Hughes, T., Lyon, R. F., and Saurous, R. A. (2017). "Trainable frontend for robust and far-field keyword spotting," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (New Orleans, LA: IEEE), 5670–5674.

Weiss, M., Hultsch, H., Adam, I., Scharff, C., and Kipper, S. (2014). The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proc. R. Soc. B. Biol. Sci.* 281, 20140460. doi: 10.1098/rspb.2014.0460

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (taeniopygia guttata). *J. Exp. Biol.* 204, 3497–3506. doi: 10.1242/jeb.204.20.3497

Wilson, B., Marslen-Wilson, W. D., and Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. *Trends Neurosci.* 40, 72–82. doi: 10.1016/j.tins.2016.11.004

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., et al. (2015). Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135. doi: 10.1016/j.neuron.2015.11.031

Wiltschko, A. B., Tsukahara, T., Zeine, A., Anyoha, R., Gillis, W. F., Markowitz, J. E., et al. (2020). Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* 23, 1433–1443. doi: 10.1038/s41593-020-00706-3

Wimmer, J., Towsey, M., Planitz, B., Roe, P., and Williamson, I. (2010). "Scaling acoustic data analysis through collaboration and automation," in *2010 IEEE Sixth International Conference on e-Science* (Brisbane, QLD: IEEE), 308–315.

Xiao, J., and Flandrin, P. (2007). Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Trans. Signal Proc.* 55, 2851–2860. doi: 10.1109/TSP.2007.893961

Xie, J., Colonna, J. G., and Zhang, J. (2020). Bioacoustic signal denoising: a review. *Artif. Intell. Rev.* 54, 3575–3597. doi: 10.07/s10462-020-09932-4

Xin, Y., Zhong, L., Zhang, Y., Zhou, T., Pan, J., and Xu, N.-L. (2019). Sensory-to-category transformation via dynamic reorganization of ensemble structures in mouse auditory cortex. *Neuron* 103, 909–921. doi: 10.1016/j.neuron.2019.06.004

Zuidema, W., French, R. M., Alhama, R. G., Ellis, K., O'Donnell, T. J., Sainburg, T., et al. (2020). Five ways in which computational modeling can help advance cognitive science: Lessons from artificial grammar learning. *Top. Cogn. Sci.* 12, 925–941. doi: 10.1111/tops.12474

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fnbeh.2021.814200

Maturation of Social-Vocal Communication in Prairie Vole (*Microtus ochrogaster*) Pups

Megan R. Warren[1,2†], *Drayson Campbell*[1,2†], *Amélie M. Borie*[1,2], *Charles L. Ford IV*[2,3], *Ammar M. Dharani*[4], *Larry J. Young*[2,3] *and Robert C. Liu*[1,2*]

[1] *Department of Biology, Emory University, Atlanta, GA, United States,* [2] *Center for Translational Social Neuroscience, Yerkes National Primate Research Center, Atlanta, GA, United States,* [3] *Department of Psychiatry and Behavioral Science, Emory University School of Medicine, Atlanta, GA, United States,* [4] *Summer Opportunities of Academic Research Program, James T. Laney School of Graduate Studies, Emory University, Atlanta, GA, United States*

Impairments in social communication are common among neurodevelopmental disorders. While traditional animal models have advanced our understanding of the physiological and pathological development of social behavior, they do not recapitulate some aspects where social communication is essential, such as biparental care and the ability to form long-lasting social bonds. Prairie voles (*Microtus ochrogaster*) have emerged as a valuable rodent model in social neuroscience because they naturally display these behaviors. Nonetheless, the role of vocalizations in prairie vole social communication remains unclear. Here, we studied the ontogeny [from postnatal days (P) 8–16] of prairie vole pup ultrasonic vocalizations (USVs), both when isolated and when the mother was present but physically unattainable. In contrast to other similarly sized rodents such as mice, prairie vole pups of all ages produced isolation USVs with a relatively low fundamental frequency between 22 and 50 kHz, often with strong harmonic structure. Males consistently emitted vocalizations with a lower frequency than females. With age, pups vocalized less, and the acoustic features of vocalizations (e.g., duration and bandwidth) became more stereotyped. Manipulating an isolated pup's social environment by introducing its mother significantly increased vocal production at older (P12–16) but not younger ages, when pups were likely unable to hear or see her. Our data provide the first indication of a maturation in social context-dependent vocal emission, which may facilitate more active acoustic communication. These results help lay a foundation for the use of prairie voles as a model organism to probe the role of early life experience in the development of social-vocal communication.

Keywords: ultrasonic vocalization, maternal potentiation, sensory development, volitional, bioacoustics, social isolation, rodent

OPEN ACCESS

Edited by:
Sara Anna Bonini,
University of Brescia, Italy

Reviewed by:
Tai Fa Dao,
Shaanxi Normal University, China
Brian Keane,
Miami University Hamilton, United States
Martha Welch,
Columbia University, United States

****Correspondence:***
Robert C. Liu
robert.liu@emory.edu

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *12 November 2021*
Accepted: *13 December 2021*
Published: *11 January 2022*

Citation:
Warren MR, Campbell D, Borie AM, Ford CL IV, Dharani AM, Young LJ and Liu RC (2022) Maturation of Social-Vocal Communication in Prairie Vole (Microtus ochrogaster) Pups. Front. Behav. Neurosci. 15:814200. doi: 10.3389/fnbeh.2021.814200

INTRODUCTION

Our understanding of the pathological and physiological development of social-vocal communication has improved through the use of traditional animal models such as rats and mice. For instance, variations in the acoustic features of mouse pup vocalizations offer insight into the emotional states of pups, while mouse models of neurodevelopmental disorders and early life stress

display long-term changes in vocal behavior (Branchi et al., 1998; Liu et al., 2003; D'Amato et al., 2005; Kessler et al., 2011; Hernandez-Miranda et al., 2017). However, the translational utility of many rodent models is limited by their expressed social traits. Traditional rodent models fail to capture the breadth and diversity of human behavior wherein social communication is essential, such as biparental care of offspring and the ability to form long-lasting, selective social bonds (Bales et al., 2021).

Prairie voles (*Microtus ochrogaster*) have emerged as a valuable rodent model in social neuroscience due to their natural repertoire of social behaviors. Prairie voles typically form lifelong, socially monogamous relationships with a single mate, exhibit biparental rearing of pups, and engage in alloparental care (McGraw and Young, 2010; Sadino and Donaldson, 2018; Walum and Young, 2018). The neural mechanisms responsible for these behaviors are modulated by neuropeptides, including oxytocin, vasopressin, and dopamine (Nair and Young, 2006; Young et al., 2008; Bosch and Young, 2017; Walum and Young, 2018), which are important for social bonding and communication in many species (Albers, 2012; Oettl et al., 2016; Marlin and Froemke, 2017; Froemke and Young, 2021; Nagasawa and Kikusui, 2021), including humans. In fact, studies using prairie voles and other vole species have been vital in furthering our understanding of the neural basis of affiliative behaviors. However, despite the volume of neurobiological and ethological studies on prairie voles, the role of specific sensory modalities in social communication during affiliative behaviors is not well-understood. It is notable that audition occupies a disproportionally large portion of the prairie vole sensory cortex as compared to other rodents, occupying about twice as much space as the mouse auditory cortex (Campi et al., 2007; Krubitzer et al., 2011). As such, vocalizations may play a larger role in prairie vole social communication as compared to other rodents.

Prairie voles, like other rodents, emit ultrasonic vocalizations (USVs) above the range of human hearing during social interactions (Holy and Guo, 2005; Arriaga and Jarvis, 2013; Rieger and Marler, 2018). Adult males, similar to other rodent models (Chabout et al., 2012, 2015; Warren et al., 2021), modify their vocal activity based upon their social context (Lepri et al., 1988; Ma et al., 2014), increasing both the number and complexity of USVs in the presence of an unfamiliar female. Males do not, however, change their vocal behavior in the presence of a same-sex sibling (Ma et al., 2014). Females, in contrast, modify vocal emission when exposed to either a same- or opposite-sex social partner. However, both male and female voles increase their vocal output when injected with amphetamine, indicating that vocal emission by adult voles may be generally linked to appetitive conditions (Ma et al., 2014).

In general, young rodent pups are highly vocal and reliably emit isolation-induced vocalizations when outside the nest. Isolation USVs function to elicit maternal retrieval (Sewell, 1970; Shapiro and Insel, 1990; Brunelli et al., 1994; Ehret, 2005; Bowers et al., 2013; Schiavo et al., 2020). Mouse and rat pups that vocalize at high rates are retrieved more rapidly than less vocal pups (Bowers et al., 2013), while silent pups are not reliably retrieved (Zippelius and Schleidt, 1956; Brooks and Banks, 1973). Thus, these sounds provide a vital means of communication between pups and mothers. Interestingly, prairie vole pups emit high numbers of USVs when separated from their mother and this increase in USV emission is highly correlated with increased levels of stress hormones (Shapiro and Insel, 1990). Non-monogamous montane voles (*Microtus montanus*), in contrast, show slightly increased corticosterone levels when separated from their mothers but not a concomitant increase in vocal emission (Shapiro and Insel, 1990). Additionally, the features of vocalizations can be a readout of early-life experience. For instance, early-life maternal separation (P2–P10) in rat pups leads to modified vocal features (Kaidbey et al., 2019). Thus, USVs may provide a useful indicator of early-life social bonds, though the precise conditions that elicit their emission and the development of context-dependent emission are not well-understood.

In this study, we characterized the ontogeny of USVs emitted by prairie vole pups and assessed how they employ them as a means of communicating with their mother. Using a novel behavioral paradigm, we measured the vocal activity of pups from P6–P16, either in isolation or when their mother was present but physically separated. We characterized how vocalizations changed over the course of development as a function of sex, as well as the role of social context in modifying vocal emissions, and mapped these changes onto developmental milestones.

MATERIALS AND METHODS

Subjects

Prairie vole pups (6–16 days after birth) were used to examine the ontogeny of social-vocal communication. 111 pups originating from 59 litters across 15 breeding pairs were used for behavioral experiments. A maximum of two pups per litter were recorded at the same age, and no pup was recorded more than once. Pups were returned to the parents after recordings to be used in other studies. All animals originated from a laboratory breeding colony that was derived from field-captured voles in Champaign, Illinois.

Animals were kept on a 14/10 h light/dark cycle at 68–72°F and 40–60% humidity with *ad libitum* access to food (Laboratory Rabbit Diet HF #5326, LabDiet, St. Louis, MO, USA) and water. Cages were filled with Bedo'cobbs Laboratory Animal Bedding (The Andersons; Maumee, Ohio) and contained environmental enrichment, including cotton pieces to allow nest building. Pups were kept in breeding cages with their parents until weaning at 20–23 days of age, then group-housed (2–3 per cage) with age-matched pups of the same sex.

All experiments were performed during the light cycle between 9 a.m. and 5 p.m. Experiments were conducted in strict accordance with the guidelines established by the National Institutes of Health and then approved by Emory University's Institutional Animal Care and Use Committee.

Data Collection

All data collection occurred in a designated behavioral-recording room separate from the animal colony, with the home cage placed at the opposite side of the room where it did not generate sounds

that could contaminate our recordings. To record isolation-induced ultrasonic vocalizations (USVs), individual pups (P6: $n = 19$; P8: $n = 19$; P10: $n = 14$; P12: $n = 16$; P14: $n = 19$; P16: $n = 14$. P = postnatal day, or days after birth, with P0 being the day of birth) were removed from their home cage and placed in a plexiglass recording chamber (14 × 16.5 × 14 cm) lined with clean Alpha-DRI bedding for 10 min while audio and video data were recorded. Approximately equal numbers of males and females were recorded at each age. Recordings began at P6 to minimize nest disturbances at very young ages. A subset of the pups, aged P8–P16 ($n = 11$, 11, 13, 12, and 12 for P8, P10, P12, P14, and P16, respectively) were also used to assess the effect of social context on pup-emitted vocalizations. Therefore, after the 10 min of isolation, pups' mothers were introduced into an adjacent, identical chamber for another 10-min recording. The two chambers were physically separated by a transparent plexiglass wall with a single hole in its center, 1 cm in diameter and 3 cm above the floor. During all recordings, animals had access to food (Laboratory Rabbit Diet HF #5326) and water gel (Clear H2O Scarborough; Scarborough, ME).

A microphone (Avisoft CM16/CMPA microphone) was placed into the pup's chamber to record audio data. Audio was sampled at 300 kHz (Avisoft-Bioacoustics; Glienicke, Germany; CM15/CMPA40-5V), and an UltraSoundGate (116H or 416H used for all recordings, Avisoft-Bioacoustics; Glienicke, Germany) data acquisition system was used and integrated with Avisoft-RECORDER software to store the data. A video camera (Canon Vixia HF R800) recorded a side-view video of the pup chamber at 30 fps.

In a set of control recordings, the same paradigm was used, but a second microphone was placed into the mother's chamber. The UltraSoundGate 416H was used to simultaneously record audio data from both microphones.

The datasets presented in this study can be found in online repositories. The name of the repository can be found in the data availability statement.

Vocal Extraction

To extract vocal segments (continuous units of sound), audio files were processed with USVSEG, an open-source MATLAB-based USV detection and analysis program (Tachibana et al., 2020). Files were band-pass filtered between 15 and 125 kHz. Audio was characterized at a time step of 0.5 ms and sounds with fewer than 6 samples (corresponding to sounds shorter than 3 ms) were excluded. USVSEG was modified to produce a structure for each vocal segment, containing the time value of each sample, the frequency of sound at each sample, and the corresponding sound amplitude. This structure is referred to as a frequency contour. Due to the high number of harmonics (sounds at integer multiples of the lowest frequency tone) present in prairie vole pup vocalizations, USVSEG was also modified to allow up to seven unique sound frequencies to be extracted at each time point. The resultant structure was organized based on sound frequency at each sample, such that each contour contained a fundamental frequency contour tracing of the lowest frequency sound at each point, and separate traces for simultaneous sounds at higher frequencies.

Frequency contours were further refined using custom-written MATLAB scripts. Manual inspection of the data allowed us to determine optimal thresholds to filter out extracted noise such that >97% of extracted sounds were true vocalizations while only excluding 3% of vocalizations (data not shown; data verified by two trained viewers). Sounds with a median fundamental (lowest) frequency below 22 kHz were also excluded, as these most often corresponded to noise. Thus, all extracted USVs fell between 22 and 125 kHz. To further refine the extraction of vocal segments, successive sounds with instantaneous jumps in frequency (changes in sound frequency between two successive sound samples) exceeding 8 kHz or periods of silence exceeding 4 ms were separated into distinct segments. All analyses were conducted based upon the fundamental contour of each segment.

For experiments where two microphones were used, USVs were separately extracted from each microphone channel. Then, if a sound was picked up simultaneously on both microphones, the sound was attributed to the chamber on which the sound amplitude was greatest. This method was verified using sound playback from speakers, and side attribution accuracy was found to be >95%. The number of USVs emitted at each age did not significantly differ based on whether mother-emitted sounds were included (paired t-tests; $-2.7 < t < -1.6$, all $p > 0.05$; data not shown), so all USVs were attributed to the pup in the one-microphone recordings.

Developmental Milestones

A separate cohort of pups ($n = 10$, 5 males and 5 females across 4 litters) was used to characterize the developmental trajectory of prairie vole pups. For these experiments, individual pups were recorded every 2 days from P6 to P16 to track their developmental trajectory (thus, each pup was recorded 6 times). For each recording, individual pups were removed from their home cage, their temperatures were measured via a PhysioSuite (Kent Scientific) temperature sensor, the pups were weighed, then placed into the plexiglass recording chamber. The chamber was lined with Alpha-DRI bedding and contained a food pellet and piece of water gel. Pups remained in the chamber for 10 min, after which their temperature was taken again, with temperature change used as a proxy for thermoregulation.

Starting at P6, a speaker was placed next to the open side of the recording chamber to assess functional hearing *via* an acoustic startle paradigm. Instead of using the plexiglass divider to seal the chamber, the open side of the chamber was enclosed using two stacked metal mesh blocks, as this would allow sound to efficiently propagate into the chamber. After a pup's second temperature reading, the pup was returned to the chamber and playback began on the speaker. The playback file consisted of 2 min of silence, followed by 50 ms of white noise. During playback, audio (Avisoft microphone, see above) and video (captured by Basler acA 1920-150uc camera) data were recorded to provide video confirmation of a startle response. Pups were assessed for a visible response to the noise, which consistently presented as the pup jumping. Playback occurred on every recording day from P6 until all pups within a litter (2–3 pups) startled.

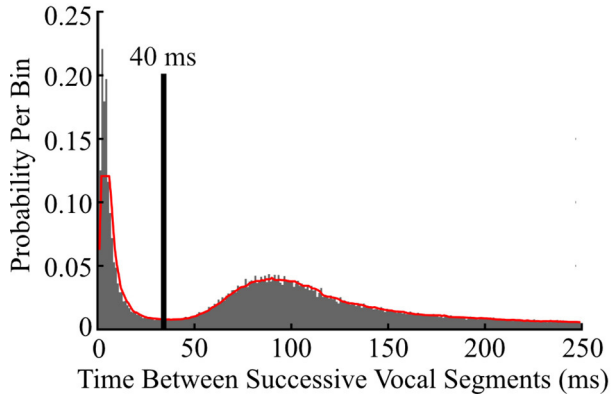


FIGURE 1 | Determining a temporal threshold to consolidate successive vocal segments into vocalizations. To determine the optimal interval for collapsing successive vocal segments into vocalizations, a histogram of the time between all pairs of successive vocal segments (gray) was generated, smoothed (red trace), and the first local minimum (black line) was used as the consolidation threshold. Consecutive segments separated by 40 ms or less were subsequently consolidated into a single vocalization.

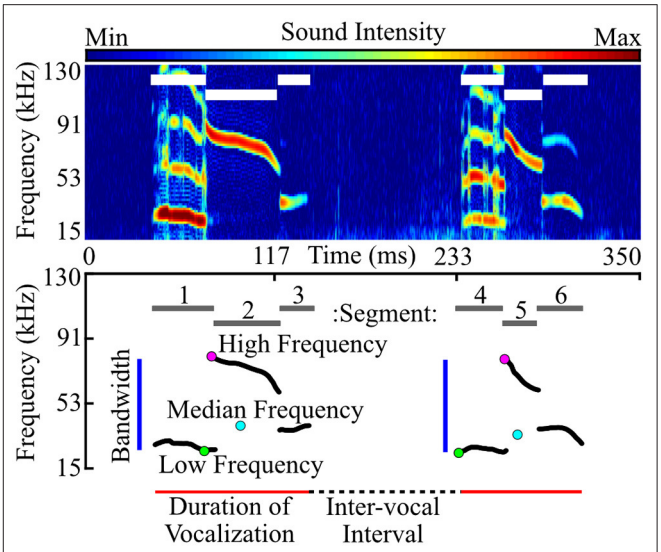


FIGURE 2 | Vocal features were characterized by generating a fundamental frequency contour of each vocalization. Using recorded audio data (represented as a spectrogram; top), we extracted vocal segments, or continuous units of sound (white lines in top image show segment duration, corresponding to gray lines in bottom). Each segment was represented by its fundamental frequency contour (bottom; black lines), tracing the fundamental (lowest) frequency at each time point. Note that harmonics (frequencies at integer multiples of the fundamental) are not represented in the fundamental trace. Segments separated by 40 ms or less were then consolidated into vocalizations (bottom; red lines show duration), and the corresponding segment contours were collapsed to generate a single contour for each vocalization. Contours were used to characterize the acoustic features of each vocalization, as represented above. Inter-vocal interval is the time between two successive vocalizations. High, median, and low frequency are the highest, median, and lowest frequency within each fundamental contour, respectively. Bandwidth is the frequency range, or high frequency minus low frequency.

Pups were also visually assessed at each age for ear and eye opening. Ear opening was defined as having a visible ear canal, and eye opening was defined as having the eyes fully open.

Analyses

Converting Vocal Segments Into Vocalizations

To determine thresholds for consolidating segments into vocalizations, a histogram of the silent intervals between successive segments was created and smoothed with a 5th-order one-dimensional median filter. The first local minimum was used as the threshold for consolidating two adjacent segments into a single vocalization (**Figure 1**). This consolidation threshold was then applied to the entire library of segments, yielding a library of vocalizations. The frequency contours of consolidated segments were unaltered during consolidation. Therefore, each vocalization in the library was described by the fundamental frequency contours of its component segments.

Quantifying Acoustic Features of Vocalizations

Fundamental contours were used to quantify the acoustic features of each vocalization (**Figure 2**). These features included duration (the end time minus the start time), median frequency, high frequency (the highest frequency across the fundamental contour), low frequency (the lowest frequency across the fundamental contour), bandwidth (the high frequency minus the low frequency), and slope (the average partial derivative across the contour). We also assessed the number of harmonics, sounds with frequencies falling approximately at integer multiples of the fundamental frequency (e.g., 1.8–2.2x, or 2.8–3.2x, etc.). Harmonics can be seen in **Figures 2**, **3**. Harmonic sounds lasting fewer than 3 ms were excluded. The mode number of harmonics, or the most common number of harmonics across the entirety of the vocalization, was used to characterize harmonicity.

To control for the possibility of inflated significance with increased sample sizes (Bakan, 1966), as well as differences in the number of vocalizations emitted by individual pups, we compared vocal features at the level of individual animals. Thus, for each acoustic feature we found the average value across all vocalizations emitted by a single pup. This process was repeated for each pup and each feature, allowing individual animals to be represented by their own average value. Group differences in vocal features were then assessed as a function of both the age and sex of each pup. If a main effect of sex was not found, male and female data were collapsed. Assessment of vocal features across social context was conducted using the same method, except data were extracted separately for the 10 min of isolation vs. the 10 min when the mother was present.

Quantifying Vocal Emission Levels

Vocal emission levels were characterized multiple ways. First, to assess the impact of age on vocal rates, we counted the total number of vocalizations emitted by pups over the 10-min isolation period. To characterize changes in vocal emission as a function of social context, the number of vocalizations emitted

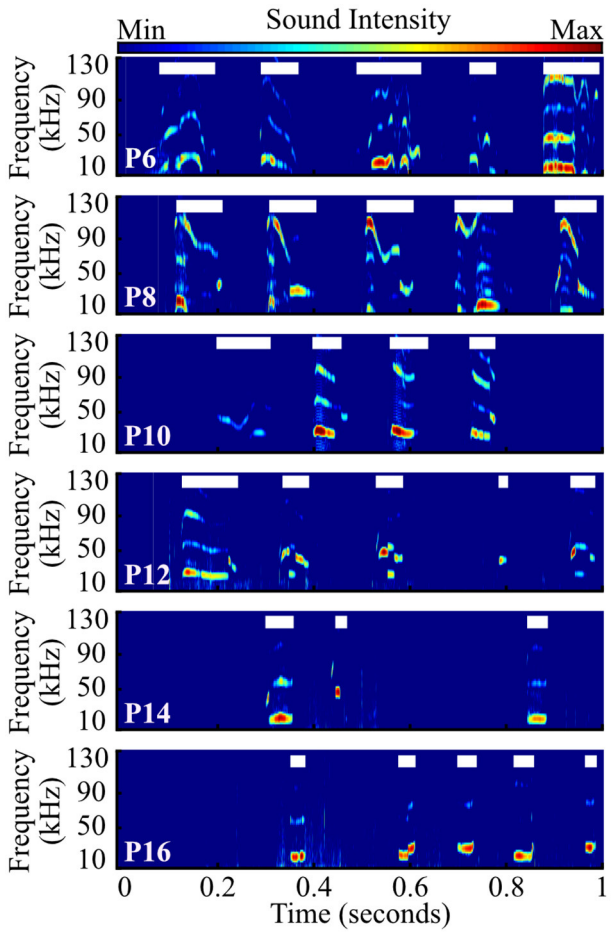


FIGURE 3 | Representative examples of pup vocal activity across development. Example spectrograms contain one second of audio data recorded from prairie vole pups aged P6–P16. Each thick white line shows time when a single vocalization was extracted.

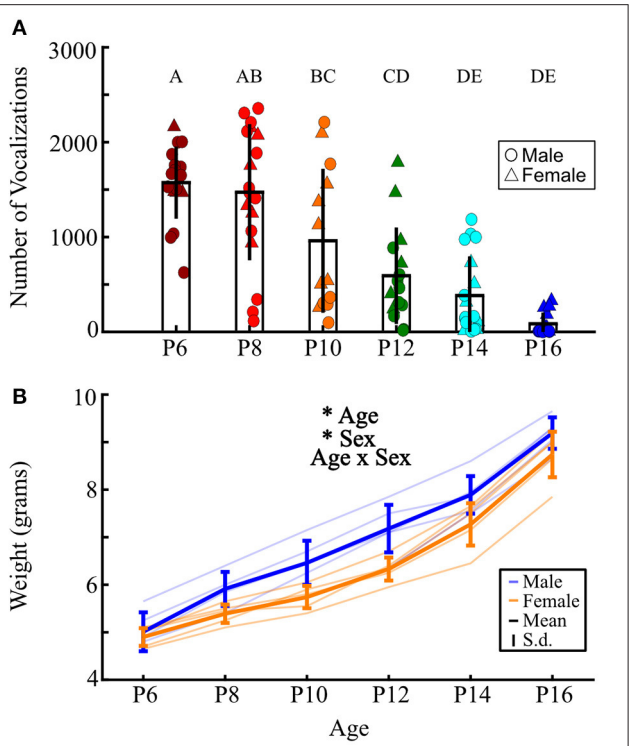


FIGURE 4 | Older pups emit fewer isolation-induced vocalizations than young pups. **(A)** Each point represents the total number of vocalizations emitted by a single pup across 10 min of isolation; males are circles, females are triangles. Data presented as mean ± 1 standard deviation. $n = 19$, 19, 14, 16, 19, and 14 for P6, 8, 10, 12, 14, and 16, respectively. There was no main effect of sex (two-way ANOVA, age-by-sex), so male and female data were collapsed. Ages sharing a letter do not significantly differ from each other ($p \geq 0.05$, Tukey's HSD *post-hoc* correction), while all other comparisons are significantly different ($p < 0.05$). **(B)** Weights of pups across development. Each thin line represents an individual pup recorded over time. Thick blue lines show mean ± 1 s.d. for males, orange show females. Two-way ANOVA found a main effect of age and sex (all $p \leq 10^{-3}$), with no significant interaction. $^*p < 0.05$.

while the mother was present was also counted. Lastly, to assess vocal emission on a finer timescale, we also determined the number of vocalizations emitted by pups in each minute of the audio recording.

Quantifying Variability in Vocal Features

To quantify the level of variability in individual acoustic features, we first extracted the feature of interest (e.g., bandwidth) from all vocalizations emitted by a single pup. We then calculated the variance (sum of squares divided by the number of vocalizations minus 1) of that feature. This analysis was repeated for each pup and each feature of interest (duration, bandwidth, and number of vocal segments per vocalization).

Statistics

Data are shown as mean ± standard deviation unless otherwise indicated. Comparisons between ages and sexes were done with two-way ANOVAs. If a main effect of sex was not found, data were collapsed across sex and then analyzed with a one-way ANOVA. All one-way ANOVAs were combined with a *post-hoc* Tukey's HSD to correct for multiple comparisons. Comparisons of changes in temperature over time used one-tailed *t*-tests to determine whether group distributions fell significantly below zero (indicating a lack of thermoregulation). The vocal emission of individual pups recorded across contexts was compared with a mixed factorial ANOVA, then pairwise comparisons were conducted via paired two-tailed *t*-tests with a Benjamini-Hochberg *post-hoc* correction. The number of vocalizations emitted per time bin was averaged across context, and then the values were compared with a paired *t*-test. Changes in acoustic features across contexts were assessed via z-tests, comparing the distributions of differences to zero, or no difference. A mixed factorial ANOVA was used to compare changes in vocal emission across social context phases for different ages, with a *post-hoc* Benjamini-Hochberg test conducted solely on paired *t*-tests comparing vocal emission across context within each age group. All alpha values were set to 0.05. All statistical analysis was performed with MATLAB (Mathworks).

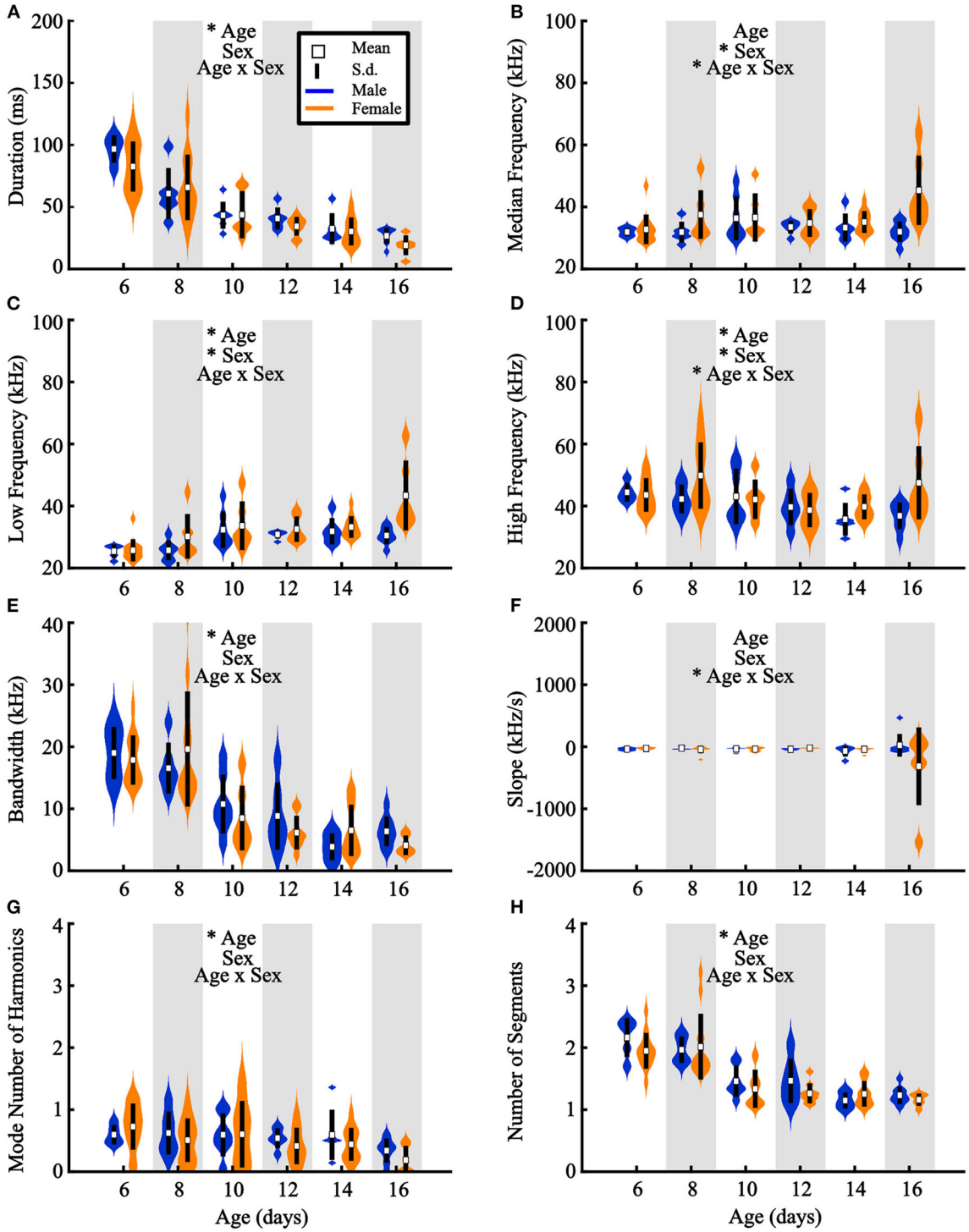


FIGURE 5 | Acoustic features of pup vocalizations differ across development. Individual pups are represented by the average value across all vocalizations emitted by that pup. Violin plots represent feature distributions for **(A)** duration, **(B)** median frequency, **(C)** low frequency, **(D)** high frequency, **(E)** bandwidth, **(F)** slope, **(G)** number of harmonics, and **(H)** number of segments per vocalization, sorted by age and sex. Wider locations indicate a higher number of animals exhibiting that average value. *Age = main effect of age, *Sex = main effect of sex, *Age x Sex = significant interaction term (two-way ANOVA). White squares represent group means and vertical black lines represent ± 1 standard deviation. Males are presented in blue, females in orange. *$p < 0.05$.

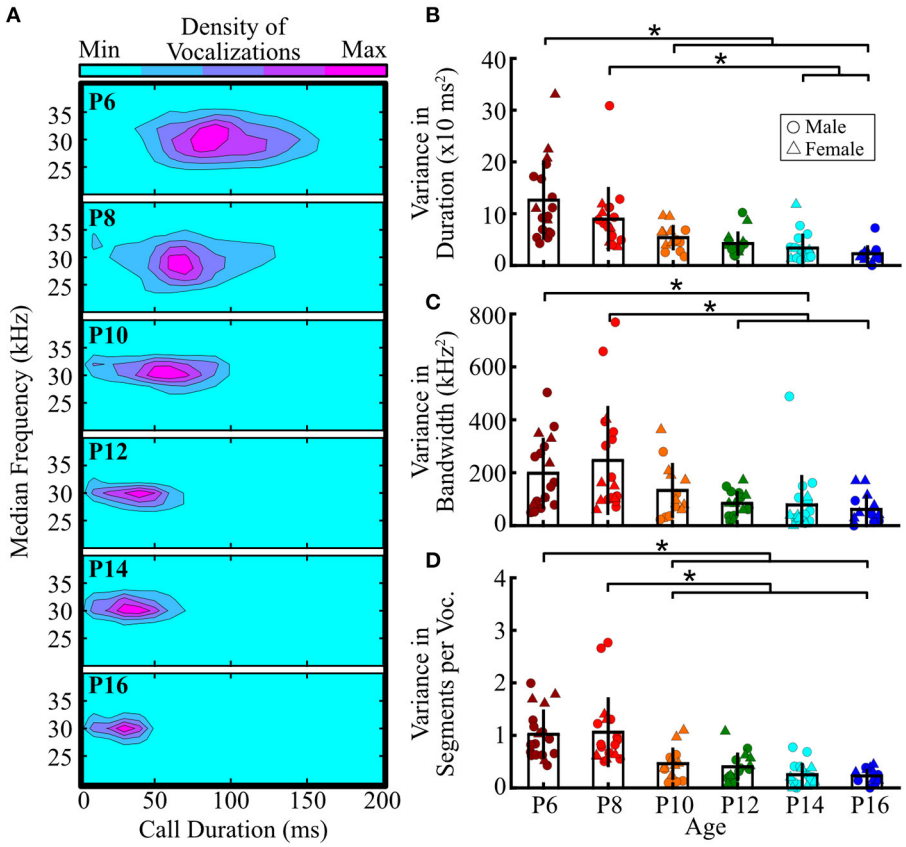


FIGURE 6 | Pup isolation vocalizations become more stereotyped with age. **(A)** Distributions of median frequency and duration of pup vocalizations across ages. Warmer colors indicate greater numbers of vocalizations falling within the region. **(B–D)** Within-animal variability across ages for **(B)** vocal duration, **(C)** bandwidth (high minus low frequency of the fundamental contour), and **(D)** number of segments per vocalization. Each dot represents the variability for a single pup, measured by finding the variance of that feature (sum of squares/n-1) across all calls emitted by that pup, with higher values indicating more variability. There was no main effect of sex for any comparison, so male and female data were collapsed. Bars represent group means, and vertical lines represent standard deviations. Males are represented by circles, females by triangles. *$p < 0.05$ (ANOVA with Tukey's HSD *post-hoc* correction).

RESULTS

Older Pups Emit Fewer Isolation-Induced Vocalizations Than Younger Pups

We recorded 84,294 isolation-induced vocalizations from 101 prairie vole pups, distributed over sex and ages P6–P16 (**Figures 3**, **4**). Males and females emitted similar numbers of vocalizations [$F_{(1, 84)} = 2.39$, $p = 0.13$, two-way ANOVA], so count data were collapsed across sex. We found that older pups emitted significantly fewer vocalizations than younger pups [**Figure 3**; $F_{(5,90)} = 21.43$, $p < 10^{-10}$; one-way ANOVA; **Figure 4**]. P6 pups emitted significantly more vocalizations than pups from P10 to P16 (all $p \leq 0.02$; Tukey's HSD), as did P8 pups compared to those aged P12–P16 (all $p \leq 10^{-5}$), and P10 pups compared to P14 and P16 pups (all $p \leq 0.038$). Thus, similar to other rodent species (Naito and Tonoue, 1987; Campbell et al., 2014; Johnson et al., 2017) and prior reports in prairie voles (Rabon Jr et al., 2001; Terleph, 2011), pup vocal emission decreases with age.

Acoustic Features Differ by Sex and Across Development

Considering all vocalizations regardless of age, USVs had median frequencies of 33 (±8 kHz standard deviation) for males and 34 kHz (±9 kHz) for females. The average range of male frequencies fell between 28 and 42 kHz (±8 and 14 kHz), while female vocalizations spanned 27 and 45 kHz (±8 and 16 kHz, respectively). 82.2% of prairie vole USVs contained at least one harmonic component, and 24.9% contained at least 3 harmonics, in contrast to USVs of other rodents like mice, which typically have no harmonics. Hence, prairie vole pups emit vocalizations with different spectral characteristics than other traditionally used rodents.

Across development, pups across species typically exhibit acoustic changes in their vocalizations (Liu et al., 2003; Yurlova et al., 2020). For prairie vole vocalizations, we ran two-way ANOVAs assessing differences in acoustic features as a function of both sex and age (see section Methods, **Figure 2**). Females emitted vocalizations that were significantly higher in low [$F_{(1, 84)}$

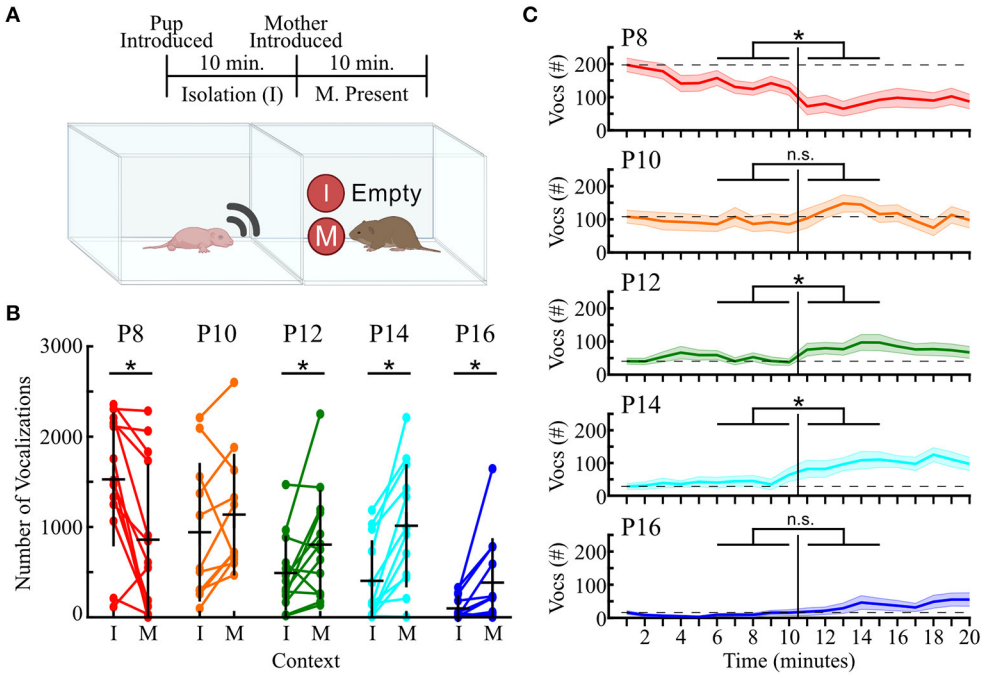


FIGURE 7 | Vocal responses to social context change with pup age. **(A)** Timeline (top) and schematic (bottom) of the two recording contexts. I, isolation; M, mother present. Pup was audio- and video-recorded in each context. **(B)** Dots indicate the number of calls emitted while a pup was in isolation (left; I) or while their mother was in an adjacent but physically distinct chamber (right; M), with lines connecting the same pup across contexts. $n = 11$, 11, 13, 12, and 12 for P8, 10, 12, 14, and 16, respectively. Horizontal black lines represent group means, and vertical black lines represent standard deviations. Mixed factorial ANOVA indicated a main effect of age and social context, and a significant age-by-context interaction (all $p \leq 0.001$). Then paired *t*-tests compared contexts within age. *Significant comparison after a Benjamini-Hochberg *post-hoc* correction. **(C)** Number of vocalizations emitted by pups in each 1-min bin over the 20-min recording. Colored horizontal lines indicate mean values across animals, with shaded areas representing ± 1 SEM. Vertical black line indicates time when the mother was introduced. Horizontal dashed line shows the average number of USVs emitted during the first minute of isolation. Colors as in **(B)**. *$p < 0.05$ when comparing the total number of vocalizations in the final 5 min of isolation (left) vs. the first 5 min after the mother was introduced (right) across all pups (paired *t*-test with Benjamini-Hochberg *post-hoc* correction for multiple comparisons).

$= 10.39$, $p = 0.002$], median [$F_{(1, 84)} = 9.94$, $p = 0.002$], and high frequency [$F_{(1, 84)} = 4.76$, $p = 0.032$] than males (**Figures 5B–D**). Interestingly, we found in a separate cohort that males also weighed significantly more than females {**Figure 4B**; two-way ANOVA, main effect of age [$F_{(5,47)} = 119.7$, $p < 10^{-5}$] and sex [$F_{(1, 47)} = 25.43$, $p < 10^{-5}$]}, suggesting a size difference might contribute to differences in vocal frequencies. Regardless, in agreement with our analysis-by-vocalization, our analysis-by-individual confirmed that the frequency distribution of male vocalizations is downshifted from that of their female littermates.

We next observed a significant main effect of age in six assessed vocal characteristics (**Figure 5**). Older pups emitted vocalizations with shorter durations [$F_{(5, 84)} = 34.34$, $p < 10^{-10}$], lower high frequencies [$F_{(5, 84)} = 3.07$, $p = 0.01$], higher low frequencies [$F_{(5,84)} = 8.17$, $p < 10^{-10}$], lower bandwidths [$F_{(5,84)} = 23.09$, $p < 10^{-10}$], fewer harmonics [$F_{(5, 84)} = 2.53$, $p = 0.04$], and fewer segments [$F_{(5, 84)} = 9.43$, $p < 10^{-10}$] than younger pups.

Additionally, three acoustic features exhibited a significant age-by-sex interaction: median frequency [$F_{(5,84)} = 2.69$, $p = 0.03$], low frequency [$F_{(5,84)} = 2.78$, $p = 0.02$], and slope [$F_{(5,84)} = 2.59$, $p = 0.03$] (**Figure 5**). Thus, along with the decrease in vocal emission at older ages, our results indicate that isolation-induced USVs become shorter, simpler in structure, and less frequent as pups age, defining a developmental trajectory of prairie vole pup vocal emission from P6.

Acoustic Features of Vocalizations Become More Stereotyped With Age

One intriguing feature of pup vocalizations is that for many acoustic features, the distributions seemed narrower for older pups, indicating a potential change in the stereotypy of vocal features with age (**Figure 6A**). Therefore, we next assessed age-related differences in the acoustic variability of duration, bandwidth, and the number of segments per vocalization – collapsing data across sex [no main sex effects, $0.4 \leq F_{(1,84)} \leq 3.0$, all $p \geq 0.09$; two-way ANOVA]. We found that the range of vocal durations significantly decreased with age [$F_{(5,84)} = 9.1$, $p < 10^{-5}$; one-way ANOVA; **Figure 6B**]. P6 pups emitted vocalizations with significantly greater variability in duration than pups aged P10 to P16 (all $p \leq 0.01$; Tukey's HSD), as did P8 pups compared to P14 and P16 pups (all $p = 0.01$). Therefore, not only do vocalizations become shorter as pups age, but vocal duration also becomes more consistent.

Vocal bandwidths also became more stereotyped with age [$F_{(5,84)} = 4.67$, $p < 0.001$; one-way ANOVA; **Figure 6C**]. P6

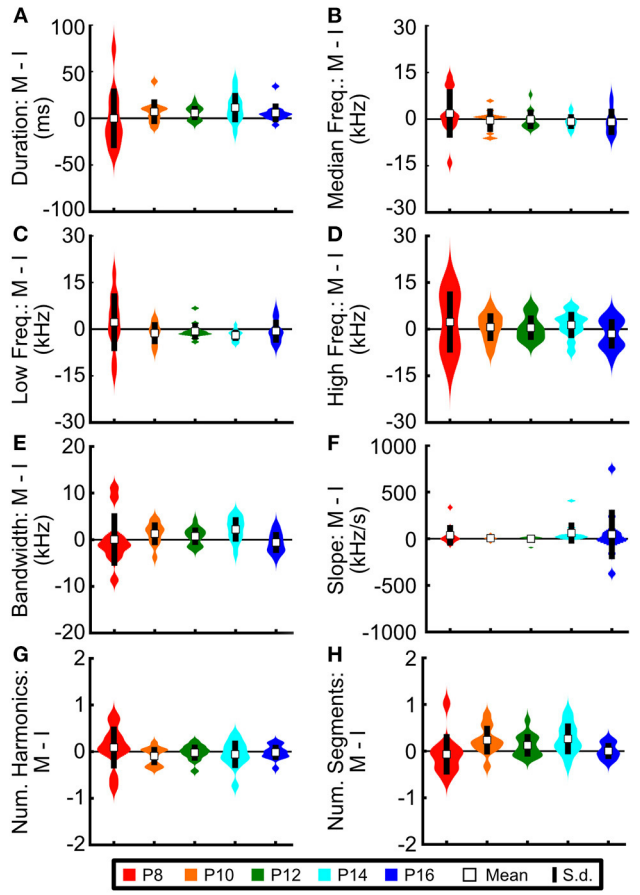


FIGURE 8 | Prairie vole pups do not modify the acoustic features of USVs across social contexts. Difference in average acoustic features for individual pups between the isolated context and the mother-present context for **(A)** duration, **(B)** median frequency, **(C)** low frequency, **(D)** high frequency, **(E)** bandwidth, **(F)** slope, **(G)** number of harmonics, and **(H)** number of segments per vocalization, calculated by subtracting the average value of each individual pup in the isolated context (I) from the same pup's average value during the mother-present condition (M). No distributions differed significantly from zero (z-test, all $z \leq 1.09$).

pups emitted vocalizations with significantly greater variability in bandwidth than P14 pups ($p < 0.05$, Tukey's HSD), as did P8 pups relative to those aged P12–P16 (all $p < 0.02$). P10 pups, however, had intermediate values that did not significantly differ from pups at any other age. Thus, pups emit vocalizations with the greatest variability in bandwidth at young ages, and the bandwidth gradually becomes more stereotyped with age.

Lastly, we assessed changes in variability in the number of segments per vocalization and found that younger pups showed the greatest variability in the number of segments within individual vocalizations [$F_{(5, 84)} = 10.5$, $p < 10^{-7}$; one-way ANOVA; **Figure 6D**]. Both P6 and P8 pups emitted vocalizations that were significantly more variable than P10–P16 pups (all $p \leq 0.02$). Taken together, these results indicate that the acoustic features of prairie vole pup vocalizations, including vocal duration, bandwidth, and the number of unique sound units per vocalization, become more stereotyped at older ages. Thus, not only do the overall acoustic features of pup vocalizations change with age, but the feature distributions also narrow, indicating a potential reliance on more standardized vocalizations at older ages.

Pups Modulate Vocalizations in an Age and Context-Dependent Fashion

In other rodent models, ultrasonic vocalizations are thought to act as a means of social communication, with adult animals modifying the rate of emission based upon social context (Whitney et al., 1974; Nyby et al., 1979; Whitney and Nyby, 1979), and isolated pups using USVs to elicit retrieval responses from adults (Zippelius and Schleidt, 1956; Sewell, 1970). Therefore, we next aimed to determine whether prairie vole pups modify their vocalizations as a function of their social context, and whether any social-context dependent changes differ as a function of age. To this end, after the isolation, a subset of pups aged P8–P16 ($n = 59$) had their mother introduced into a chamber that was attached to the pup's recording chamber (**Figures 7A,B**).

We found that introduction of the pup's mother led to different effects on vocal activity based on pup age. A mixed factorial ANOVA (age-by-context) uncovered a main effect of age [$F_{(4, 56)} = 5.34$, $p < 10^{-4}$] and social context [$F_{(1, 56)} = 5.35$, $p = 0.02$], as well as a significant age-by-context interaction [$F_{(4, 56)} = 11.95$; $p < 10^{-6}$]. At P8, introducing the mother significantly decreased pup vocal output, with pups dropping from an average of 1,527 ± 742 vocalizations in isolation to only 859 ± 841 vocalizations with the mother present [$t_{(13)} = 3.72$, $p = 0.003$, paired t-test]. In contrast, pups aged P12–P16 significantly increased their vocal emissions (all $p \leq 0.03$). P10 was the only age group that did not show a significant change in vocal emission when transitioning from isolation to the mother-present condition [$t_{(10)} = -1.37$, $p = 0.20$]. Taken together, these results provide the first indication of a developmental maturation in the social context-dependent vocal emission of prairie voles.

To determine whether introducing the mother led to rapid changes in vocal activity, we next compared the total number of vocalizations emitted in the 5 min directly preceding the introduction of the mother to the 5 min directly following (**Figure 7C**). At P8, we found a significant reduction in the number of USVs between these periods (paired t-test, Benjamini-Hochberg *post-hoc* correction; $p < 0.002$). In contrast, no significant difference existed for P10 animals. P12 and P14 pups, however, exhibited a significant increase in vocal emission upon introduction of the mother ($p < 0.02$). P16 pups did not show a significant modification in vocal emission ($p = 0.11$), but the number of USVs emitted by P16 animals was lower than at any other age point. Therefore, this lack of significance may be due to low emission levels, as we do see a significant difference if we expand our temporal window to 10 min (**Figure 7B**). Overall, these data indicate that introduction of the mother leads to a rapid and lasting change in the vocal emission levels of prairie vole pups that had just endured a short-term social isolation.

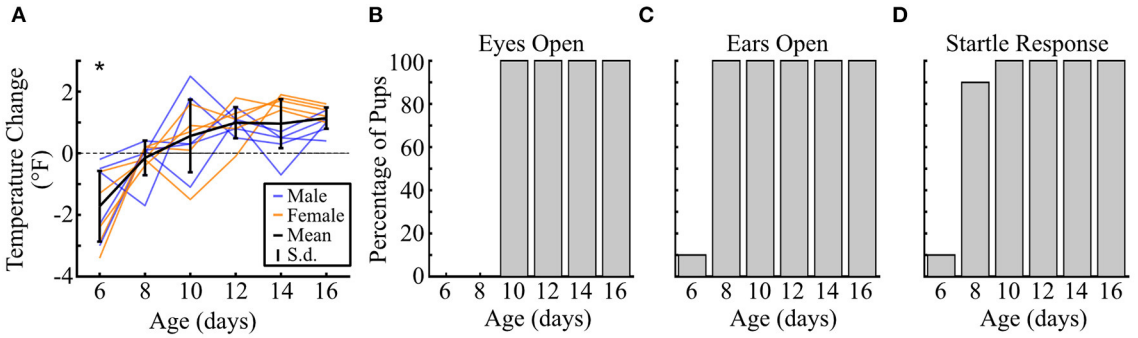


FIGURE 9 | Prairie vole pup developmental timeline. **(A)** Change in temperature of individual pups after 10 min of isolation. Negative values indicate a decrease in temperature. Colored lines represent individual animals over time, with males in blue ($n = 5$) and females in orange ($n = 5$). Solid horizontal black line indicates mean value, with vertical lines showing standard deviation. Dashed line shows location of 0, or no change in temperature. *$p < 0.05$, distribution compared to zero with a one-tailed t-test. **(B)** Bar plots show the percentage of pups at each age **(B)** with open eyes, **(C)** with open ears, and **(D)** exhibiting a startle response to a loud sound.

Acoustic Features of Vocalizations Do Not Differ Across Social Contexts

Other rodent species show changes in the acoustic features of their vocalizations based upon their social context (Chabout et al., 2012, 2015; Warren et al., 2021). Therefore, we next aimed to determine whether this also held for prairie vole pups. In fact, no feature showed significant differences across context (**Figure 8**; all z-scores ≤ 1.09). Thus, while pups modulate the number of vocalizations emitted based upon social context, context alone is not sufficient to modify those vocalizations' acoustic features.

Context-Dependent Vocal Changes Are Preceded by Key Developmental Milestones

One core need experienced by many rodent pups is body temperature maintenance, and temperature decreases have been associated with increased USV production in other rodents (Okon, 1970; Hofer and Shair, 1992). Vocal emission may be linked to the process of generating heat itself (Blumberg and Alberts, 1990; Hofer and Shair, 1993) or simply serve as a communication signal to mothers for retrieval back to the warmth of the nest (Ehret, 2005). The onset of thermoregulation is therefore a vital developmental milestone that could influence whether and how much vole pups vocalize when isolated. To assess this possibility, we used a separate cohort of pups ($n = 10$) to longitudinally track developmental markers every 2 days from P6 to P16 (**Figure 9**). We found that pups could not maintain their temperature across 10 min at P6 [$t_{(9)} = -4.50, p < 10^{-3}$, one-tailed t-test], but were able to do so from P8 onwards (all $p \geq 0.22$; **Figure 9A**). Hence, over the age range where we observed changing context-dependent USV emission, the need to thermoregulate during isolation was likely not a major factor in driving this vocal modulation.

Another documented driver of pup USV emission in some rodent species is brief maternal reunion and contact, which potentiates USV calling if the mother is subsequently removed (Hofer et al., 1998; Shair, 2014). In our paradigm, unlike standard maternal potentiation paradigms, pups could not physically contact the mother. Instead, pups could likely use auditory or visual cues to recognize her presence. We therefore checked when these sensory modalities were developmentally functional. For audition, a single pup's ears were open at P6, while all other pups' ears opened between P6 and P8 (**Figure 8B**). Ear opening effectively corresponded with the presence of functional hearing, as confirmed through an acoustic startle paradigm. Hence, the ability to hear the mother on the other side of the barrier was likely possible before P10. Meanwhile, eye opening did not occur until between P8 and P10. Thus, by P12 when the addition of the mother significantly increased the number of vocalizations emitted, vole pups would have already passed key sensory developmental milestones that could enable better sampling of changes in their social environment.

DISCUSSION

Our study demonstrates a developmental maturation in both the features of prairie vole pup vocalizations and the role that social context plays in modifying vocal emission. Specifically, we found that older prairie vole pups emit fewer vocalizations that are simpler and more stereotyped than younger pups. Additionally, and more importantly, how pups vocally respond to changes in their social context switches around P10. Whereas, young pups decrease their vocal activity when their mother is present but physically unattainable and potentially unnoticed, older pups who can sense the mother's presence increase their vocal emissions. Interestingly, this change in vocal emission is not accompanied by changes in the vocalizations' acoustic features. Together, these results define a developmental trajectory for the use of stereotyped pup USVs as a means of social communication, opening the door for leveraging prairie vole pup USVs as a developmental readout of social motivation and affiliative behavior.

To our knowledge, this work provides the first report of a sex difference in the acoustic features of prairie vole pup USVs. Vocalizations as a whole centered around 32 kHz, in line with

previous work (Colvin, 1973). However, the frequency content of male pups was downshifted from age-matched females. This finding is consistent with work in rat pups showing that females emit higher frequency vocalizations on average compared to males. Since females were consistently smaller than males, we considered whether size-frequency allometry—hypothesized (Morton, 1977) and found (Ey et al., 2007; Bowling et al., 2017) to explain differences in the fundamental frequency of harmonically structured vocalizations across primate and carnivore species—might explain our results. While pup weight significantly increased over time, our data showed no significant effect of age on the median frequency of vocalizations. In fact, low frequency significantly increased with age, opposite of what a size-frequency allometry might predict. This discrepancy likely reflects the fact that rodent USVs are probably generated by a different mechanism (Roberts, 1975; Mahrt et al., 2016; Riede et al., 2017) than the vocal fold oscillations of larger species. The presence of a size-independent sex difference should therefore be factored into models of rodent vocal emission.

Acoustic variability is reduced over vocal development in several species that learn to refine their vocalizations with experience [e.g., humans (Lee et al., 1999), birds (Ölveczky et al., 2011), and bats (Fernandez et al., 2021)]. We found that prairie voles, a presumed vocal non-learning species, exhibited a similar pattern. Across development, pup vocalizations became shorter, had narrower bandwidths, contained fewer harmonics, and had fewer segments [corresponding to previous findings (Terleph, 2011)]. Additionally, vocalizations became more stereotyped in duration, bandwidth, and the number of segments per vocalization with age. Another vocal non-learning species, the laboratory mouse [*Mus musculus* (Hammerschmidt et al., 2012; Arriaga and Jarvis, 2013; Mahrt et al., 2013)], also emits vocalizations that become more stereotyped with age – narrowing in frequency, duration, and repetition rate (Liu et al., 2003). These data from rodents thus suggest that, irrespective of the capacity for vocal learning, age-dependent reductions in the acoustic variability of vocalizations are likely a natural part of vocal development.

Why pups emit USVs has been a question of long-standing interest in rodent bioacoustics, with few studies investigating the communicative function of prairie vole pup USVs in particular. One hypothesized driver of vocal emission in other rodent species is pup temperature. Moving young pups to a colder location reliably evokes USVs (Okon, 1970, 1971, 1972; Blake, 1992), leading mothers to retrieve pups back to the warmth of the nest (Sewell, 1970; Shapiro and Insel, 1990; Brunelli et al., 1994; Ehret, 2005; Bowers et al., 2013). Kept at a thermoneutral temperature, however, relocated pups do not increase their vocal rate (Allin and Banks, 1971). The influence of external temperature on vocal activity diminishes as pups achieve thermoregulatory competence, with older pups exhibiting no vocal response to temperature changes (Okon, 1970). In our study of prairie voles, however, pups vocalized even after they reached an age when thermoregulatory competence was achieved. Thus, while undeveloped thermoregulation may explain the vocal emission of young pups, vocalizations from older pups must be differentially driven.

A second hypothesis in the field is that USV emission indicates a pup's emotional state (Noirot and Pye, 1969; Hahn et al., 1998; Ehret, 2005; Hahn and Lavooy, 2005; Portfors, 2007). Social isolation is a known stressor for prairie vole pups, leading to a hormonal stress response and a correlated increase in vocal emission (Shapiro and Insel, 1990). However, rodent pups consistently vocalize more upon experiencing a second isolation from their mother than when they are first isolated (Hofer et al., 1998; Shair, 2014; Robison et al., 2016). These maternally potentiated vocalizations cannot be a direct response to the immediate environmental conditions, since they are identical. Instead, these experiences may produce different stress responses, leading to distinct vocal patterns. Additionally, drugs that increase stress also increase USV production, while stress-reducing drugs diminish vocal emission (Hahn and Lavooy, 2005; Costantini and D'amato, 2006). Thus, instead of reflecting acute physical conditions, vocal activity may provide an accurate readout of the internal state experienced by pups.

In our study, we manipulated social context to impact prairie vole pups' internal states. We placed a pup's mother on the opposite side of a transparent barrier and characterized the pup's vocal responses. While rodent pups exhibit olfactory competence as early as P2 (Cornwell-Jones and Sobrian, 1977; Geyer, 1979), ear and eye opening occur during our window of observation. Young pups (P8) significantly decreased their vocal emissions when their mother was introduced. This may be due to habituation to the environment, as vocalization levels consistently decreased over 11 min, then stabilized. Interestingly, only after pups were competent in both hearing and seeing their surroundings did they vocalize more after adding the mother. This increased calling cannot be attributed to a desire for warmth, as pups could thermoregulate by this age. While pups at all ages were fully covered in fur, older pups also had thicker fur, likely aiding their thermoregulation. Therefore, our results may indicate that older pups' stress levels increase when the mother is nearby but not physically accessible. Alternatively, assuming that pups vocalize to elicit contact from their mother, our results may reveal the developmental emergence of using vocalizations to satisfy the drive for social contact.

In fact, social drive becomes the most reliable modulator of rodent vocal emission by adulthood. For example, adult mice, while typically silent in isolation (Whitney et al., 1973), are highly vocal in social settings. Female mice vocalize when exposed to conspecifics of either sex (D'Amato and Moles, 2001; Neunuebel et al., 2015; Hoier et al., 2016), while males are most vocal toward sexually-receptive females (Sewell, 1970; Whitney et al., 1974; Nyby et al., 1979). Enhancing social drive via social isolation leads to increased vocal emission, with pairs of isolated females vocalizing four times more than group-housed pairs (Zhao et al., 2021). Adult prairie vole vocalizations also indicate social motivation, with males and females both vocalizing toward conspecifics, and males vocalizing more toward sexually receptive females than non-receptive females (Ma et al., 2014). Thus, vocal activity by adult prairie voles is a reliable readout of social environment and social drive.

Together, our study provides a thorough assessment of the acoustic development of prairie vole pup vocalizations

and the development of social-vocal communication. We find that vocalizations become shorter and more stereotyped with age. Additionally, pups exhibit an age-dependent modification in vocal emission rates, but not acoustic features, based on social context. These results uncover a potential developmental regulation in pups' ability to modify vocal emission based on their social drive and social environment, with pups not increasing social vocalizations until after they can hear and see. Future work is necessary to explore the neural circuitry underlying social communication, including the brain regions and developmental influences involved in the control of vocal behavior.

Furthermore, given the unique behavioral repertoire of adult prairie voles and their importance in furthering our understanding of the neuromodulatory regulation of social behaviors, prairie vole pups provide a novel avenue for assessing the role of neuromodulators in early-life social bonds and the development of social communication. For instance, oxytocin can be involved in social motivation in rodent pups. Oxytocin knockout (OTKO) mouse pups (P10) take longer than wildtype pups to approach their mother following separation, and show no preference for their mother over other females [P15 (Ross and Young, 2009)]. Concomitantly, both OTKO (Winslow et al., 2000) and oxytocin-receptor knockout [OTRKO (Takayanagi et al., 2005)] mouse pups vocalize less during social isolation than wildtype pups, consistent with USVs as a readout for pups' social drive for maternal contact. Very young OTRKO prairie voles (P2–P5) have also been assessed for isolation-induced USV and found not to vocalize less (Horie et al., 2019), but this may reflect the influence of thermoregulation rather than social motivation at these ages, as our data suggest. Interestingly, rats selectively bred for lower USV emission exhibit increased oxytocin receptor expression in the nucleus accumbens (Brunelli et al., 2015). Thus, further work is necessary to determine the role of oxytocin in the development of social-vocal communication in prairie voles and the relationship between early-life oxytocinergic signaling and social-vocal communication during pair bonding in adulthood.

Finally, social-vocal communication is vital for social interaction and synchrony from birth. For instance, exposing human infants in the NICU to their mother's voice leads to enhanced wakefulness (Filippa et al., 2013) and can help ameliorate later deficits in mother-child interaction (Welch and Ludwig, 2017; Beebe et al., 2018). In other species, vocal communication enhances both behavioral and neural synchrony between individuals (Rose et al., 2021). Thus, while the possibilities of the prairie vole as a model of neurodevelopmental disorders and early life stressors are only beginning to be explored, work should begin to assess dyadic vocal interplay between pairs of animals. Our work sets the stage for future studies determining the neural and neurochemical bases of early-life social communication, and the impact of early-life manipulations on the development and use of social-vocal communication in rodent species.

DATA AVAILABILITY STATEMENT

The datasets present in this study can be found in a GitHub repository: https://github.com/rcbliu/Pup-USV-Data.

ETHICS STATEMENT

Experiments were conducted in strict accordance with the guidelines established by the National Institutes of Health and then approved by Emory University's Institutional Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

AB, CF, and RL designed the study. CF, AB, MW, and DC developed the methodology. MW, DC, and AD produced all necessary software. MW and DC conducted formal analyses. AB, CF, DC, AD, and MW conducted the research. RL and LY provided all necessary resources. MW, DC, and RL wrote the original manuscript draft, with all authors providing reviews and edits. MW, DC, and RL generated data visualizations. LY, RL, CF, and AB acquired project funding. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NIH grant P50MH100023 (LY and RL), R01MH115831 (RL and LY), 5R01DC008343 (RL), and P51OD11132 (YNPRC).

ACKNOWLEDGMENTS

We thank Lorra Julian and the Yerkes Veterinary and Animal Care staff for vole care. We also thank Danial Arslan and Jim Kwon for setting up the USV recording system. The schematic in **Figure 7A** was generated in Biorender.

REFERENCES

Albers, H. E. (2012). The regulation of social recognition, social communication and aggression: vasopressin in the social behavior neural network. *Horm. Behav.* 61, 283–292. doi: 10.1016/j.yhbeh.2011.10.007

Allin, J. T., and Banks, E. M. (1971). Effects of temperature on ultrasound production by infant albino rats. *Dev. Psychobiol.* 4, 149–156. doi: 10.1002/dev.420040206

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

Bakan, D. (1966). The test of significance in psychological research. *Psychol. Bull.* 66:423. doi: 10.1037/h0020412

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

Beebe, B., Myers, M. M., Lee, S. H., Lange, A., Ewing, J., Rubinchik, N., et al. (2018). Family nurture intervention for preterm infants facilitates positive mother–infant face-to-face engagement at 4 months. *Dev. Psychol.* 54:2016. doi: 10.1037/dev0000557

Blake, B. H. (1992). Ultrasonic vocalization and body temperature maintenance in infant voles of three species (Rodentia: Arvicolidae). *Dev. Psychobiol.* 25, 581–596. doi: 10.1002/dev.420250805

Blumberg, M. S., and Alberts, J. R. (1990). Ultrasonic vocalizations by rat pups in the cold: an acoustic by-product of laryngeal braking? *Behav. Neurosci.* 104:808. doi: 10.1037/0735-7044.104.5.808

Bosch, O. J., and Young, L. J. (2017). Oxytocin and social relationships: from attachment to bond disruption. *Behav. Pharmacol. Neuropept. Oxytocin* 35, 97–117. doi: 10.1007/7854_2017_10

Bowers, J. M., Perez-Pouchoulen, M., Edwards, N. S., and McCarthy, M. M. (2013). Foxp2 mediates sex differences in ultrasonic vocalization by rat pups and directs order of maternal retrieval. *J. Neurosci.* 33, 3276–3283. doi: 10.1523/JNEUROSCI.0425-12.2013

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K.-H., et al. (2017). Body size and vocalization in primates and carnivores. *Sci. Rep.* 7, 1–11. doi: 10.1038/srep41070

Branchi, I., Santucci, D., Vitale, A., and Alleva, E. (1998). Ultrasonic vocalizations by infant laboratory mice: a preliminary spectrographic characterization under different conditions. *Dev. Psychobiol.* 33, 249–256. doi: 10.1002/(SICI)1098-2302(199811)33:3<249::AID-DEV5>3.0.CO;2-R

Brooks, R. J., and Banks, E. M. (1973). Behavioural biology of the collared lemming [Dicrostonyx groenlandicus (Trail)]: an analysis of acoustic communication. *Anim. Behav. Monogr.* 6, 1–83. doi: 10.1016/0003-3472(73)90003-1

Brunelli, S. A., Curley, J. P., Gudsnuk, K., Champagne, F. A., Myers, M. M., Hofer, M. A., et al. (2015). Variations in maternal behavior in rats selected for infant ultrasonic vocalization in isolation. *Horm. Behav.* 75, 78–83. doi: 10.1016/j.yhbeh.2015.08.007

Brunelli, S. A., Shair, H. N., and Hofer, M. A. (1994). Hypothermic vocalizations of rat pups (Rattus norvegicus) and direct maternal search behavior. *J. Comp. Psychol.* 108:298. doi: 10.1037/0735-7036.108.3.298

Campbell, P., Pasch, B., Warren, A. L., and Phelps, S. M. (2014). Vocal ontogeny in neotropical singing mice (Scotinomys). *PLoS ONE* 9:e113628. doi: 10.1371/journal.pone.0113628

Campi, K. L., Karlen, S. J., Bales, K. L., and Krubitzer, L. (2007). Organization of sensory neocortex in prairie voles (*Microtus ochrogaster*). *J. Compar. Neurol.* 502, 414–426. doi: 10.1002/cne.21314

Chabout, J., Sarkar, A., Dunson, D. B., and Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Front. Behav. Neurosci.* 9:76. doi: 10.3389/fnbeh.2015.00076

Chabout, J., Serreau, P., Ey, E., Bellier, L., Aubin, T., Bourgeron, T., et al. (2012). Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS ONE* 7:e29401. doi: 10.1371/journal.pone.0029401

Colvin, M. A. (1973). Analysis of acoustic structure and function in ultrasounds of neonatal Microtus. *Behaviour* 44, 234–263. doi: 10.1163/156853973X00418

Cornwell-Jones, C., and Sobrian, S. K. (1977). Development of odor-guided behavior in Wistar and Spraque-Dawley rat pups. *Physiol. Behav.* 19, 685–688. doi: 10.1016/0031-9384(77)90044-0

Costantini, F., and D'amato, F. (2006). Ultrasonic vocalizations in mice and rats: social contexts and functions. *Dong Wu Xue Bao.* 52, 619–633.

D'Amato, F. R., and Moles, A. (2001). Ultrasonic vocalizations as an index of social memory in female mice. *Behav. Neurosci.* 115:834. doi: 10.1037/0735-7044.115.4.834

D'Amato, F. R., Scalera, E., Sarli, C., and Moles, A. (2005). Pups call, mothers rush: does maternal responsiveness affect the amount of ultrasonic vocalizations in mouse pups? *Behav. Genet.* 35, 103–112. doi: 10.1007/s10519-004-0860-9

Ehret, G. (2005). Infant rodent ultrasounds–a gate to the understanding of sound communication. *Behav. Genet.* 35, 19–29. doi: 10.1007/s10519-004-0853-8

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

Fernandez, A. A., Burchardt, L. S., Nagy, M., and Knörnschild, M. (2021). Babbling in a vocal learning bat resembles human infant babbling. *Science* 373, 923–926. doi: 10.1126/science.abf9279

Filippa, M., Devouche, E., Arioni, C., Imberty, M., and Gratier, M. (2013). Live maternal speech and singing have beneficial effects on hospitalized preterm infants. *Acta Paediatr.* 102, 1017–1020. doi: 10.1111/apa.12356

Froemke, R. C., and Young, L. J. (2021). Oxytocin, neural plasticity, and social behavior. *Annu. Rev. Neurosci.* 44, 359–381. doi: 10.1146/annurev-neuro-102320-102847

Geyer, L. A. (1979). Olfactory and thermal influences on ultrasonic vocalization during development in rodents. *Am. Zool.* 19, 419–431. doi: 10.1093/icb/19.2.419

Hahn, M. E., Karkowski, L., Weinreb, L., Henry, A., Schanz, N., and Hahn, E. M. (1998). Genetic and developmental influences on infant mouse ultrasonic calling. II. Developmental patterns in the calls of mice 2–12 days of age. *Behav. Genet.* 28, 315–325. doi: 10.1023/A:1021679615792

Hahn, M. E., and Lavooy, M. J. (2005). A review of the methods of studies on infant ultrasound production and maternal retrieval in small rodents. *Behav. Genet.* 35, 31–52. doi: 10.1007/s10519-004-0854-7

Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenzke, N., and Fischer, J. (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci.* 13:40. doi: 10.1186/1471-2202-13-40

Hernandez-Miranda, L. R., Ruffault, P.-L., Bouvier, J. C., Murray, A. J., Morin-Surun, M.-P., Zampieri, N., et al. (2017). Genetic identification of a hindbrain nucleus essential for innate vocalization. *Proc. Nat. Acad. Sci.* 114, 8095–8100. doi: 10.1073/pnas.1702893114

Hofer, M., Masmela, J., Brunelli, S., and Shair, H. (1998). The ontogeny of maternal potentiation of the infant rats' isolation call. *Dev. Psychobiol.* 33, 189–201. doi: 10.1002/(SICI)1098-2302(199811)33:3<189::AID-DEV1>3.0.CO;2-N

Hofer, M. A., and Shair, H. N. (1992). Ultrasonic vocalization by rat pups during recovery from deep hypothermia. *Dev. Psychobiol.* 25, 511–528. doi: 10.1002/dev.420250705

Hofer, M. A., and Shair, H. N. (1993). Ultrasonic vocalization, laryngeal braking, and thermogenesis in rat pups: a reappraisal. *Behav. Neurosci.* 107:354. doi: 10.1037/0735-7044.107.2.354

Hoier, S., Pfeifle, C., von Merten, S., and Linnenbrink, M. (2016). Communication at the garden fence–context dependent vocalization in female house mice. *PLoS ONE* 11:e0152255. doi: 10.1371/journal.pone.0152255

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

Horie, K., Inoue, K., Suzuki, S., Adachi, S., Yada, S., Hirayama, T., et al. (2019). Oxytocin receptor knockout prairie voles generated by CRISPR/Cas9 editing show reduced preference for social novelty and exaggerated repetitive behaviors. *Horm. Behav.* 111, 60–69. doi: 10.1016/j.yhbeh.2018.10.011

Johnson, S. A., Painter, M. S., Javurek, A. B., Murphy, C. R., Howald, E. C., Khan, Z. Z., et al. (2017). Characterization of vocalizations emitted in isolation by California mouse (*Peromyscus californicus*) pups throughout the postnatal period. *J. Comp. Psychol.* 131:30. doi: 10.1037/com0000057

Kaidbey, J. H., Ranger, M., Myers, M. M., Anwar, M., Ludwig, R. J., Schulz, A. M., et al. (2019). Early life maternal separation and maternal behaviour modulate acoustic characteristics of rat pup ultrasonic vocalizations. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-54800-z

Kessler, M. S., Bosch, O. J., Bunck, M., Landgraf, R., and Neumann, I. D. (2011). Maternal care differs in mice bred for high vs. low trait anxiety: impact of brain vasopressin and cross-fostering. *Soc. Neurosci.* 6, 156–168. doi: 10.1080/17470919.2010.495567

Krubitzer, L., Campi, K. L., and Cooke, D. F. (2011). All rodents are not the same: a modern synthesis of cortical organization. *Brain Behav. Evol.* 78, 51–93. doi: 10.1159/000327320

Lee, S., Potamianos, A., and Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* 105, 1455–1468. doi: 10.1121/1.426686

Lepri, J. J., Theodorides, M., and Wysocki, C. J. (1988). Ultrasonic vocalizations by adult prairie voles, *Microtus ochrogaster*. *Experientia* 44, 271–273. doi: 10.1007/BF01941736

Liu, R. C., Miller, K. D., Merzenich, M. M., and Schreiner, C. E. (2003). Acoustic variability and distinguishability among mouse ultrasound vocalizations. *J. Acoust. Soc. Am.* 114, 3412–3422. doi: 10.1121/1.1623787

Ma, S. T., Resendez, S. L., and Aragona, B. J. (2014). Sex differences in the influence of social context, salient social stimulation and amphetamine

on ultrasonic vocalizations in prairie voles. *Integr. Zool.* 9, 280–293. doi: 10.1111/1749-4877.12071

Mahrt, E., Agarwal, A., Perkel, D., Portfors, C., and Elemans, C. P. (2016). Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. *Curr. Biol.* 26, R880–R881. doi: 10.1016/j.cub.2016.08.032

Mahrt, E. J., Perkel, D. J., Tong, L., Rubel, E. W., and Portfors, C. V. (2013). Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J. Neurosci.* 33, 5573–5583. doi: 10.1523/JNEUROSCI.5054-12.2013

Marlin, B. J., and Froemke, R. C. (2017). Oxytocin modulation of neural circuits for social behavior. *Dev. Neurobiol.* 77, 169–189. doi: 10.1002/dneu.22452

McGraw, L. A., and Young, L. J. (2010). The prairie vole: an emerging model organism for understanding the social brain. *Trends Neurosci.* 33, 103–109. doi: 10.1016/j.tins.2009.11.006

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855–869. doi: 10.1086/283219

Nagasawa, M., and Kikusui, T. (2021). Neuroendocrine mechanisms of social bonds and separation stress in rodents, dogs, and other species. *Curr Top Behav Neurosci*. doi: 10.1007/7854_2021_257 [Epub ahead of print].

Nair, H. P., and Young, L. J. (2006). Vasopressin and pair-bond formation: genes to brain to behavior. *Physiology* 21, 146–152. doi: 10.1152/physiol.00049.2005

Naito, H., and Tonoue, T. (1987). Sex difference in ultrasound distress call by rat pups. *Behav. Brain Res.* 25, 13–21. doi: 10.1016/0166-4328(87)90041-6

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., and Egnor, S. R. (2015). Female mice ultrasonically interact with males during courtship displays. *Elife* 4:e06203. doi: 10.7554/eLife.06203.013

Noirot, E., and Pye, D. (1969). Sound analysis of ultrasonic distress calls of mouse pups as a function of their age. *Anim. Behav.* 17, 340–349. doi: 10.1016/0003-3472(69)90020-7

Nyby, J., Wysocki, C. J., Whitney, G., Dizinno, G., and Schneider, J. (1979). Elicitation of male mouse (Mus musculus) ultrasonic vocalizations: I. Urinary cues. *J. Compar. Physiol. Psychol.* 93:957. doi: 10.1037/h0077623

Oettl, L.-L., Ravi, N., Schneider, M., Scheller, M. F., Schneider, P., Mitre, M., et al. (2016). Oxytocin enhances social recognition by modulating cortical control of early olfactory processing. *Neuron* 90, 609–621. doi: 10.1016/j.neuron.2016.03.033

Okon, E. E. (1970). The effect of environmental temperature on the production of ultrasounds by isolated non-handled albino mouse pups. *J. Zool.* 162, 71–83. doi: 10.1111/j.1469-7998.1970.tb01258.x

Okon, E. E. (1971). The temperature relations of vocalization in infant golden hamsters and Wistar rats. *J. Zool.* 164, 227–237. doi: 10.1111/j.1469-7998.1971.tb01308.x

Okon, E. E. (1972). Factors affecting ultrasound production in infant rodents. *J. Zool.* 168, 139–148. doi: 10.1111/j.1469-7998.1972.tb01344.x

Ölveczky, B. P., Otchy, T. M., Goldberg, J. H., Aronov, D., and Fee, M.S. (2011). Changes in the neural control of a complex motor sequence during learning. *J. Neurophysiol.* 106, 386–397. doi: 10.1152/jn.00018.2011

Portfors, C. V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J. Am. Assoc. Lab. Anim. Sci.* 46, 28–34.

Rabon Jr, D. R., Sawrey, D. K., and Webster, W. D. (2001). Infant ultrasonic vocalizations and parental responses in two species of voles (Microtus). *Can. J. Zool.* 79, 830–837. doi: 10.1139/z01-043

Riede, T., Borgard, H. L., and Pasch, B. (2017). Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. *R. Soc. Open Sci.* 4:170976. doi: 10.1098/rsos.170976

Rieger, N. S., and Marler, C. A. (2018). The function of ultrasonic vocalizations during territorial defence by pair-bonded male and female California mice. *Anim. Behav.* 135, 97–108. doi: 10.1016/j.anbehav.2017.11.008

Roberts, L. (1975). The rodent ultrasound production mechanism. *Ultrasonics* 13, 83–88. doi: 10.1016/0041-624X(75)90052-9

Robison, W. T., Myers, M. M., Hofer, M. A., Shair, H. N., and Welch, M. G. (2016). Prairie vole pups show potentiated isolation-induced vocalizations following isolation from their mother, but not their father. *Dev. Psychobiol.* 58, 687–699. doi: 10.1002/dev.21408

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

Ross, H. E., and Young, L. J. (2009). Oxytocin and the neural mechanisms regulating social cognition and affiliative behavior. *Front. Neuroendocrinol.* 30, 534–547. doi: 10.1016/j.yfrne.2009.05.004

Sadino, J. M., and Donaldson, Z. R. (2018). Prairie voles as a model for understanding the genetic and epigenetic regulation of attachment behaviors. *ACS Chem. Neurosci.* 9, 1939–1950. doi: 10.1021/acschemneuro.7b00475

Schiavo, J. K., Valtcheva, S., Bair-Marshall, C. J., Song, S. C., Martin, K. A., and Froemke, R. C. (2020). Innate and plastic mechanisms for maternal behaviour in auditory cortex. *Nature* 587, 426–431. doi: 10.1038/s41586-020-2807-6

Sewell, G. D. (1970). Ultrasonic communication in rodents. *Nature* 227, 410–410. doi: 10.1038/227410a0

Shair, H. N. (2014). Parental potentiation of vocalization as a marker for filial bonds in infant animals. *Dev. Psychobiol.* 56, 1689–1697. doi: 10.1002/dev.21222

Shapiro, L. E., and Insel, T. R. (1990). Infant's response to social separation reflects adult differences in affiliative behavior: a comparative developmetal study in prairie and montane voles. *Dev. Psychobiol.* 23, 375–393. doi: 10.1002/dev.420230502

Tachibana, R. O., Kanno, K., Okabe, S., Kobayasi, K. I., and Okanoya, K. (2020). USVSEG: a robust method for segmentation of ultrasonic vocalizations in rodents. *PLoS ONE* 15:e0228907. doi: 10.1371/journal.pone.0228907

Takayanagi, Y., Yoshida, M., Bielsky, I. F., Ross, H. E., Kawamata, M., Onaka, T., et al. (2005). Pervasive social deficits, but normal parturition, in oxytocin receptor-deficient mice. *Proc. Nat. Acad. Sci.* 102, 16096–16101. doi: 10.1073/pnas.0505312102

Terleph, T. A. (2011). A comparison of prairie vole audible and ultrasonic pup calls and attraction to them by adults of each sex. *Behaviour* 148, 1275–1294. doi: 10.1163/000579511X600727

Walum, H., and Young, L. J. (2018). The neural mechanisms and circuitry of the pair bond. *Nat. Rev. Neurosci.* 19, 643–654. doi: 10.1038/s41583-018-0072-6

Warren, M. R., Spurrier, M. S., Sangiamo, D. T., Clein, R. S., and Neunuebel, J. P. (2021). Mouse vocal emission and acoustic complexity do not scale linearly with the size of a social group. *J. Exp. Biol.* 224:jeb239814. doi: 10.1242/jeb.243045

Welch, M. G., and Ludwig, R. J. (2017). "Mother/infant emotional communication through the lens of visceral/autonomic learning," in *Early Vocal Contact and Preterm Infant Brain Development,* eds M. Filippa, P. Kuhn, and B. Westrup (Cham: Springer), 271–294.

Whitney, G., Alpern, M., Dizinno, G., and Horowitz, G. (1974). Female odors evoke ultrasounds from male mice. *Anim. Learn. Behav.* 2, 13–18. doi: 10.3758/BF03199109

Whitney, G., Coble, J. R., Stockton, M. D., and Tilson, E. F. (1973). Ultrasonic emissions: do they facilitate courtship of mice? *J. Comp. Physiol. Psychol.* 84:445. doi: 10.1037/h0034899

Whitney, G., and Nyby, J. (1979). Cues that elicit ultrasounds from adult male mice. *Am. Zool.* 19, 457–463. doi: 10.1093/icb/19.2.457

Winslow, J. T., Hearn, E. F., Ferguson, J., Young, L. J., Matzuk, M. M., and Insel, T. R. (2000). Infant vocalization, adult aggression, and fear behavior of an oxytocin null mutant mouse. *Horm. Behav.* 37, 145–155. doi: 10.1006/hbeh.1999.1566

Young, K. A., Liu, Y., and Wang, Z. (2008). The neurobiology of social attachment: a comparative approach to behavioral, neuroanatomical, and neurochemical studies. *Comp. Biochem. Physiol. C. Toxicol. Pharmacol.* 148, 401–401. doi: 10.1016/j.cbpc.2008.02.004

Yurlova, D. D., Volodin, I. A., Ilchenko, O. G., and Volodina, E. V. (2020). Rapid development of mature vocal patterns of ultrasonic calls in a fast-growing rodent, the yellow steppe lemming (Eolagurus luteus). *PLoS ONE* 15:e0228892. doi: 10.1371/journal.pone.0228892

Zhao, X., Ziobro, P., Pranic, N. M., Chu, S., Rabinovich, S., Chan, W., et al. (2021). Sex- and context-dependent effects of acute isolation on vocal and non-vocal social behaviors in mice. *PLoS ONE* 16:e0255640. doi: 10.1371/journal.pone.0255640

Zippelius, H.-M., and Schleidt, W. M. (1956). Ultraschall-laute bei jungen mäusen. *Naturwissenschaften* 43, 502–502. doi: 10.1007/BF00632534

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fnbeh.2021.812142

Response Calls Evoked by Playback of Natural 50-kHz Ultrasonic Vocalizations in Rats

Annuska C. Berz[1,2†], Markus Wöhr[1,2,3,4†‡] and Rainer K. W. Schwarting[1,2†‡]*

[1] Behavioral Neuroscience, Experimental and Biological Psychology, Faculty of Psychology, Philipps-University Marburg, Marburg, Germany, [2] Center for Mind, Brain and Behavior, Philipps-University Marburg, Marburg, Germany, [3] Research Unit Brain and Cognition, Laboratory of Biological Psychology, Social and Affective Neuroscience Research Group, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium, [4] Leuven Brain Institute, KU Leuven, Leuven, Belgium

OPEN ACCESS

Edited by:
Susanna Pietropaolo,
UMR 5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Stefan Brudzynski,
Brock University, Canada
Sergio Marcello Pellis,
University of Lethbridge, Canada

***Correspondence:**
Annuska C. Berz
Annuska.berz@staff.uni-marburg.de

†ORCID:
Annuska C. Berz
orcid.org/0000-0002-5522-0188
Markus Wöhr
orcid.org/0000-0001-6986-5684
Rainer K. W. Schwarting
orcid.org/0000-0002-4686-3974

‡These authors share last authorship

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: 09 November 2021
Accepted: 24 December 2021
Published: 14 January 2022

Citation:
Berz AC, Wöhr M and Schwarting RKW (2022) Response Calls Evoked by Playback of Natural 50-kHz Ultrasonic Vocalizations in Rats.
Front. Behav. Neurosci. 15:812142.
doi: 10.3389/fnbeh.2021.812142

Rats are highly social animals known to communicate with ultrasonic vocalizations (USV) of different frequencies. Calls around 50 kHz are thought to represent a positive affective state, whereas calls around 22 kHz are believed to serve as alarm or distress calls. During playback of natural 50-kHz USV, rats show a reliable and strong social approach response toward the sound source. While this response has been studied in great detail in numerous publications, little is known about the emission of USV in response to natural 50-kHz USV playback. To close this gap, we capitalized on three data sets previously obtained and analyzed USV evoked by natural 50-kHz USV playback in male juvenile rats. We compared different rat stocks, namely Wistar (WI) and Sprague-Dawley (SD) and investigated the pharmacological treatment with the dopaminergic D2 receptor antagonist haloperidol. These response calls were found to vary broadly inter-individually in numbers, mean peak frequencies, durations and frequency modulations. Despite the large variability, the results showed no major differences between experimental conditions regarding call likelihood or call parameters, representing a robust phenomenon. However, most response calls had clearly lower frequencies and were longer than typical 50-kHz calls, i.e., around 30 kHz and lasting generally around 0.3 s. These calls resemble aversive 22-kHz USV of adult rats but were of higher frequencies and shorter durations. Moreover, blockade of dopamine D2 receptors did not substantially affect the emission of response calls suggesting that they are not dependent on the D2 receptor function. Taken together, this study provides a detailed analysis of response calls toward playback of 50-kHz USV in juvenile WI and SD rats. This includes calls representing 50-kHz USV, but mostly calls with lower frequencies that are not clearly categorizable within the so far known two main groups of USV in adult rats. We discuss the possible functions of these response calls addressing their communicative functions like contact or appeasing calls, and whether they may reflect a state of frustration. In future studies, response calls might also serve as a new read-out in rat models for neuropsychiatric disorders, where acoustic communication is impaired, such as autism spectrum disorder.

Keywords: ultrasonic vocalizations, animal communication, playback, stock, strain, haloperidol, Wistar, Sprague-Dawley

INTRODUCTION

Acoustic communication among conspecifics is an important aspect of the social life of many species and often essential for maintaining stable social structures. A characteristic feature of acoustic communication in several species is its reciprocal nature where a signal emitted by the sender frequently evokes the emission of a response signal in the receiver (Seyfarth and Cheney, 2003).

Many rodent species communicate through so-called ultrasonic vocalizations (USV), i.e., within frequencies not audible for humans (Brudzynski, 2010). In juvenile and adult rats, two main types of vocalizations are typically distinguished (Brudzynski, 2013a; Wöhr and Schwarting, 2013). Vocalizations with frequencies around 22 kHz are referred to as aversive or distress calls, presumably representing a negative affective state (Blanchard et al., 1991; Fendt et al., 2018). Vocalizations with frequencies around 50 kHz are thought to represent a positive affective state usually emitted during appetitive situations like play or mating (Knutson et al., 1998; Panksepp, 2005). These appetitive calls are typically characterized by frequencies between 35 and 80 kHz and durations in a range of 10–150 ms (Burgdorf et al., 2008; Wöhr et al., 2008; Takahashi et al., 2010). Often, such 50-kHz USV are categorized and the call categories flat, step, trill, and mixed are commonly differentiated (Kisko et al., 2018). Aversive 22-kHz USV, in contrast, have been defined between frequencies of 18 and 32 kHz (Brudzynski, 2001) and within this frequency range, short (<300 ms) and long (>300 ms) calls were identified (Brudzynski et al., 1993). Long 22-kHz calls were found to be emitted during situations of external danger, such as during the presence of a predator or during predator odor exposure, and are usually associated with freezing behavior (Blanchard et al., 1991; Fendt et al., 2018; Simmons et al., 2018). Short 22-kHz USV, however, are much more ambiguous and their function has not been identified yet (Brudzynski, 2021). It was suggested that short 22-kHz USV represent internal distress without external influence, like frustration (Taylor et al., 2019). In addition, they were repeatedly reported to occur during drug withdrawal (Ma et al., 2010; Simmons et al., 2018).

The communicative functions of 22- and 50-kHz USV can be studied by means of playback experiments (Seffer et al., 2014) and it was shown that they elicit distinct behavioral responses pattern in the receiver (Wöhr et al., 2016). Playback of natural 22-kHz USV usually induces a defensive response, including avoidance behavior and behavioral inhibition (Brudzynski and Chiu, 1995; Fendt et al., 2018). Playback of natural 50-kHz USV, in contrast, evokes social approach behavior toward the sound source (Wöhr and Schwarting, 2007). At the physiological level, playback of 22- and 50-kHz USV entail to distinct alterations. While playback of 22-kHz leads to a decrease in heart rate during behavioral inhibition, heart rate is increased during social approach behavior in response to playback of 50-kHz USV (Olszyński et al., 2020). Likewise, distinct brain activation patterns are observed. Playback of 22-kHz USV induces increased activity in the amygdala (Sadananda et al., 2008; Parsana et al., 2012), whereas playback of 50-kHz USV results in an activation of the nucleus accumbens (Sadananda et al., 2008), where it causes a phasic release of dopamine (Willuhn et al., 2014).

At the behavioral level, the social approach response toward 50-kHz USV playback can be accompanied by the emission of response calls (Wöhr and Schwarting, 2007, 2009; Willadsen et al., 2014; Willuhn et al., 2014; Engelhardt et al., 2017, 2018; Berg et al., 2018, 2021; Kisko et al., 2020; Olszyński et al., 2020, 2021). Although echoing the reciprocal nature of acoustic communication and repeatedly observed in studies applying the 50-kHz USV playback paradigm, still little is known about such response calls. In previous studies, response calls toward 50-kHz USV were observed in males and females (Berg et al., 2018, 2021), albeit the emission of calls in response to 50-kHz USV playback was found to be more prominent in males than females in one study (Kisko et al., 2020). A developmental study further suggests that age is another relevant factor, with juvenile rats emitting more response calls than adult rats (Wöhr and Schwarting, 2009). Finally, prior experiences (Olszyński et al., 2021) and inter-individual differences (Engelhardt et al., 2018) were also reported to play a role. However, the function of response calls remains elusive, which is why we wanted to shed light onto the meaning and the importance of response calls in social situations like the 50-kHz USV playback.

To close this gap, we capitalized on a previously obtained large data set and analyzed USV evoked by natural 50-kHz USV playback in male juvenile rats (Berz et al., 2021). In our previous study, we showed, amongst others, that the social approach response toward 50-kHz calls is a stable phenomenon that occurs in Wistar (WI) and Sprague-Dawley (SD) rats and that it can be modulated by administration of the dopaminergic D2 receptor antagonist haloperidol (Halo; Berz et al., 2021). Here, we present three new data sets from these previous experiments. Data set 1 was comprised of WI rats exposed to 50-kHz USV playback. We analyzed it in an initial attempt to better understand the emission of response calls and to test whether response calls occur specifically in reaction toward 50-kHz USV but not noise and whether stimulus order of 50-kHz USV and noise plays a role. Data set 2 consisted of WI and SD rats and their response calls were compared to see whether there was a difference between the stocks. In the final data set 3, rats received either Halo or saline (Sal) to investigate whether Halo treatment not only affects social approach behavior but also the emission of response calls toward 50-kHz USV playback. Our comprehensive analysis approach included a detailed investigation of the temporal emission pattern and an examination of acoustic features, focusing on numbers of calls, latencies to start calling, mean peak frequencies, call durations, and frequency modulations.

MATERIALS AND METHODS

Animals and Housing

In total, 108 experimentally naïve juvenile male rats around 5–7 weeks of age (Charles River Laboratories, Sulzfeld, Germany) were analyzed. The sample consisted of 90 Wistar (WI) wildtype rats and 18 Sprague-Dawley (SD) wildtype rats. The animals were kept in a vivarium with a 12-hour light/dark cycle with lights on

at 7 am and 32–50% humidity. They were housed in groups of five to six rats in polycarbonate cages (macrolon type IV, size 380 × 200 × 590 mm with high steel covers) where food and water were provided *ad libitum*. After arrival from the breeder, the animals had seven days to acclimate to the vivarium, followed by a standardized protocol of handling for three consecutive days, each day for 5 min. The procedures had been approved by the ethical committee of the local government (Regierungspräsidium Gießen, Germany, TVA Nr. 6 35-2018).

Overview

Response calls were analyzed in three data sets. These sets were obtained as part of a recently published study focusing on the habituation of the social approach response to repeated playback of 50-kHz USV (Berz et al., 2021). In this previous study, rats were exposed twice to playback of 50-kHz USV and their behavioral response was quantified, i.e., locomotor activity and approach behavior. Here, we now analyzed response calls evoked by playback of 50-kHz USV that were also recorded in this study. We focused on the emission of response calls during the first playback exposure because preliminary data indicate that call emission decreases with repeated playback presentations similar to social approach behavior (Berz et al., 2021). In the first data set, we analyzed response calls in WI rats (N = 24) and tested whether their emission occurs specifically during playback of 50-kHz USV but not noise and whether their emission depends on stimulus order. Rats were weighing 144.25 ± 1.88 g (range 128.5–164.5 g). In the second data set, we compared the production of response calls between WI rats (N = 18) to that of SD rats (N = 18). Rats were weighing 163.47 ± 2.85 g (range 138.5–205 g). In the third data set, we studied the role of the dopaminergic system in regulating the emission of response calls and compared response calls emitted by WI rats systemically treated with the dopaminergic D2 receptor antagonist Halo (N = 24) and saline treated controls (N = 24). Rats were weighing 189.57 ± 2.95 g (range 147.5–233 g).

Drug Treatment

In the third data set, rats received the dopaminergic D2 receptor antagonist Halo (0.5 mg/kg; Haldol, Janssen, Belgium) or saline (Sal, 0.9% NaCl solution, Braun, Germany). The ip injection took place 60 min before the start of the playback experiment and during the time between the injection and the playback experiment, rats were kept singly (in a small cage with bedding and water *ad libitum*) in a dark room (according to Tonelli et al., 2017).

50-kHz Ultrasonic Vocalizations Playback: Setup

As experimental setups, an eight-arm radial maze (data sets 1 and 2) and a squared platform (data set 3), each elevated 52 cm above the ground, were employed. On two opposite sides of the given apparatus, an ultrasonic speaker (ScanSpeak, Avisoft Bioacoustics, Berlin, Germany) and an ultrasonic condenser microphone (CM16, Avisoft Bioacoustics) were placed 20 cm away from the end of the arm or platform. Only one of the speakers was active, whereas the other one served as a visual control. Experiments were conducted under red light (∼10 lux).

50-kHz Ultrasonic Vocalizations Playback: Acoustic Stimuli

We presented two types of acoustic stimuli: (A) 50-kHz USV recorded from an adult male WI rat (ca. 350 g) during exploration of a cage containing scents from a recently removed cage mate (for details see Wöhr et al., 2008). This recording was composed of a sequence of 3.5 s with 13 different 50-kHz calls (total calling time 0.9 s) presented in a loop (for details see Wöhr and Schwarting, 2007). The peak amplitude was 70 dB (measured from a distance of 40 cm), being in the typical range of 50-kHz USV (Kisko et al., 2020). (B) Time- and amplitude-matched noise was generated with SASLab Pro (Version 4.2, Avisoft Bioacoustics) by replacing each 50-kHz call by noise with matching duration and amplitude modulation. Accordingly, each noise playback series had the same temporal pattern and all call features were identical, except that the sound energy was not in a certain frequency range as in the natural 50-kHz USV playback (for details see Wöhr and Schwarting, 2012). The acoustic stimuli were presented *via* an ultrasonic speaker (ScanSpeak, Avisoft Biosacoustics) with a frequency range of 1–120 kHz and a flat frequency response (±12 dB) between 15 and 80 kHz. Sounds were played *via* a portable ultrasonic power amplifier with a frequency range of 1–125 kHz (Avisoft Bioacoustics) and *via* an external sound card with a sampling rate of 192 kHz (Fire Wire Audio Capture FA-101, Edirol, London, United Kingdom).

50-kHz Ultrasonic Vocalizations Playback: Paradigm

At the beginning of the playback experiment, rats were placed individually in the center of the eight-arm radial maze (data sets 1 and 2) or the squared platform (data set 3). After an initial habituation period of 15 min, the first playback presentation of 5 min duration commenced. The second playback presentation of 5 min duration followed after an inter-stimulus interval of 10 min. Acoustic stimuli (i.e., 50-kHz USV, noise) were presented in a counterbalanced manner. The trial ended with a post-stimulus interval of 10 min. The whole paradigm lasted 45 min.

Recording and Analysis of Response Calls

For recording response calls emitted by the given experimental rat, two ultrasonic microphones were placed symmetrically on two sides of the maze (data sets 1 and 2) or the platform (data set 3) next to the speakers. They were connected *via* an UltraSoundGate 416H USB audio device (Avisoft Bioacoustics) to a computer, where acoustic data were recorded with a sampling rate of 250 kHz (16-bit format; recording range 0–125 kHz) using RECORDER USGH (Avisoft Bioacoustics). For acoustical analysis, recordings were transferred to DeepSqueak (version 2.6.1, Windows standalone), a deep learning-based system for detection and analysis of USV (Coffey et al., 2019). Recorded files were converted into high-resolution spectrograms and were analyzed using the pre-trained automated "short rat call network

V2." The settings for call detection were "high recall," with an overlap of 0.001 s. This setting was chosen because it minimizes the possibility that a call is missed, albeit at the cost of false positives by including noise. Therefore, a custom trained network for denoising was applied afterward. The detected events were then transferred into the DeepSqueak Screener (Fork on GitHub by L. Lara-Valderrábano and R. Ciszek: 10.5281/zenodo.3690137),[1] where the files were reviewed and denoised again manually by an experienced observer accepting (response calls) or rejecting (noise or playback calls) events. All response calls, irrespective of frequencies and durations, were counted. For later analysis, response calls during the 5 min before, during, and after the playback presentations (50-kHz USV or noise) were taken into account (referred to as stimulus phase). Outside this time window, calls occurred rarely. Acoustic features, i.e., call duration, peak frequency, and frequency modulation (difference between highest and lowest frequency), were defined and analyzed as described previously (Kisko et al., 2018). For classifying response calls, we applied previously established frequency thresholds (Brudzynski, 2001). Calls with frequencies higher than 32 kHz were classified as 50-kHz USV and calls below 32 kHz were defined as 22-kHz USV.

Recording and Analysis of Overt Behavior

As pointed out above, the behavioral data (locomotion, approach) were part of a recently published study focusing on the habituation of the social approach response to playback of 50-kHz USV (Berz et al., 2021). Here, we reconsidered these data in the context of the new data on response calls in order to address the question whether locomotor activity and approach behavior evoked by playback of 50-kHz USV are associated with the emission of response calls. Briefly, overt behavior was recorded and analyzed using EthoVision XT (Version 13, Noldus, The Netherlands). Locomotion was measured by the distance traveled. For quantifying approach behavior on the maze (data sets 1 and 2), the numbers of entries into the three arms proximal and distal to the active speaker and the time spent thereon were measured. For quantifying approach behavior on the platform (data set 3), it was virtually divided into 25 equal quadrants, with the six quadrants close to the active speaker serving as proximal zone, while the six quadrants close to the inactive speaker were defined as distal zones. Entries and time spent in these zones were measured (for details see Berz et al., 2021).

Statistical Analysis

Analyses of variance (ANOVAs) for repeated measurements were calculated with the between-subject factors playback order (50-kHz USV first vs. second), stocks (WI vs. SD), or drug treatment (Halo vs. Sal), and the within-subject factors stimulus phase (5 min before, during, or after playback) and playback stimulus (50-kHz USV or time- and amplitude-matched noise). This was followed by two-tailed *t*-tests for comparing individual experimental groups. The ratio between calling and non-calling rats was evaluated by a χ^2-test (calculated using https://www.socscistatistics.com/tests/chisquare2/default2.aspx). Approach behavior was quantified by subtracting the times spent on proximal arms (or in proximal zones) before the 5 min of 50-kHz USV playback from the time spent there during the 5 min of playback. The same was done with the entries into proximal arms or zones. Pearson correlation coefficients (bivariate) were calculated for the correlation between numbers of emitted calls and approach behavior. For testing a possible correlation with locomotor behavior, locomotion (distance traveled in cm) during the 5 min before playback were subtracted from that during the 5 min during playback. This number was then correlated with the numbers of response calls emitted using the Pearson correlation coefficient. For general locomotor activity correlations, the distance traveled during the initial 15-min habituation period were taken into account. All *t*-tests, ANOVAs, and correlations were calculated with IBM SPSS Statistics (version 25). Graphs were made using GraphPad Prism (version 8). Data are represented as means ± SEM (standard error of mean). A p-value of < 0.050 was considered statistically significant.

RESULTS

Data Set 1: Response Calls

Call Numbers and Latencies

Playback of 50-kHz USV induced response calls in the majority of WI rats. Among the 24 rats of data set 1, 23 of them emitted response calls. The mean number of response calls was 123.5 ± 26.21, ranging between 0 and 414 calls in total per rat (**Figure 1A**). During the 5 min before 50-kHz USV playback, no calls were emitted. The occurrence of response calls was not dependent on whether 50-kHz USV were presented as the first or the second stimulus ($t_{22} = 0.82$, $p = 0.21$). Importantly, high levels of response calls were emitted specifically in reaction toward playback of 50-kHz USV but not noise, irrespective of whether 50-kHz USV were presented as the first ($t_{11} = 2.8$, $p = 0.017$) or the second stimulus ($t_{11} = 4.013$, $p = 0.002$; **Figure 1A**). The latency to start calling after onset of 50-kHz USV was 20.17 ± 88.17 s (**Figure 1B**). Stimulus order did not affect call latency ($t_{21} = 0.52$, $p = 0.61$). Therefore, we abstained from differentially considering stimulus order further in all following analyses.

Data Set 2: Stock Differences

Call Numbers and Latencies

Consistent with data set 1, response calls were seen in the majority of rats in data set 2 focusing on possible stock differences between WI and SD rats. From the two different stocks, 10 out of the 18 WI rats emitted calls in response to 50-kHz USV playback and 12 out of 18 SD rats did. The ratios between calling and non-calling rats did not differ between stocks ($x^2_{1,\,36} = 0.468$, $p = 0.49$). Likewise, the mean numbers of response calls (**Figure 2A**; $t_{34} = 0.032$, $p = 0.975$; WI: 44.39 ± 17.81; SD: 45.17 ± 16.45) as well as the latencies to start calling (**Figure 2B**; $t_{20} = 0.547$, $p = 0.590$; WI: 50.56 ± 41.16 s; SD: 29.88 ± 4.93 s) did not differ between WI and SD. In both stocks, high levels of response calls were exclusively evoked by playback of 50-kHz USV, while response calls rarely

[1] https://github.com/UEFepilepsyAIVI/DeepSqueak.git

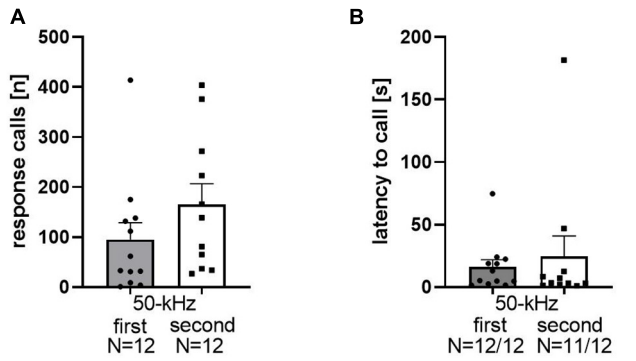


FIGURE 1 | Total numbers of response calls emitted **(A)** for playback of 50-kHz as the first stimulus [95.25 ± 33.62 (mean ± SEM)] or as the second stimulus (152 ± 40.07) of WI rats. Latencies after stimulus onset **(B)**: 50-kHz first: 16.04 ± 5.67; 50-kHz second: 24.67 ± 16.19.

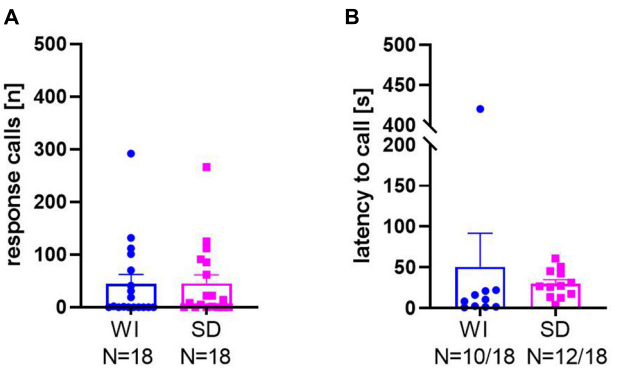


FIGURE 2 | Total numbers of response calls **(A)** and latencies to call **(B)** in Wistar (WI) and Sprague-Dawley (SD) rats.

occurred during noise playback (WI: $t_{17} = 2.717$, $p = 0.015$; SD: $t_{17} = 2.727$, $p = 0.014$).

Data Set 1 and 2: Detailed Analyses

Temporal Emission Pattern

We next pooled the data sets 1 and 2 and performed more detailed analyses. First, a detailed temporal analysis revealed that the emission of response calls was strongly dependent on stimulus ($F_{1,\ 58} = 21.260$, $p < 0.001$) and stimulus phase ($F_{2,\ 116} = 21.120$, $p < 0.001$), with an interaction between stimulus and stimulus phase ($F_{2,\ 116} = 21.002$, $p < 0.001$), while stock had no major impact (stock: $F_{1,\ 58} = 2.311$, $p = 0.134$; stock × stimulus: $F_{1,\ 58} = 2.253$, $p = 0.139$; stock × stimulus phase: $F_{2,\ 116} = 2.308$, $p = 0.104$; stock × stimulus × stimulus phase: $F_{1,\ 116} = 2.290$, $p = 0.106$; **Figure 3**). Specifically, playback of 50-kHz USV but not noise led to a prominent increase in response calls, which occurred during the 5 min of 50-kHz USV playback and up to 5 min thereafter. The peak of vocalization typically occurred in the second or third minute after 50-kHz USV playback onset. With onset of the 50-kHz USV playback, the numbers of emitted response calls increased significantly in WI ($F_{1,\ 41} = 27.940$, $p < 0.001$) and SD rats ($F_{1,\ 17} = 7.436$, $p = 0.014$). After that, calling rate decreased to zero at the latest 5 min after the playback had ended. In both stocks, substantial calling only occurred in response to 50-kHz USV playback and not in response to noise, reflecting high specificity of response call emission (WI: $F_{1,\ 41} = 25.387$, $p < 0.001$; SD: $F_{1,\ 17} = 7.538$, $p = 0.014$). Furthermore, the call emission sequence showed that most animals started calling with higher frequencies around 50 kHz and quickly changed to emit calls of frequencies around 22 kHz (**Supplementary Figure 1A**).

Response Call Features

Secondly, detailed analyses of acoustic features revealed that the calls in response to 50-kHz USV playback were heterogeneous since they were characterized by a large variability in acoustic features and shapes. Both, WI and SD rats emitted calls below and above 32 kHz. These calls had rather different durations and shapes, and the temporal spaces between them varied substantially.

For a further quantification of the response calls, mean peak frequencies, mean call durations, and mean frequency modulations were quantified (**Figure 4**; for examples of response calls, see **Figure 5**). Peak frequencies of WI rats (32.48 ± 1.46 kHz) and SD rats (37.82 ± 3.2 kHz; **Figure 4A**) did not differ significantly from each other ($t_{15.82} = 1.52$, $p = 0.149$). Call durations of WI rats (0.34 ± 0.03 s) tended to be longer than those of SD rats (0.24 ± 0.05 s; $t_{43} = 1.859$, $p = 0.07$). Frequency modulations did not differ between stocks ($t_{43} = 0.98$, $p = 0.33$; WI: 6.68 ± 0.51 kHz; SD: 7.68 ± 0.97 kHz).

To visualize the different call parameters and the distribution of individual calls, scatter plots for either call durations or frequency modulations were plotted vs. peak frequencies (**Figure 6**). This analysis showed that most calls were below 32 kHz, with durations above and below 0.3 s. Frequency modulations were mainly below 5 kHz. The main distribution of the calls was around mean peak frequencies below 32 kHz in both stocks, but in SD rats also another distribution peak occurred around 50 kHz, with call durations typically shorter than 0.3 s and frequency modulations below 5 kHz (**Figures 6B,D**).

Next, we quantified call numbers depending on acoustic call features and divided response calls into those with mean peak frequencies below or above 32 kHz, durations shorter or longer than 0.3 s, and frequency modulations below or above 5 kHz (**Table 1**). This analysis showed that in both stocks the majority of response calls was below 32 kHz. Considering durations, most calls were shorter than 0.3 s, particularly in SD rats. Frequency modulations were mainly below 5 kHz. When comparing the percentages of calls with mean peak frequencies below 32 kHz among stocks, WI rats were found to have higher percentages of calls below 32 kHz ($t_{43} = 2.137$, $p = 0.038$). Considering percentages of calls with durations below 0.3 s, stocks did not differ ($t_{43} = -1.95$, $p = 0.058$). The same was true for frequency modulations. Similar percentages of calls were emitted with modulations below 5 kHz in both stocks ($t_{43} = 0.173$, $p = 0.864$).

In addition, we asked whether response calls below or above 32 kHz were related to each other in individual

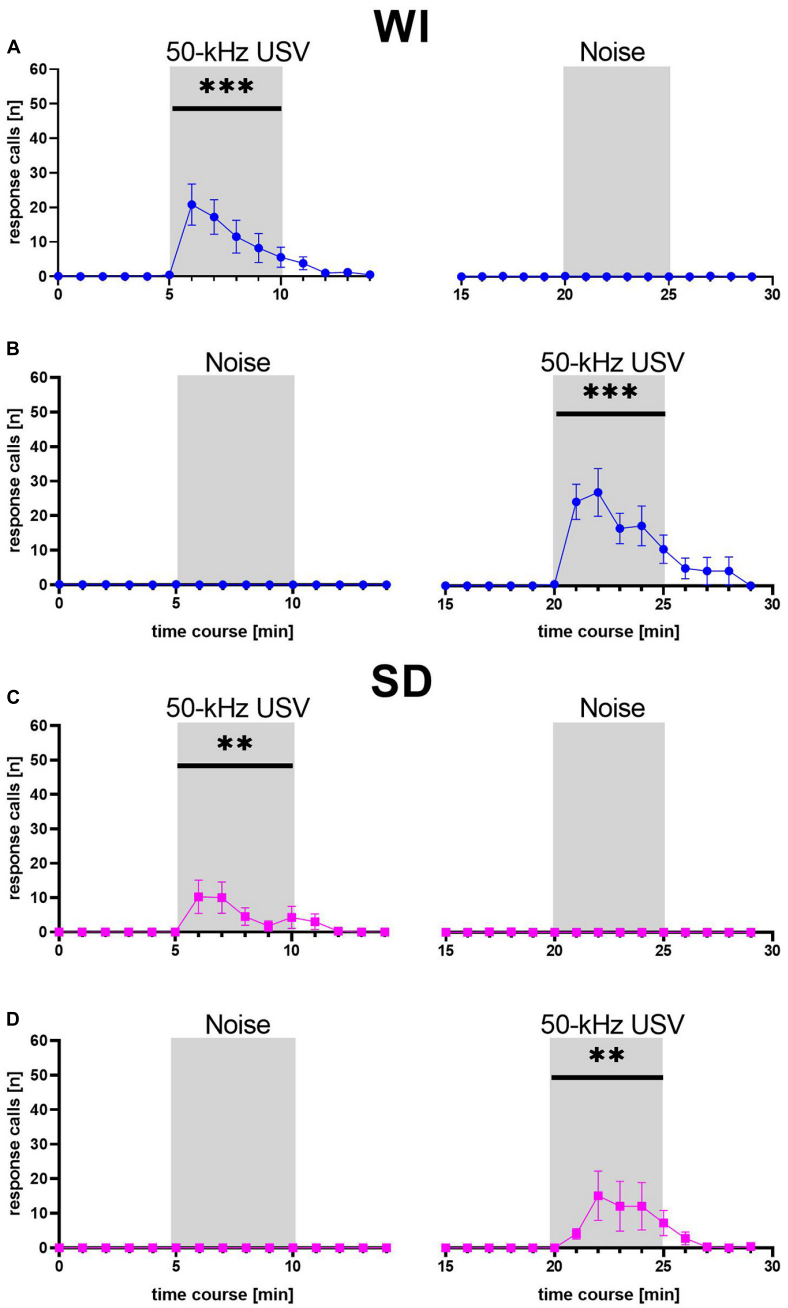


FIGURE 3 | Mean numbers (±SEM) of response calls emitted during each minute of WI (blue dots; **A,B**) and SD (magenta squares; **C,D**) rats. $^{**}p < 0.01$, $^{***}p < 0.001$.

animals (**Figure 7**), but did not find significant correlations between the two in WI ($r = 0.08$, $p = 0.66$) or SD rats ($r = -0.26$, $p = 0.44$).

Relationships Between Response Calls and Playback-Induced Approach

Thirdly, we asked whether the emission of the response calls was correlated with social approach behavior evoked by playback of 50-kHz USV. As stated in the Introduction, the present response call data sets were obtained in a study where social approach behavior evoked by 50-kHz USV playback was examined (Berz et al., 2021). In that study, approach behavior was quantified by subtracting the time spent on the proximal arms (i.e., close to the speaker) before playback from the time spent thereon during the presentation of 50-kHz USV. The same was done for the proximal entries (see detailed analysis in Berz et al., 2021). These numbers were now correlated with the total amount of response calls evoked by playback of 50-kHz USV to see

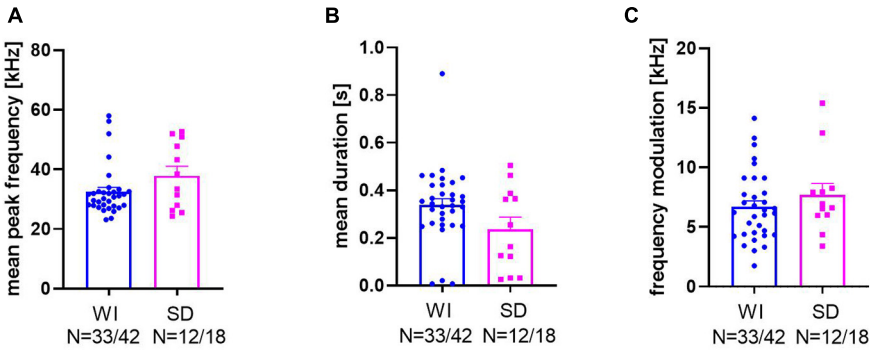


FIGURE 4 | Bar graphs and individual data points of mean peak frequency **(A)**, mean duration **(B)**, and frequency modulation **(C)** of WI (blue dots) and SD rats (magenta squares).

whether social approach behavior was related to the emission of response calls across individual rats. In WI rats, this tended to be the case. The more time the rats spent close to the active speaker, the more calls in response to 50-kHz USV playback they tended to emit (r = 0.314, p = 0.075). A more prominent correlation was evident in SD rats, where social approach and the emission of response calls were strongly associated (SD: r = 0.662, p = 0.019). No such correlations were found with respect to proximal arm entries (WI: r = 0.01, p = 0.952; SD: r = −0.017, p = 0.948). To test whether these correlations were only a byproduct of locomotor activity during playback, the total numbers of response calls were correlated with the degree of locomotor activation using the distance traveled during playback in comparison to the distance traveled before playback. Neither in WI nor SD rats a correlation was found (r = 0.065, p = 0.681; r = 0.151, p = 0.551, respectively). Also, the numbers of response calls were not correlated with locomotor activity during the first 15 min on the maze as a measure of general locomotor activity (WI: r = 0.031, p = 0.864; SD: r = 0.187, p = 0.540).

Data Set 3: Effects of Drug Treatment

Call Numbers and Latencies

In the third data set, rats were treated either with the dopaminergic D2 receptor antagonist Halo or saline as a control. The pharmacological treatment had no prominent effect on the emission of response calls and the proportion of vocalizing rats (saline: 15 out of 24, Halo: 20 out of 24) did not differ between Sal and Halo ($x^2_{1,\ 48}$ = 2.64, p = 0.104). Moreover, treatment did not affect response call numbers (t_{46} = 0.465, p = 0.644; **Figure 8A**; Sal: 66.5 ± 31.18; Halo: 86 ± 31.53) and latencies to start calling (t_{33} = 0.578, p = 0.567; **Figure 8B**; Sal: 19.41 ± 4.18 s; Halo: 26.33 ± 9.86 s).

Temporal Emission Pattern

Similar to the previous data sets 1 and 2, the emission of response calls was strongly dependent on stimulus ($F_{1,\ 46}$ = 11.771, p = 0.001) and stimulus phase ($F_{2,\ 92}$ = 14.443, p < 0.001), with an interaction between stimulus and stimulus phase ($F_{2,\ 92}$ = 14.373, p < 0.001), while treatment had no major impact (treatment: $F_{1,\ 46}$ = 0.194, p = 0.662; treatment × stimulus: $F_{1,\ 46}$ = 0.232, p = 0.632; treatment × stimulus phase: $F_{2,\ 92}$ = 0.842, p = 0.434; treatment × stimulus × stimulus phase: $F_{1,\ 92}$ = 0.797, p = 0.454; **Figure 9**). Specifically, as in the previous data sets 1 and 2, playback of 50-kHz USV but not noise led to a prominent increase in response calls, which occurred during the 5 min of 50-kHz USV playback and up to 5 min thereafter. The peak was again typically seen during the second or third minute after 50-kHz USV playback onset. With onset of 50-kHz USV playback, the numbers of emitted response calls increased significantly in rats treated with Sal ($F_{1,\ 23}$ = 6.443, p = 0.018) but also in rats treated with Halo ($F_{1,\ 23}$ = 8.068, p = 0.009). After that, calling rate decreased to zero at the latest 5 min after playback had ended. Substantial calling only occurred in response to 50-kHz USV and not in response to noise and was therefore specific to the 50-kHz USV playback in both treatment groups (Sal: $F_{1,\ 23}$ = 4.687, p = 0.041; Halo: $F_{1,\ 18}$ = 7.613, p = 0.013). Furthermore, the call emission sequence showed that most animals started calling with higher frequencies around 50 kHz and quickly changed to emit calls of frequencies around 22 kHz (**Supplementary Figure 1B**).

Response Call Features

For a further characterization of response calls in the third data set, their mean peak frequencies, durations, and frequency modulations were analyzed. Sal-treated animals had peak frequencies around 33.76 ± 2.8 kHz, which was not significantly different from Halo-treated animals (30.89 ± 2.49 kHz; t_{33} = 0.898, p = 0.376; **Figure 10A**). Call durations in controls were 0.282 ± 0.036 s, which was significantly shorter than those of Halo-treated rats (0.395 ± 0.039 s; t_{33} = 2.048, p = 0.049, **Figure 10B**). Frequency modulation did not differ between treatment groups and Sal-treated rats called with a frequency modulation of 5.33 ± 0.56 kHz compared to 6.16 ± 0.66 kHz in HALO-treated rats (t_{33} = 0.919, p = 0.365; **Figure 10C**).

The response calls were various in shape and differed in call parameters (for examples of response calls, see **Figure 11**). For better visualization of the different call parameters and the distribution of the individual calls, scatter plots for either call durations or frequency modulations were plotted vs. peak frequencies (**Figure 12**). The accompanying histograms show

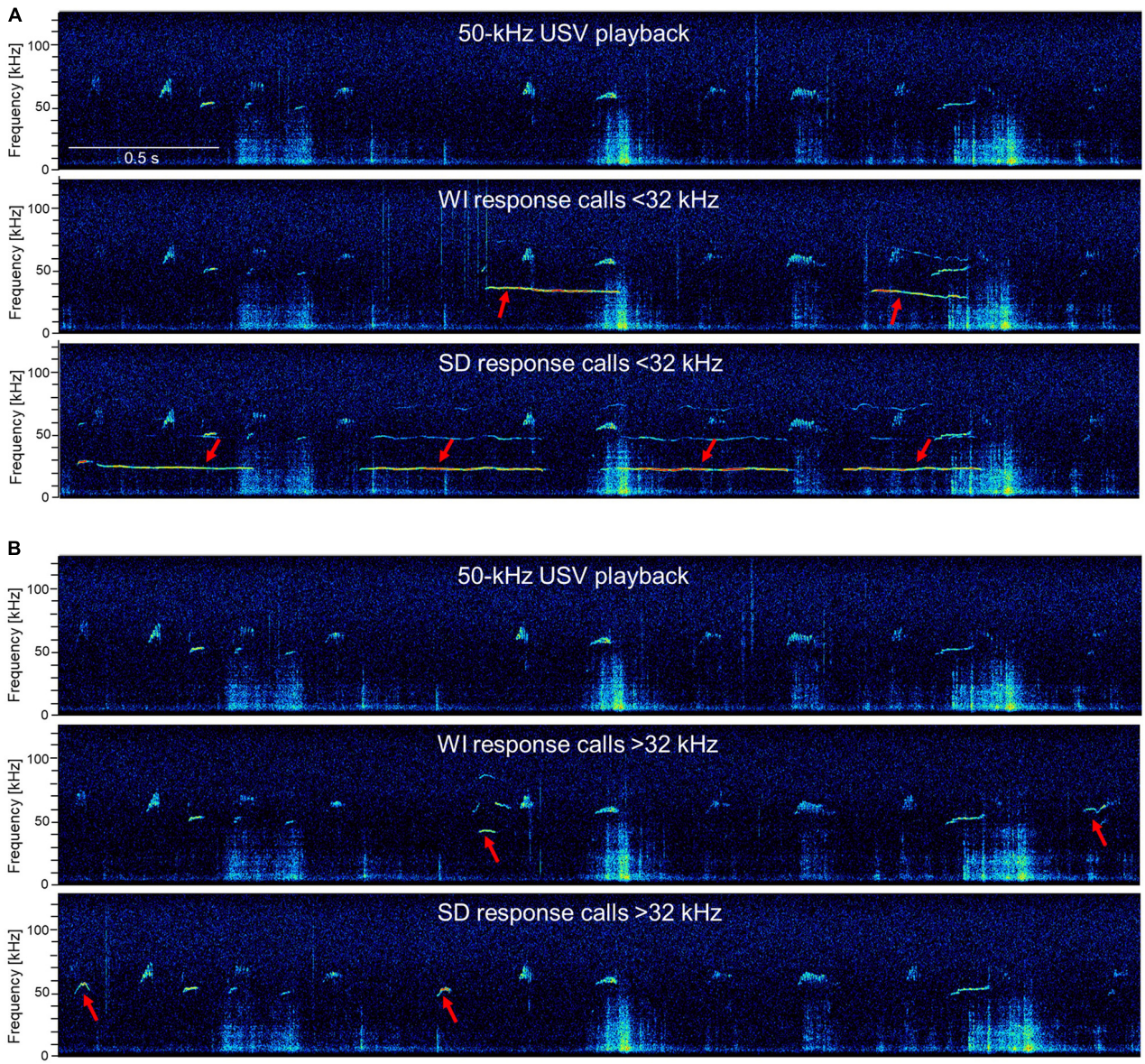


FIGURE 5 | Exemplary response calls during 50-kHz USV playback. The first picture is always the 50-kHz USV playback sequence and the following pictures show response calls in addition to the 50-kHz USV playback sequence (red arrows) < 32 kHz **(A)** or > 32 kHz **(B)** of WI and SD rats. Exemplary high-resolution spectrograms (frequency resolution 488 Hz; time resolution 0.512 ms) were generated with SASLab Pro software 5.2.09 (Avisoft Bioacoustics) *via* fast Fourier transformation (512 FFT length, 100% frame, Hamming window, and 75% time-window overlap).

the main distribution at mean peak frequencies around 25 kHz in both treatment groups; meaning that the majority of calls were below 32 kHz. Especially in Halo-treated rats, few calls were above 32 kHz. Call durations were as well above as below 0.3 s in Sal- and Halo-treated rats. Frequency modulation was mainly below 5 kHz.

Next, we again quantified call numbers depending on acoustic call features and divided response calls into those with mean peak frequencies below or above 32 kHz, durations shorter or longer than 0.3 s, and frequency modulations below or above 5 kHz (**Table 2**). When comparing the percentages of calls with mean peak frequencies below 32 kHz among treatment groups, no significant difference was detected ($t_{33} = -0.978$, $p = 0.335$). Considering durations below 0.3 s, there was likewise no difference ($t_{33} = 1.996$, $p = 0.054$). The same was true for frequency modulations, since similar percentages of calls were emitted with modulations smaller than 5 kHz in both groups ($t_{33} = 0.979$, $p = 0.335$).

In addition, we again asked whether response calls below or above 32 kHz were related in individual animals (**Figure 13**), but found no significant correlations in Sal- ($r = -0.161$, $p = 0.566$) or Halo-treated rats ($r = 0.123$, $p = 0.606$).

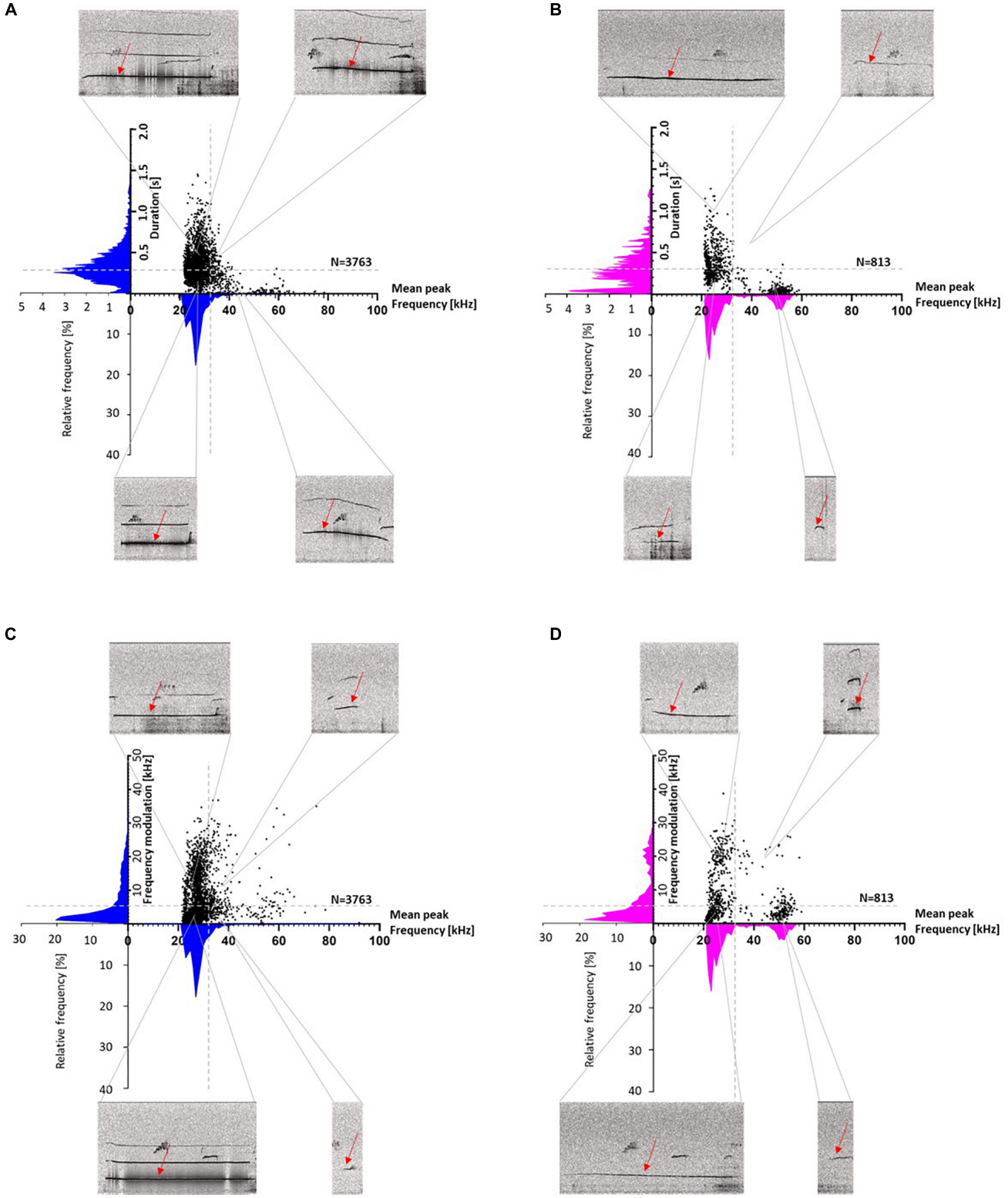


FIGURE 6 | Wistar (WI) **(A,C)** and Sprague-Dawley (SD) rats **(B,D)** scatter plots with histograms (blue for WI and magenta for SD rats) of duration or frequency modulation vs. mean peak frequency. Duration is divided into <> 0.3 s (**A,B**: horizontal gray dashed lines), frequency modulation is divided into <> 5 kHz (**C,D**: horizontal gray dashed lines) and mean peak frequencies are divided into <> 32 kHz (vertical gray dashed line). For each section, an exemplary call with the respective parameters is shown (red arrows).

TABLE 1 | Scatter plot distributions for Wistar (WI) and Sprague-Dawley (SD) rats.

WI ***N* = 33/42**			**Mean peak frequency**		
			=32 kHz	**> 32 kHz**	**Total calls**
	Total numbers (percentages) means ± SEM		3,328 (88.44%) 69.34 ± 5.61	435 (11.56%) 30.66 ± 5.61	3,763 (100%)
Duration	<0.3 s	1,936 (51.44%) 48.69 ± 4.5	1,644 (43.7%)	292 (7.8%)	
	>0.3 s	1,827 (48.56%) 51.34 ± 4.5	1,684 (44.8%)	143 (3.8%)	
Modulation	<5 kHz	2,272 (60.38%) 51.62 ± 4.18	2,104 (55.9%)	168 (4.5%)	
	>5 kHz	1,491 (39.62%) 48.38 ± 4.18	1,224 (32.5%)	267 (7.1%)	
SD ***N* = 12/18**			**Mean peak frequency**		
			=32 kHz	**>32 kHz**	**Total calls**
			599 (73.7%) 44.29 ± 11.95	214 (26.3%) 55.71 ± 11.95	813 (100%)
Duration	<0.3 s	479 (58.9%) 66.79 ± 9.2	273 (33.6%)	206 (25.3%)	
	>0.3 s	334 (41.1%) 33.21 ± 9.2	326 (40.1%)	8 (1%)	
Modulation	<5 kHz	464 (57.1%) 50.21 ± 7.62	324 (39.9%)	140 (17.2%)	
	>5 kHz	349 (42.9%) 49.79 ± 7.62	275 (33.8%)	74 (9.1%)	

Mean peak frequencies < or > 32 kHz, Durations = or > 0.3 s, frequency modulations = or > 5 kHz.

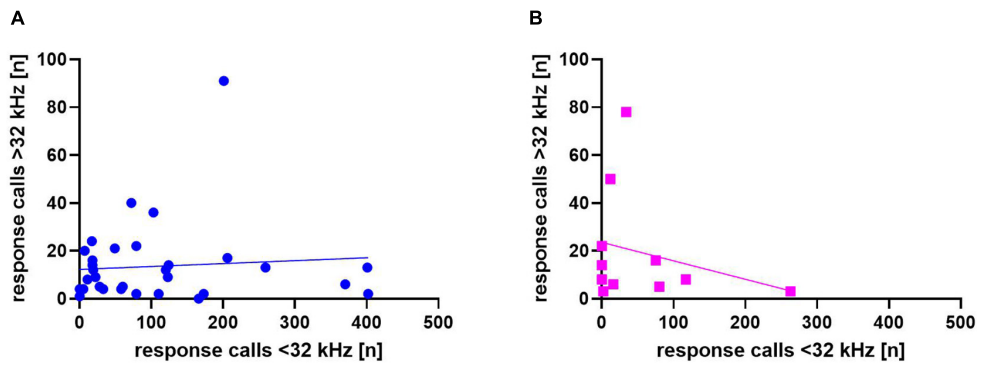


FIGURE 7 | Correlation between calls <32 kHz and >32 kHz for Wistar (WI) **(A)** and Sprague-Dawley (SD) **(B)** rats. Each data point represents response calls below and above 32 kHz of one animal.

Relationships Between Response Calls and Playback-Induced Approach

To see whether social approach was associated with the emission of response calls, these two parameters were again correlated. The results were the same in both treatment conditions. In Sal-treated rats, there were no significant correlations, neither between the time spent in the proximal arms close to the active speaker nor between the entries into those with the amount of response calls (Sal time: $r = -0.0195$, $p = 0.487$; Sal entries: $r = 0.059$, $p = 0.783$). In Halo-treated animals, likewise no significant correlations between proximal time or entries and number of emitted calls were detected (Halo time: $r = 0.143$, $p = 0.547$; Halo entries: $r = -0.112$, $p = 0.602$). Moreover, locomotor activity during 50-kHz USV playback in comparison to the distance traveled before playback was not correlated with the total numbers of response calls, irrespective of treatment condition (Sal: $r = -0.101$, $p = 0.639$; Halo: $r = -0.113$, $p = 0.598$). In addition, locomotor activity during the first 15 min on the platform was not correlated with the number of response calls (Sal: $r = -0.224$, $p = 0.421$; Halo: $r = 0.238$, $p = 0.312$).

DISCUSSION

In this study, we characterized response calls emitted by rats exposed to playback of appetitive 50-kHz USV, previously shown to function as social contact calls (Wöhr, 2018). The phenomenon that rats respond to playback of species-specific 50-kHz calls by emitting response calls has been repeatedly reported before, but has not been described in detail yet (Wöhr and Schwarting, 2007, 2009; Willadsen et al., 2014; Willuhn et al., 2014; Engelhardt et al., 2017, 2018; Berg et al., 2018, 2021; Kisko et al., 2020; Olszyński et al., 2020, 2021; for an overview see **Supplementary Table 1**). First, we described the emission of response calls in reaction toward 50-kHz USV playback in

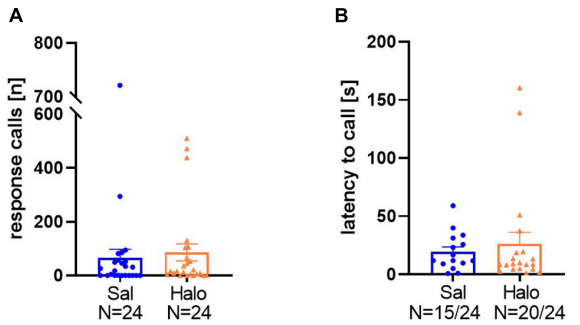


FIGURE 8 | Total numbers of response calls **(A)** and latencies to call **(B)** in Sal- and Halo-treated rats. Data are presented as individual results and as means ± SEM.

WI rats. Secondly, we compared these to SD rats. Thirdly, we analyzed the effect of blocking DA receptors on response calls using Halo, as compared to vehicle-injected WI subjects.

Through these means, we could demonstrate that most rats emitted response calls. Importantly, the emission of response calls was clearly linked to 50-kHz USV playback. In fact, response calls were seen specifically in response to 50-kHz USV but not in response to time- and amplitude-matched noise, replicating previous results (Willadsen et al., 2014; Willuhn et al., 2014; Engelhardt et al., 2017, 2018; Berg et al., 2018, 2021; Kisko et al., 2020; Olszyński et al., 2020, 2021). When exposed to 50-kHz USV, receiver rats often started emitting response calls within the first minute of playback and emission rates were typically peaking after around 2–3 min, often outlasting playback for up to 5 min. This certainly supports naming these calls "response calls."

Most response calls were characterized by peak frequencies below 32 kHz, the threshold typically applied to differentiate between 22- and 50-kHz USV (Brudzynski, 2001). Although peak frequencies were highly variable and ranged roughly between 20 and 80 kHz, the vast majority of response calls occurred in a frequency range of 20–32 kHz. Similarly, call durations were characterized by large variability, ranging from a few milliseconds to up to 1.5 s. Call durations of about 0.3 s occurred at a particularly high rate. Frequency modulations were typically below 5 kHz. When comparing these values to the parameters of typical 22- and 50-kHz USV, our values correspond more to 22-kHz USV; more precisely the short 22-kHz USV type since the durations were rarely longer than 0.3 s (Brudzynski, 2021).

The emission of response calls was seen in WI and SD rats, suggesting that this is a robust phenomenon not dependent on stocks. Specifically, we found that there were no substantial differences between WI and SD rats, concerning numbers of emitted calls, latencies to start calling, and call likelihood. In both stocks there was a large variability among response calls. However, their mean peak frequencies, call durations, and frequency modulations did not differ significantly between experimental conditions. SD rats only differed in one aspect by clearly showing calls around frequencies of 50 kHz, which was not that prominent in WI rats. This is somehow in line with other studies that also showed higher emission of 50-kHz USV and elevated rough-and-tumble play behavior in SD compared to WI rats (Manduca et al., 2014). Other studies, however, found that WI rats emitted more 50-kHz USV compared to SD rats (Schwarting, 2018a,b), indicating that WI rats are more prone to emit USV in general, which is also not represented by our data. If at all, on a descriptive level, WI rats emit slightly less response calls compared to SD rats. Regarding call parameters, previous studies showed marginal differences between stocks, i.e., shorter call durations in SD rats compared to WI rats (Schwarting, 2018b). On a descriptive level again, this aligns with our results, albeit this difference in call duration did not yield significance. Apart from stock differences, various other factors like breeding or experience have to be taken into account. Moreover, inter-individual differences should not be neglected, as our results also suggest (Schwarting, 2018a,b).

In our study, the pharmacological treatment with the D2 antagonist Halo did not affect call likelihood, call rates, latencies, temporal distribution, peak frequency, and frequency modulation. In Sal-treated WI rats, the majority of calls was again below 32 kHz, however, in Halo-treated rats this was even more prominent and Halo treatment also led to longer call durations. Previous studies showed that exposure to 50-kHz USV playback under the influence of systemically applied amphetamine, a catecholaminergic agonist, resulted in response calls with frequencies around 50 kHz at the expense of 22 kHz (Engelhardt et al., 2017). Specifically, calls of lower frequencies decreased drastically under the influence of amphetamine. In contrast, response calls in the 50 kHz range increased dose-dependently following amphetamine administration. This is in line with a large number of studies showing that the emission of 22- and 50-kHz USV are associated with the activation of distinct neurotransmitter systems (for review: Brudzynski, 2021). While 22-kHz USV are associated with the cholinergic system (Brudzynski, 2001; Kroes et al., 2007; Willadsen et al., 2018), the dopaminergic system plays an important role in the regulation of 50-kHz USV (Wöhr, 2021). For instance, electrolytic or 6-hydroxydopamine lesions of the ventral tegmental area reduce the emission of 50-kHz USV (Burgdorf et al., 2007). Conversely, emission of 50-kHz USV can be evoked by electrical stimulation of the ventral tegmental area or the nucleus accumbens (Burgdorf et al., 2000, 2007). Moreover, psychostimulants, most notably amphetamine, lead to a robust increase in 50-kHz USV emission (Rippberger et al., 2015). Additionally, playback of 50-kHz USV was shown to induce enhanced levels of activity in the nucleus accumbens (Sadananda et al., 2008), where it elicits a rapid phasic release of dopamine (Willuhn et al., 2014). Based on these findings, one could have assumed that the dopaminergic receptor blockade with Halo should decrease response call numbers, especially those above 32 kHz, which was apparently not the case. Possibly, these calls are not critically dependent on dopamine D2 receptor function, and might be dependent on endogenous opiates, as indicated by an earlier playback study with the opiate receptor antagonist naloxone (Wöhr and Schwarting, 2009).

Together, the present findings indicate that the emission of response calls is a robust phenomenon that is seen specifically in response to playback of 50-kHz USV independent of stock and despite blocking dopamine neurotransmission. These

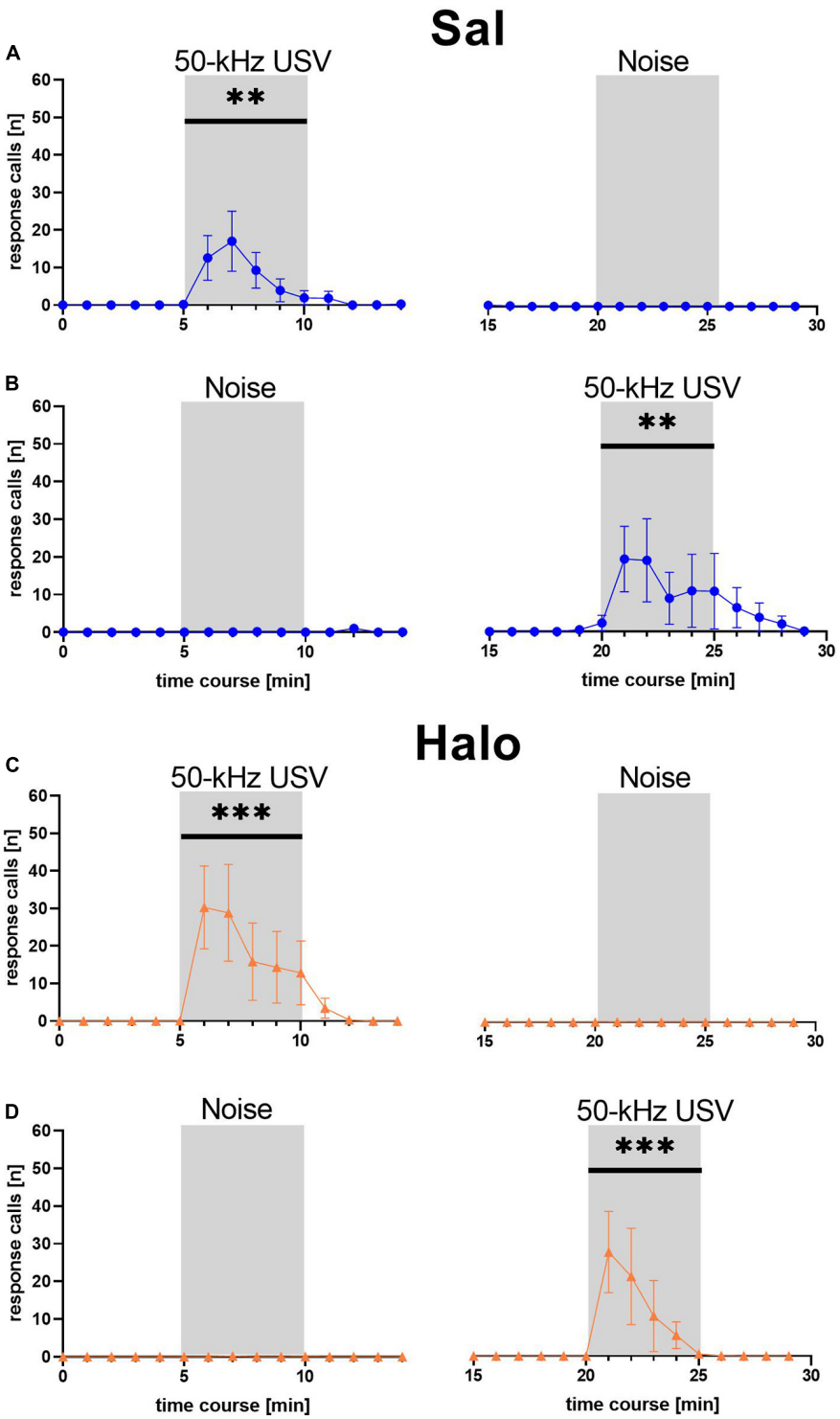


FIGURE 9 | Mean number of response calls emitted during each minute of Sal- (blue; **A,B**) and Halo-treated (orange; **C,D**) rats. Most calls were emitted during 50-kHz USV stimulus and almost no calls were emitted during noise. $^{**}p < 0.01$, $^{***}p < 0.001$.

observations are in line with the idea that the emission of response calls reflects changes in affect that are caused by playback of 50-kHz USV. For example, one might expect the induction of a positive affective state in response to appetitive 50-kHz USV. On the other hand, it was suggested that response calls reflect frustration induced by the inability to reach the

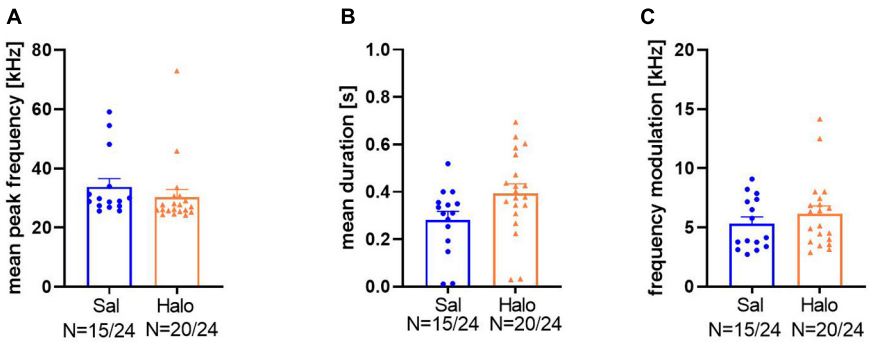


FIGURE 10 | Call parameters of Sal- (blue) and Halo-treated (orange) rats for mean peak frequency **(A)**, mean duration **(B)**, and frequency modulation **(C)**.

conspecific emitting 50-kHz USV. Alternatively, response calls might serve communicative functions as social contact calls or as appeasement signals. While the present findings do not allow drawing strong conclusions about causes and functional significance of response calls, they provide first insights into potential mechanisms underlying their emission.

In support of the idea that response calls might reflect an affective state we hypothesize that the rats are not solely in one affective state, but rather in an ambivalent state. There is convincing evidence in support of the notion that USV emission reflects prominent affective states (Brudzynski, 2021) and that different call types are associated with distinct states (Brudzynski, 2013b). Because USV below 32 kHz are typically believed to function as alarm or distress calls reflecting a negative affective state, this would suggest that playback of 50-kHz USV induced a negative state in the receiver rats. However, the strong level of social approach behavior and the emission of 50-kHz response calls, at least in SD rats, evoked by playback of 50-kHz USV speaks against the induction of a solely negative affective state through 50-kHz USV playback (Wöhr, 2018). Furthermore, the positive and negative emotional states in rats were proposed to be mutually exclusive and acting in an antagonistic manner (Brudzynski, 2021). It is possible, however, that the two states quickly alternate which leads to the hypothesis of an ambivalent state, with negative and positive phases present in an oscillating manner. This is also reminiscent of an approach/avoidance conflict, i.e., a situation characterized by choices leading to either reward or punishment (Aupperle et al., 2015). Interestingly, it was shown that rats emit 22-kHz as well as 50-kHz USV during neutral situations and not only aversive ones (Robakiewicz et al., 2019). The study by Robakiewicz et al. (2019) also showed that both call types and hence presumably both emotional states can be present during an emotional neutral task of performing nose pokes in order to change the light of the experimental apparatus. Both call types were also found in a cocaine self-administration task (Barker et al., 2010), where animals received either high or low doses of cocaine. Low dose rats predominantly emitted short 22-kHz calls and high dose rats emitted mostly 50-kHz calls. Nevertheless, both groups showed calls of both emotional states and this supports the hypothesis of the ambivalent state. In the present study, however, only SD rats emitted 50-kHz USV to a higher extent and all other experimental groups mainly emitted calls with frequencies below 32 kHz. Additionally, the emissions of response calls below and above 32 kHz were not correlated across individual rats, suggesting that there was no general tendency for emitting response calls in both frequency ranges, which speaks against the hypothesis of an ambivalent state.

With respect to the emission of 22-kHz calls, this phenomenon might be explained by the hypothesis of a frustrated state in the receiver rat, possibly induced by the violated expectation of another rat being present. Other studies suggested that short 22-kHz calls (<0.3 s) represent a dysphoric state or displeasure without any external threat (Simmons et al., 2018), which is in line with the mean peak frequencies, durations, and low frequency modulations of the response calls found in our study. This might also be an indication that calls with low frequencies in response toward 50-kHz USV playback are an expression of internal distress, i.e., frustration, as suggested before (Wöhr and Schwarting, 2009). Frustration is defined as a result of behavior after an expected but not received reward (Scull et al., 1970; Burokas et al., 2012). In our playback paradigm, the rat probably realized that there was no rat physically present for interaction after hearing the 50-kHz USV playback, and this could have led to a state of frustration in the approaching rat. This might also explain why the majority of response calls was emitted within 2 or 3 min after the onset of the 50-kHz USV playback. At first, the animals heard and recognized the stimulus, exhibited a strong social approach immediately afterward and as soon as the rats realized that there was no conspecific present, the emission of response calls increased as an expression of a frustrated state. In line with the frustrated state hypothesis is our finding that the first calls of most animals of data set 2 and 3 were of higher frequencies, i.e., around 50 kHz and quickly changed to calls with frequencies in the 22-kHz USV range (**Supplementary Figure 1**).

On the other hand, the positive correlation of response calls and approach behavior might serve the hypothesis that the response calls could also be characterized as social contact calls. 50-kHz USV have been postulated to fulfill an affiliative communication function to, for example, maintain a playful state during rough-and-tumble play or as social contact calls to reestablish social proximity after separation of conspecifics (Wöhr et al., 2016). An indication that the response calls in

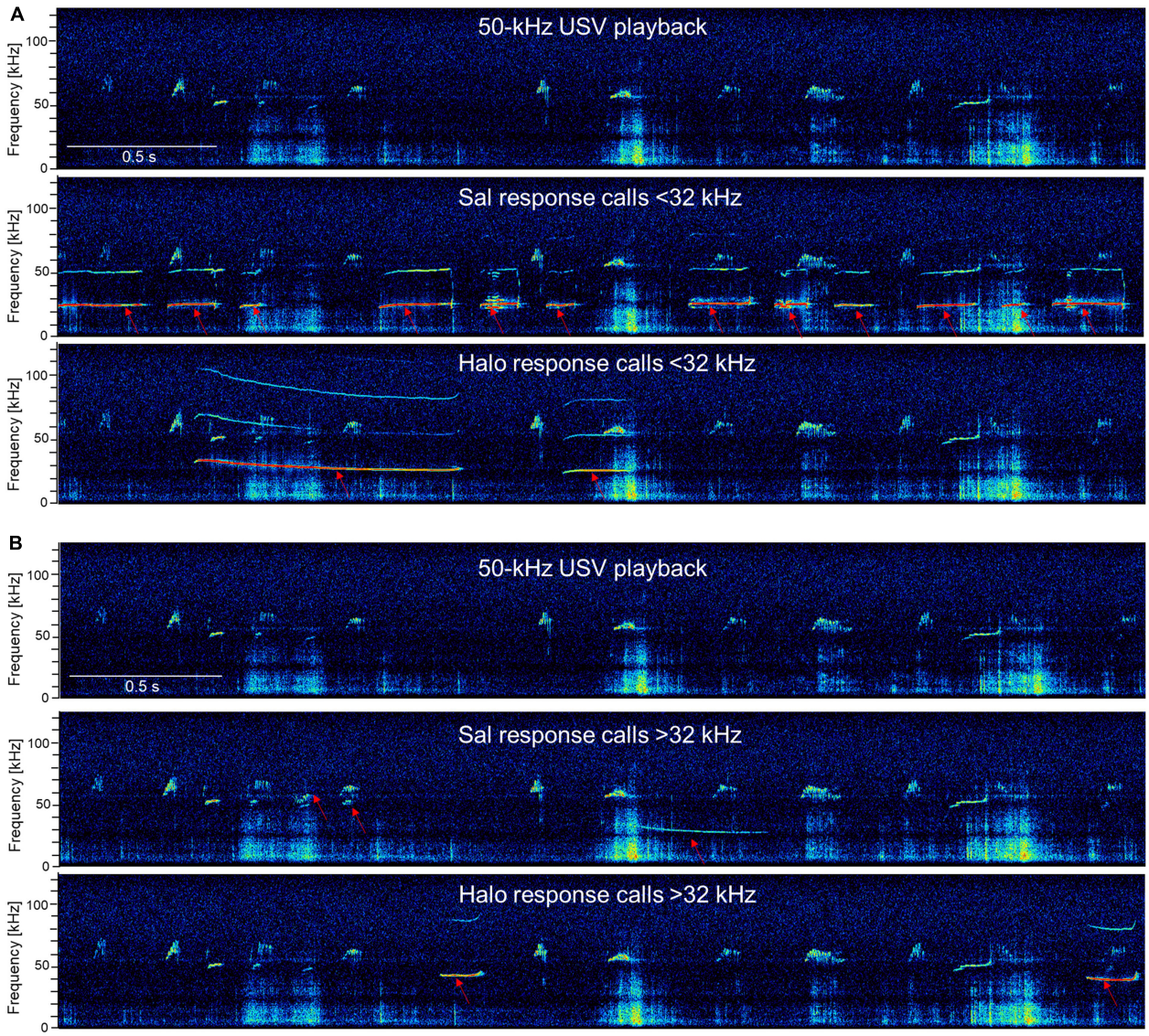


FIGURE 11 | Exemplary response calls during 50-kHz USV playback. The first picture is always the 50-kHz USV playback sequence and the following pictures show response calls in addition to the 50-kHz USV playback sequence (red arrows) <32 kHz **(A)** or >32 kHz **(B)** of Sal- and Halo-treated rats. Note that the calls depicted for Sal- or Halo-treated rats are not specific to the treatment groups and calls were descriptively similar in all groups.

our study serve as social contact calls is that they are emitted during social approach behavior. Further, such calls are emitted frequently during the approach behavior like 50-kHz USV during rough-and-tumble play (Knutson et al., 1998). In our study we found a moderate positive correlation between response calls and approach behavior, i.e., the time spent close to the active speaker, in SD and, at least to some extent, in WI rats. Apparently, the more the animals tried to reach a possible conspecific signaled by the 50-kHz USV playback, the more calls they emitted, supporting the hypothesis of response calls being contact calls. For Sal- and Halo-treated WI rats, however, this was not the case. In Halo-treated rats, the absence of a positive correlation between approach behavior and response call emission was probably due to the drug-induced immobility (Berz et al., 2021). Since Sal-treated rats also received an i.p. injection 60 min prior to testing, this might have influenced their approach response, as well as their calling behavior; even though Sal-treated rats significantly approached the sound source (Berz et al., 2021) and emitted similar numbers of response calls as WI rats. No correlation was found, however, between overall activity and call numbers in any group. Also or alternatively, the positive correlation between approach behavior and response calls especially observed in SD rats might not be in order to establish contact, but rather due to hypervigilance. Olszyński et al. (2021) showed that in response to 50-kHz USV playback, heart rate and locomotor activity increased as well as the emission of USV. The USV in response

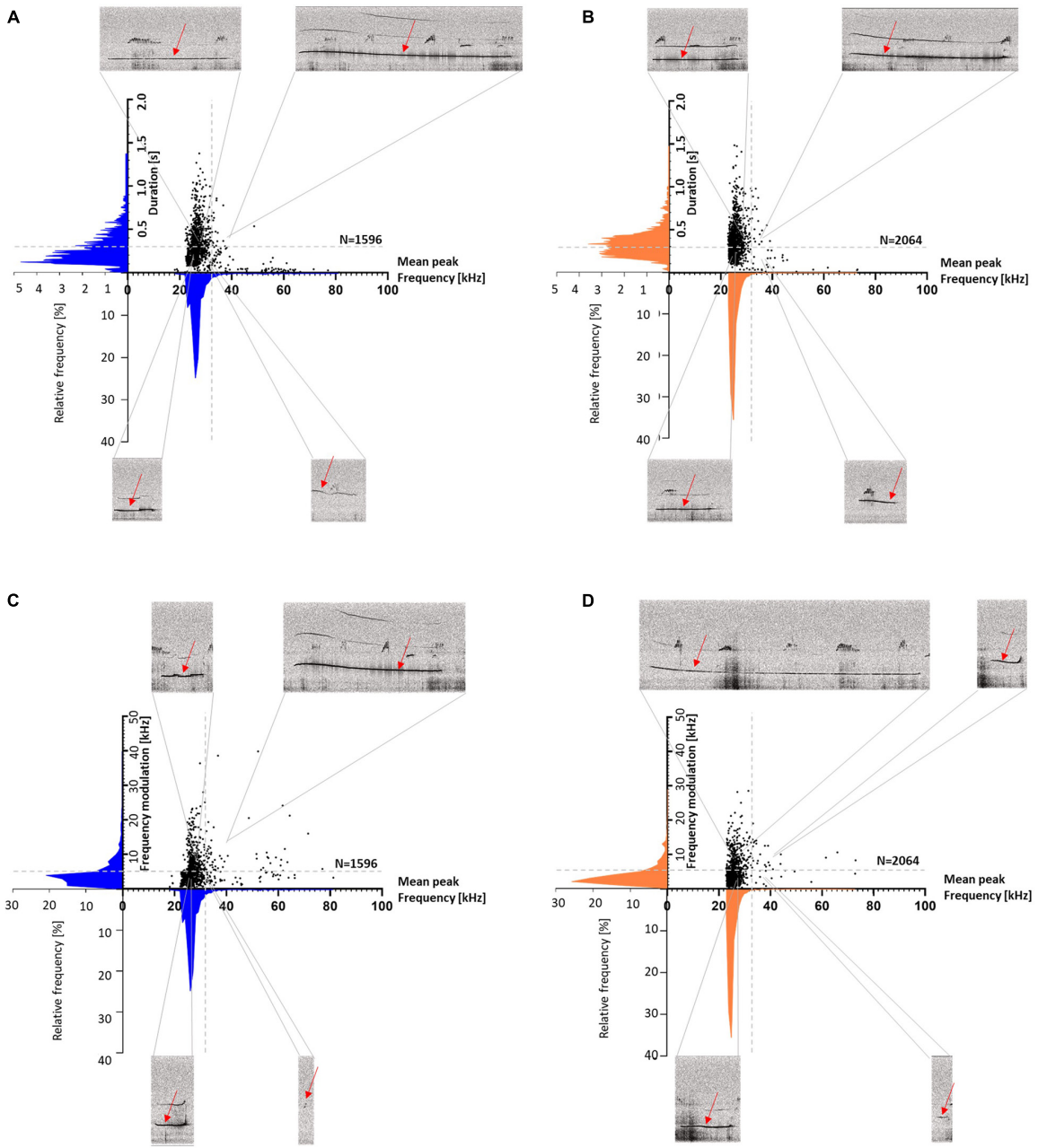


FIGURE 12 | Sal- **(A,C)** and Halo-treated rats **(B,D)** scatter plots with histograms (blue for Sal- and orange for Halo-treated rats) of duration or frequency modulation vs. mean peak frequency. Duration is divided into <> 0.3 s (**A,B**: horizontal gray dashed lines), frequency modulation is divided into <> 5 kHz (**C,D**: horizontal gray dashed lines) and mean peak frequencies are divided into <> 32 kHz (vertical gray dashed line). For each section, an exemplary call with the regarding parameters is shown (red arrows).

to 50-kHz USV playback in that study were mainly 50-kHz calls, possibly representing contact calls, in contrast to our study here, where the animals mostly emitted calls of lower frequencies. Also, the peak of call emission occurred shortly after the recipient of the playback was in proximity to the sound source and ceased after playback has stopped, which suggests that these calls could function to establish social contact or in search of it. However, the response calls linked to the 50-kHz USV playback do not classify as 50-kHz calls because their mean peak frequencies are much lower, the duration is longer, and there is hardly any frequency modulation compared to 50-kHz calls.

Alternatively, response calls could serve appeasing purposes. The age difference between the rat of the recorded playback and the test subject might be of interest, because in our study, a juvenile rat heard 50-kHz USV playback recorded from an adult rat and accordingly, it seems plausible for the

TABLE 2 | Scatter plot distributions for Sal- and Halo-treated rats.

Sal ***N* = 15/24**			**Mean peak frequency**		
			=32 kHz	**>32 kHz**	**Total calls**
	Total numbers (percentages) means ± SEM		1,490 (93.4%) 72.77 ± 8.78	106 (6.6%) 27.23 ± 8.78	1,596 (100%)
Duration	<0.3 s	1,044 (65.4%) 59.39 ± 5.91	960 (60.2%)	84 (5.3%)	
	>0.3 s	552 (34.6%) 40.61 ± 5.91	530 (33.2%)	22 (1.4%)	
Modulation	<5 kHz	1,147 (71.9%) 60.27 ± 6.73	1,098 (68.8%)	49 (3.1%)	
	>5 kHz	449 (28.1%) 39.73 ± 6.73	392 (24.6%)	57 (3.6%)	
Halo ***N* = 20/24**			**Mean peak frequency**		
			=32 kHz	**> 32 kHz**	**Total calls**
			2,021 (97.9%) 83.44 ± 6.78	43 (2.1%) 16.56 ± 6.78	2,064 (100%)
Duration	<0.3 s	919 (44.5%) 40.48 ± 6.89	893 (43.3%)	26 (1.3%)	
	>0.3 s	1,145 (55.5%) 59.52 ± 6.89	1,128 (54.7%)	17 (0.8%)	
Modulation	<5 kHz	1,633 (79.1%) 50.76 ± 6.73	1,624 (78.7%)	9 (0.4%)	
	>5 kHz	431 (20.9%) 49.24 ± 6.73	397 (19.2%)	34 (1.6%)	

Mean peak frequencies < or > 32 kHz, Durations = or > 0.3 s, frequency modulations = or > 5 kHz.

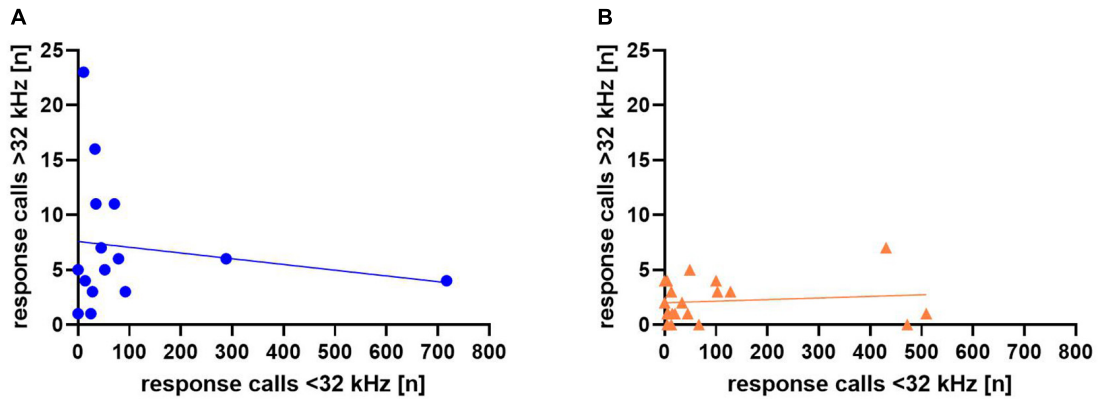


FIGURE 13 | Correlation between calls <32 kHz and >32 kHz for Sal- **(A)** and Halo-treated **(B)** rats. Each data point represents response calls below and above 32 kHz of one animal.

subject rat to cautiously approach the potential conspecific. Supporting this hypothesis, is the fact that in adult male rats, USV calls of lower frequencies were found during play fighting (Burke et al., 2017, 2020). In social situations that were at risk to escalate into aggression, the play partners lowered their calls gradually from 50 kHz to around 30 kHz with increasing durations (Burke et al., 2017). The authors hypothesized that this group of calls might be a transition from 50-kHz flats to 22-kHz flats or a unique new type of calls. The function of these calls is probably the induction of appeasement, i.e., to de-escalate a situation at risk to turn into aggression (see also Sales, 1972; Lore et al., 1976). Our results seem to support this hypothesis since we tested juvenile rats subjected to calls from an older adult rat and the response calls were in similar frequencies. Moreover, the response calls had also similar frequency modulations, like the calls in the study by Burke et al. (2017) and were not exclusively flat as the common 22-kHz USV. So far, however, it is not known whether receiver rats can gain information about the age of the sender based on their USV.

Importantly, the response call phenomenon studied here in detail appears sufficiently robust to be used as a measure for the reciprocal nature of acoustic communication and can easily be applied in rat model systems for neuropsychiatric disorders, where acoustic communication is impaired, such as autism spectrum disorder (Lai and Baron-Cohen, 2015). In preclinical studies examining USV with the aim to reveal communication deficits in rodent model systems, most laboratories have focused exclusively on the sender. Although there is now an increasing number of preclinical studies including playback paradigms to learn about the responses evoked in the receiver as well (Berg et al., 2018, 2020a,b; Kisko et al., 2018, 2020; Wöhr et al., 2020),

an important aspect of acoustic communication that is often still neglected is its reciprocal nature and the fact that a signal emitted by the sender frequently evokes the emission of a response signal in the receiver (Seyfarth and Cheney, 2003). Measuring response calls offers a unique opportunity to overcome this limitation. It offers a new approach to studying the reciprocal nature of communication in rodent models for neuropsychiatric disorders.

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 Tierschutzbehoerde, Regierungspraesidium Giessen, Germany, TVA No. 35-2018.

AUTHOR CONTRIBUTIONS

RS and MW designed the study, acquired resources and funding, and oversaw the project. AB performed the experiments. AB with substantial help from MW analyzed the data. AB, RS, and MW wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grant SCHW 559/15-1 from the Deutsche Forschungsgemeinschaft (DFG) and the DFG funded Research Training Group "Breaking Expectation" (GRK 2271).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2021.812142/full#supplementary-material

REFERENCES

Aupperle, R. L., Melrose, A. J., Francisco, A., Paulus, M. P., and Stein, M. B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Hum. Brain Mapp.* 36, 449–462. doi: 10.1002/HBM.22639

Barker, D. J., Root, D. H., Ma, S., Jha, S., Megehee, L., Pawlak, A. P., et al. (2010). Dose-dependent differences in short ultrasonic vocalizations emitted by rats during cocaine self-administration. *Psychopharmacology* 211, 435–442. doi: 10.1007/s00213-010-1913-9

Berg, E. L., Ching, T. M., Bruun, D. A., Rivera, J. K., Careaga, M., Ellegood, J., et al. (2020a). Translational outcomes relevant to neurodevelopmental disorders following early life exposure of rats to chlorpyrifos. *J. Neurodev. Disord.* 12:40. doi: 10.1186/s11689-020-09342-1

Berg, E. L., Pride, M. C., Petkova, S. P., Lee, R. D., Copping, N. A., Shen, Y., et al. (2020b). Translational outcomes in a full gene deletion of ubiquitin protein ligase E3A rat model of Angelman syndrome. *Transl. Psychiatry* 10:39. doi: 10.1038/s41398-020-0720-2

Berg, E. L., Copping, N. A., Rivera, J. K., Pride, M. C., Careaga, M., Bauman, M. D., et al. (2018). Developmental social communication deficits in the Shank3 rat model of phelan-mcdermid syndrome and autism spectrum disorder. *Autism Res.* 11, 587–601. doi: 10.1002/aur.1925

Berg, E. L., Jami, S. A., Petkova, S. P., Berz, A., Fenton, T. A., Segal, D. J., et al. (2021). Excessive laughter-like vocalizations, microcephaly, and translational outcomes in the Ube3a deletion rat model of Angelman Syndrome. *J. Neurosci.* 41, 8801–8814. doi: 10.1523/JNEUROSCI.0925-21.2021

Berz, A., Pasquini de Souza, C., Wöhr, M., and Schwarting, R. K. W. (2021). Limited generalizability, pharmacological modulation, and state-dependency of habituation towards pro-social 50-kHz calls in rats. *IScience* 24:102426. doi: 10.1016/j.isci.2021.102426

Blanchard, R. J., Blanchard, D. C., Agullana, R., and Weiss, S. M. (1991). Twenty-two kHz alarm cries to presentation of a predator, by laboratory rats living in visible burrow systems. *Physiol. Behav.* 50, 967–972. doi: 10.1016/0031-9384(91)90423-L

Brudzynski, S. M. (2001). Pharmacological and behavioral characteristics of 22 kHz alarm calls in rats. *Neurosci. Biobehav. Rev.* 25, 611–617. doi: 10.1016/S0149-7634(01)00058-6

Brudzynski, S. M. (2010). Handbook of mammalian vocalization. *Handb. Behav. Neurosci.* 19, 13–21. doi: 10.1016/B978-0-12-374593-4.00002-4

Brudzynski, S. M. (2013a). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Curr. Opin. Neurobiol.* 23, 310–317. doi: 10.1016/j.conb.2013.01.014

Brudzynski, S. M. (2013b). "Vocalizations as indicators of emotional states in rats and cats," in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford: Oxford University Press), 75–91.

Brudzynski, S. M. (2021). Biological functions of rat ultrasonic vocalizations, arousal mechanisms, and call initiation. *Brain Sci.* 11:605. doi: 10.3390/brainsci11050605

Brudzynski, S. M., Bihari, F., Ociepa, D., and Fu, X. W. (1993). Analysis of 22 kHz ultrasonic vocalization in laboratory rats: long and short calls. *Physiol. Behav.* 54, 215–221. doi: 10.1016/0031-9384(93)90102-L

Brudzynski, S. M., and Chiu, E. M. C. (1995). Behavioural responses of laboratory rats to playback of 22 kHz ultrasonic calls. *Physiol. Behav.* 57, 1039–1044. doi: 10.1016/0031-9384(95)00003-2

Burgdorf, J., Knutson, B., and Panksepp, J. B. (2000). Anticipation of rewarding electrical brain stimulation evokes ultrasonic vocalization in rats. *Behav. Neurosci.* 114, 320–327. doi: 10.1037/0735-7044.114.2.320

Burgdorf, J., Kroes, R. A., Moskal, J. R., Pfaus, J. G., Brudzynski, S. M., and Panksepp, J. B. (2008). Ultrasonic vocalizations of rats (*Rattus norvegicus*) during mating, play, and aggression: behavioral concomitants, relationship to reward, and self-administration of playback. *J. Comp. Psychol.* 122, 357–367. doi: 10.1037/a0012889

Burgdorf, J., Wood, P. L., Kroes, R. A., Moskal, J. R., and Panksepp, J. B. (2007). Neurobiology of 50-kHz ultrasonic vocalizations in rats: electrode mapping, lesion, and pharmacology studies. *Behav. Brain Res.* 182, 274–283. doi: 10.1016/j.bbr.2007.03.010

Burke, C. J., Euston, D. R., and Pellis, S. M. (2020). International Journal of Play What do you hear, what do you say? Ultrasonic calls as signals during play fighting in rats. *Int. J. Play* 9, 92–107. doi: 10.1080/21594937.2020.1720126

Burke, C. J., Kisko, T. M., Pellis, S. M., and Euston, D. R. (2017). Avoiding escalation from play to aggression in adult male rats: the role of ultrasonic calls. *Behav. Process.* 144, 72–81. doi: 10.1016/j.beproc.2017.09.014

Burokas, A., Gutiérrez-Cuesta, J., Martín-García, E., and Maldonado, R. (2012). Operant model of frustrated expected reward in mice. *Addict. Biol.* 17, 770–782. doi: 10.1111/J.1369-1600.2011.00423.X

Coffey, K. R., Marx, R. G., and Neumaier, J. F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 859–868. doi: 10.1038/s41386-018-0303-6

Engelhardt, K.-A., Fuchs, E., Schwarting, R. K. W., and Wöhr, M. (2017). Effects of amphetamine on pro-social ultrasonic communication in juvenile rats: implications for mania models. *Eur. Neuropsychopharmacol.* 27, 261–273. doi: 10.1016/j.euroneuro.2017.01.003

Engelhardt, K. A., Schwarting, R. K. W., and Wöhr, M. (2018). Mapping trait-like socio-affective phenotypes in rats through 50-kHz ultrasonic vocalizations. *Psychopharmacology* 235, 83–98. doi: 10.1007/s00213-017-4746-y

Fendt, M., Brosch, M., Wernecke, K. E. A., Willadsen, M., and Wöhr, M. (2018). Predator odour but not TMT induces 22-kHz ultrasonic vocalizations in rats that lead to defensive behaviours in conspecifics upon replay. *Sci. Rep.* 8:11041. doi: 10.1038/s41598-018-28927-4

Kisko, T. M., Braun, M. D., Michels, S., Witt, S. H., Rietschel, M., and Culmsee, C. (2018). Cacna1c haploinsufficiency leads to pro-social 50-kHz ultrasonic communication deficits in rats. *Dis. Model Mech.* 11:dmm034116. doi: 10.1242/dmm.034116

Kisko, T. M., Braun, M. D., Michels, S., Witt, S. H., Rietschel, M., Culmsee, C., et al. (2020). Sex-dependent effects of Cacna1c haploinsufficiency on juvenile social play behavior and pro-social 50-kHz ultrasonic communication in rats. *Genes Brain Behav.* 19:e12552. doi: 10.1111/gbb.12552

Knutson, B., Burgdorf, J., and Panksepp, J. B. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *J. Comp. Psychol.* 112, 65–73. doi: 10.1037/0735-7036.112.1.65

Kroes, R. A., Burgdorf, J., Otto, N. J., Panksepp, J., and Moskal, J. R. (2007). Social defeat, a paradigm of depression in rats that elicits 22-kHz vocalizations, preferentially activates the cholinergic signaling pathway in the periaqueductal gray. *Behav. Brain Res.* 182, 290–300. doi: 10.1016/J.BBR.2007.03.022

Lai, M. C., and Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *Lancet Psychiatry* 2, 1013–1027. doi: 10.1016/S2215-0366(15)00277-1

Lore, R., Flannelly, K., and Farina, P. (1976). Ultrasounds produced by rats accompany decreases in intraspecific fighting. *Aggress. Behav.* 2, 175–181.

Ma, S. T., Maier, E. Y., Ahrens, A. M., Schallert, T., and Duvauchelle, C. L. (2010). Repeated intravenous cocaine experience: development and escalation of pre-drug anticipatory 50-kHz ultrasonic vocalizations in rats. *Behav. Brain Res.* 212, 109–114. doi: 10.1016/J.BBR.2010.04.001

Manduca, A., Campolongo, P., Palmery, M., Vanderschuren, L. J. M. J., Cuomo, V., and Trezza, V. (2014). Social play behavior, ultrasonic vocalizations and their modulation by morphine and amphetamine in Wistar and Sprague-Dawley rats. *Psychopharmacology* 231, 1661–1673. doi: 10.1007/s00213-013-3337-9

Olszyński, K. H., Polowy, R., Małż, M., Boguszewski, P. M., and Filipkowski, R. K. (2020). Playback of alarm and appetitive calls differentially impacts vocal, heart-rate and motor response in rats. *IScience* 23:101577. doi: 10.1016/j.isci.2020.101577

Olszyński, K. H., Polowy, R., Wardak, A. D., Grymanowska, A. W., and Filipkowski, R. K. (2021). Increased vocalization of rats in response to ultrasonic playback as a sign of hypervigilance following fear conditioning. *Brain Sci.* 11:970.

Panksepp, J. (2005). Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* 14, 30–80. doi: 10.1016/J.CONCOG.2004.10.004

Parsana, A. J., Li, N., and Brown, T. H. (2012). Positive and negative ultrasonic social signals elicit opposing firing patterns in rat amygdala. *Behav. Brain Res.* 226, 77–86. doi: 10.1016/J.BBR.2011.08.040

Rippberger, H., Van Gaalen, M. M., Schwarting, R. K. W., and Wöhr, M. (2015). Environmental and pharmacological modulation of amphetamine- induced 50-kHz ultrasonic vocalizations in rats. *Curr. Neuropharmacol.* 13, 220–232.

Robakiewicz, I., Polak, M., Rawska, M., Alberski, D., Polowy, R., Wytrychiewicz, K., et al. (2019). Stimulus-seeking in rats is accompanied by increased c-Fos expression in hippocampal CA1 as well as short 22 kHz and flat 50 kHz calls. *Acta Neurobiol. Exp.* 79, 309–317. doi: 10.21307/ANE-2019-029

Sadananda, M., Wöhr, M., and Schwarting, R. K. W. (2008). Playback of 22-kHz and 50-kHz ultrasonic vocalizations induces differential c-fos expression in rat brain. *Neurosci. Lett.* 435, 17–23. doi: 10.1016/j.neulet.2008.02.002

Sales, G. D. (1972). Ultrasound and aggressive behaviour in rats and other small mammals. *Anim. Behav.* 20, 88–100. doi: 10.1016/S0003-3472(72)80177-5

Schwarting, R. K. W. (2018a). Ultrasonic vocalization in female rats: a comparison among three outbred stocks from pups to adults. *Physiol. Behav.* 196, 59–66. doi: 10.1016/J.PHYSBEH.2018.08.009

Schwarting, R. K. W. (2018b). Ultrasonic vocalization in juvenile and adult male rats: a comparison among stocks. *Physiol. Behav.* 191, 1–11. doi: 10.1016/J.PHYSBEH.2018.03.023

Scull, J., Davies, K., and Amsel, A. (1970). Behavioral contrast and frustration effect in multiple and mixed fixed-interval schedules in the rat. *J. Comp. Physiol. Psychol.* 71, 478–483. doi: 10.1037/H0029160

Seffer, D., Schwarting, R. K. W., and Wöhr, M. (2014). Pro-social ultrasonic communication in rats: insights from playback studies. *J. Neurosci. Methods* 234, 73–81. doi: 10.1016/j.jneumeth.2014.01.023

Seyfarth, R. M., and Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annu. Rev. Psychol.* 54, 145–173. doi: 10.1146/annurev.psych.54.101601.145121

Simmons, S. J., Barker, D. J., and West, M. O. (2018). Ultrasonic vocalizations capture opposing affective states during drug self-administration: revisiting the opponent-process model of addiction. *Handb. Behav. Neurosci.* 25, 389–399. doi: 10.1016/B978-0-12-809600-0.00037-8

Takahashi, N., Kashino, M., and Hironaka, N. (2010). Structure of rat ultrasonic vocalizations and its relevance to behavior. *PLoS One* 5:e14115. doi: 10.1371/journal.pone.0014115

Taylor, J. E., Ficzere, B., St. Louis, J., and Schoenfeld, T. J. (2019). Examining the effects of exercise on frustration-induced anxiety-like behavior in rats. *J. Psychol. Res.* 24, 210–221. doi: 10.24839/2325-7342.JN24.4.210

Tonelli, L. C., Wöhr, M., Schwarting, R. K. W., and Melo-Thomas, L. (2017). Awakenings in rats by ultrasounds: a new animal model for paradoxical kinesia. *Behav. Brain Res.* 337, 204–209. doi: 10.1016/j.bbr.2017.09.021

Willadsen, M., Best, L. M., Wöhr, M., and Clarke, P. B. S. (2018). Effects of anxiogenic drugs on the emission of 22- and 50-kHz ultrasonic vocalizations in adult rats. *Psychopharmacology* 235, 2435–2445. doi: 10.1007/s00213-018-4942-4

Willadsen, M., Seffer, D., Schwarting, R. K. W., and Wöhr, M. (2014). Rodent ultrasonic communication: male prosocial 50-kHz ultrasonic vocalizations elicit social approach behavior in female rats (*Rattus norvegicus*). *J. Comp. Psychol.* 128, 56–64. doi: 10.1037/a0034778

Willuhn, I., Tose, A., Wanat, M. J., Hart, A. S., Hollon, N. G., Phillips, P. E. M., et al. (2014). Phasic dopamine release in the nucleus accumbens in response to pro-social 50 kHz ultrasonic vocalizations in rats. *J. Neurosci.* 34, 10616–10623. doi: 10.1523/JNEUROSCI.1060-14.2014

Wöhr, M. (2018). Ultrasonic communication in rats: appetitive 50-kHz ultrasonic vocalizations as social contact calls. *Behav. Ecol. Sociobiol.* 72:14. doi: 10.1007/s00265-017-2427-9

Wöhr, M. (2021). Measuring mania-like elevated mood through amphetamine-induced 50-kHz ultrasonic vocalizations in rats. *Br. J. Pharmacol.* [Epub ahead of print]. doi: 10.1111/bph.15487

Wöhr, M., Houx, B., Schwarting, R. K. W., and Spruijt, B. (2008). Effects of experience and context on 50-kHz vocalizations in rats. *Physiol. Behav.* 93, 766–776. doi: 10.1016/J.PHYSBEH.2007.11.031

Wöhr, M., and Schwarting, R. K. W. (2007). Ultrasonic communication in rats: can playback of 50-kHz calls induce approach behavior? *PLoS One* 2:e1365. doi: 10.1371/journal.pone.0001365

Wöhr, M., and Schwarting, R. K. W. (2009). Ultrasonic communication in rats: effects of morphine and naloxone on vocal and behavioral responses to playback of 50-kHz vocalizations. *Pharmacol. Biochem. Behav.* 94, 285–295. doi: 10.1016/j.pbb.2009.09.008

Wöhr, M., and Schwarting, R. K. (2012). Testing social acoustic memory in rats: effects of stimulus configuration and long-term memory on the induction of social approach behavior by appetitive 50-kHz ultrasonic vocalizations. *Neurobiol. Learn. Memory* 98, 154–164.

Wöhr, M., and Schwarting, R. K. W. (2013). Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation. *Cell Tissue Res.* 354, 81–97. doi: 10.1007/s00441-013-1607-9

Wöhr, M., Seffer, D., and Schwarting, R. K. W. (2016). Studying socio-affective communication in rats through playback of ultrasonic vocalizations. *Curr. Protoc. Neurosci.* 75, 8.35.1–8.35.17.

Wöhr, M., Willadsen, M., Kisko, T. M., Schwarting, R. K. W., and Fendt, M. (2020). Sex-dependent effects of Cacna1c haploinsufficiency on behavioral inhibition evoked by conspecific alarm signals in rats. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 99:109849. doi: 10.1016/j.pnpbp.2019.109849

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.

The handling editor declared a past co-authorship with one of the authors MW.

frontiers in Behavioral Neuroscience

TECHNOLOGY REPORT
published: 25 January 2022
doi: 10.3389/fnbeh.2021.810590

HybridMouse: A Hybrid Convolutional-Recurrent Neural Network-Based Model for Identification of Mouse Ultrasonic Vocalizations

Yizhaq Goussha[1], Kfir Bar[2], Shai Netser[1], Lior Cohen[1], Yacov Hel-Or[2] and Shlomo Wagner[1]*

[1] Sagol Department of Neurobiology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel, [2] School of Computer Science, The Interdisciplinary Center, Herzliya, Israel

OPEN ACCESS

Edited by:
Susanna Pietropaolo, UMR5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Aaron M. Johnson, NYU Grossman School of Medicine, United States
David James Barker, Rutgers, The State University of New Jersey, United States

***Correspondence:**
Yizhaq Goussha
gutzcha@gmail.com

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *07 November 2021*
Accepted: *16 December 2021*
Published: *25 January 2022*

Citation:
Goussha Y, Bar K, Netser S, Cohen L, Hel-Or Y and Wagner S (2022) HybridMouse: A Hybrid Convolutional-Recurrent Neural Network-Based Model for Identification of Mouse Ultrasonic Vocalizations. Front. Behav. Neurosci. 15:810590. doi: 10.3389/fnbeh.2021.810590

Mice use ultrasonic vocalizations (USVs) to convey a variety of socially relevant information. These vocalizations are affected by the sex, age, strain, and emotional state of the emitter and can thus be used to characterize it. Current tools used to detect and analyze murine USVs rely on user input and image processing algorithms to identify USVs, therefore requiring ideal recording environments. More recent tools which utilize convolutional neural networks models to identify vocalization segments perform well above the latter but do not exploit the sequential structure of audio vocalizations. On the other hand, human voice recognition models were made explicitly for audio processing; they incorporate the advantages of CNN models in recurrent models that allow them to capture the sequential nature of the audio. Here we describe the HybridMouse software: an audio analysis tool that combines convolutional (CNN) and recurrent (RNN) neural networks for automatically identifying, labeling, and extracting recorded USVs. Following training on manually labeled audio files recorded in various experimental conditions, HybridMouse outperformed the most commonly used benchmark model utilizing deep-learning tools in accuracy and precision. Moreover, it does not require user input and produces reliable detection and analysis of USVs recorded under harsh experimental conditions. We suggest that HybrideMouse will enhance the analysis of murine USVs and facilitate their use in scientific research.

Keywords: animal communication, social interactions, ultrasonic vocalizations, neural networks, machine learning, CNN–convolutional neural networks, LSTM–long short-term memory

INTRODUCTION

Many vertebrates use Species-specific vocal communications for social interactions (Todt and Naguib, 2000; Wilkins et al., 2013; Chen and Wiens, 2020). Specifically, mice and rats use ultrasonic vocalizations (USVs) to convey a variety of information types: identification of the emitter and its group, its status within the group and the status of the group, the environmental conditions, such as the presence of predators, and whereabouts of food and water (Lahvis et al., 2011). The emitter and the receiver of the vocal cue can be found at various types of social engagements. For example, pups emit either wriggling calls (WC) to gain their parents' attention or isolation calls following a

drop in their body temperature when kept away from their mothers (Haack et al., 1983). An adult male may vocalize to ward off other male intruders or peruse a female in courtship by singing to her (i.e., mating calls, MC), while a female may emit vocalizations to answer her male suiter or to call other females (Neunuebel et al., 2015).

Current tools used to detect and analyze murine USVs rely on user input and classical image processing algorithms to identify and clean USVs, thus requiring manual adjustments and ideal recording environments. More recent tools utilize machine learning and convolutional neural networks (CNN) models (Van Segbroeck et al., 2017; Fonseca et al., 2021) to identify vocalization segments. For example, DeepSqueak (DS), a recent benchmark tool for detecting USVs, relies on neural networks to detect USVs (Coffey et al., 2019). It implements an object detection architecture, namely: regional convolutional neural networks (Faster-RCNN) (Ren et al., 2017). CNN-based models such as DeepSqueak are powerful tools for image processing due to their capability to take advantage of the local spatial coherence of images. In the case of vocalizations, the images are 2D representations of the audio signal transformed to the frequency domain. However, CNN models do not exploit the temporal correlations of audio signals and underperform under noisy recording conditions. On the other hand, recurrent neural networks (RNNs) can compensate for these weaknesses by capturing long contextual dependencies, using prior knowledge, thus offering better performance.

In this study, we describe a USV Extractor software (termed HybridMouse), an audio analysis tool that allows labeling of murine ultrasonic vocalizations both manually or automatically. The automatic detection of USVs is performed by the HybridMouse model, which combines convolutional (CNN) and Bidirectional Long-Short Term Memory (BiLSTM) recurrent neural networks to identify USVs, label them, and extract their denoised representations.

METHODS

Training a Hybrid CNN-BiLSTM Model to Detect USVs

Animals

Mice from three distinct strains: BALBc, C57BL/6J, and CD-1 (ICR) were purchased from Envigo (Rehovot, Israel). The mice were kept in clean plastic chambers (GM500, Tecniplast, Italy) at 22°C and a 12-h light/12-h dark cycle (light on at 7 am) and received food and water *ad libitum*. All cages contained standard wood chip bedding, cotton wool bedding material, and a mock nest of a 12 cm plastic tube. All the recorded adult mice were 3–5 months old, and all the recorded pups were healthy, 3–5 days old pups.

Recording the Training Data

Adult vocal communications were recorded using a 1/4 inch microphone (Type 4939-A-011), connected to a preamplifier (Type 2670), and an amplifier (Type 2690-0S1, Bruel and Kjaer) in a custom-built sound-shielded box. Vocalizations were sampled at 250 kHz with a CED Micro 1401-3 recording device (Cambridge Electronic Design Limited, Sunnyvale, CA). The mice were placed in their home cage or a fresh cage during recording sessions and received food and water *ad libitum*. The lid was removed and replaced with a metallic grid to allow passage of sound and prevent escape, while the microphone was placed hovering just above the cage. The system records 10 min every hour for 12 h, during the 1st, 2nd, and 8th nights. In each strain, we recorded four types of pairs: male-male (both naive), female-female (both naive), male-female (both naive), experienced female (dam following weening)-naïve female. In total, we recorded $S \times G \times P \times D \times H \times M = 216$ h. When S is the number of strains (3), G is the number of social groups (4), P is the number of pairs in each group (3), D is the number of recorded days (3), H is the number of hours recorded each night (12), and M is the number of minutes sampled every hour (10).

Pup calls were recorded by placing them in a 200 ml cardboard cup and recording for 5 min each. In total, we recorded ten pups per strain from at least two different litters for a total of $S \times P \times M = 2.5$ h, where S is the number of strains (3), P is the number of pups in each group (10), and M is the number of minutes recorded for each pup (5). Even though this group's recorded hours are low, we extracted ∼4k pups isolation USVs. This is because isolated pups emit vocalizations rigorously as their temperature drops almost throughout the entire (5 min) recordings, resulting in a high number of USVs per recording, while other groups tend to vocalize fewer hours. As a comparison, we also recorded interactions between naïve male dyads, and we did not use these recordings in our training because the number of USVs emitted was too low.

Dealing With Imbalanced Data

Only 0.44% of the data contained USVs (57 min of 216 h recorded), making the data highly imbalanced and unsuitable for a classification task. To overcome this, we trained the data only on short (1–20 s) audio segments containing consecutive USVs with short (<1 s) intervals between them, which effectively condensed our training data from ∼200 h to ∼2 h with more balanced training examples (45% USV and 55% background).

Noise Samples

Condensing the data meant that we trained the model only on the background noise between the USVs, which posed a new problem; it was insufficient to adequately train the model, resulting in a high rate of false-positive events. To overcome this problem, we randomly sampled noise segments from our recordings and added them to the training data. We also added background recordings made in other labs to diversify noise patterns and thus generalize the model. In total, we added 2600 files (1-s long each) without USVs, for a total of ∼43 min.

To overcome the imbalance reintroduced to our data due to adding noise samples, we used a weighted classifier in our training to represent the imbalance in the data between the two classes (USV 25%, background 75%).

Manual Labeling

A trained team of analytics manually identified each USV by screening spectrograms of all the recorded data using a custom-made Matlab script. When labeling the USVs, our main priority was to extract temporal features (such as syllable duration, call duration, syllable rate per sentence and second, etc.). For that reason and to save time and effort, we only extracted the beginning and ending timestamps of each USV, which means that the minimal and maximal frequencies for each USV were not extracted.

Evaluation

Evaluation Metrics

Model evaluation was made by comparing the model predictions to manually labeled data and extracting recall, precision, and *F*1 scores, as follows:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$F1 = 2^* \frac{Presicion^* Recall}{Presicion + Recall}$$

Where

True-positive—The number of correctly identified USVs

False-positive—The number of noise segments mislabeled as USVs

False-negative—The number of USVs the were not detected.

Local and Global Signal to Noise Ratio Calculations

The magnitude of an audio segment was calculated as the mean sum of squares of the raw audio values. To calculate the local signal-to-noise ratio (SNR), we first calculated $M_{sig+bg,i}$, the magnitude of the ith audio segment containing USV. We then calculated $M_{bg,i}$, the mean magnitude of the audio segments preceding and following the i^{th} USV segment (that do not contain USVs). Then we calculate R_i, the local SNR of the i^{th} USV, as follows:

$$R_i = \frac{M_{sig+bg,i} - M_{bg,i}}{M_{bg,i}}$$

Where

R_i–The local SNR of USV segment *i*

$M_{sig+bg,i}$–The magnitude of the audio segment containing USV segment *i*

$M_{bg,i}$–The magnitude of the audio segments preceding and following USV segment i.

To calculate the global SNR of the entire audio file, we calculated $M_{sig+bg,}$ the magnitude of all segments containing USVs combined, and M_{bg}, the magnitude of all the segments not containing USVs combined, and repeated the calculation above.

Testing Data and Labeling

The analysis was performed on audio files downloaded from the mouseTube database (https://mousetube.pasteur.fr/). The files were recorded by two laboratories during various social tasks to ensure model generalization. Files from the first laboratory (Duke University Medical Center, Social context comparisons) included recordings of adult male B6D2F1/J mice with either awake or anesthetized male or female conspecifics on different menstrual states or only with urine from males or females. Files from the second laboratory (Washington University in St. Louis, Pup USV Day 3–14 Monitoring) included recordings of C57BL/6J pups (p3–p14). All recordings were made using Avisoft CM16/CMPA microphones at a 250 kHz sampling rate (for more details, see **Supplementary Table 1**). USVs detected by the different models were labeled as "USV." Subsequent tagging of the same USVs by a model were labeled "Partial," and single taggings encompassing multiple USVs were labeled "Multi." Each of these labels was counted only once. In addition, tagged alarm calls were labeled as "AC" and excluded from further analysis.

Comparison With DeepSqueak

We analyzed the same test files using DeepSqueak (Coffey et al., 2019), employing its default network for mouse USVs. We performed the analysis with two different settings; balanced (DS_B) and high recall (DS_H); all other settings were kept unchanged in their default settings: Total analysis length = 0 (full duration), Analysis chunk length = 3 s, Overlap = 0.1 s, frequency cut off high = 120 kHz, frequency cut off low = 1 kHz, score threshold = 0. The analysis was performed using the built-in "Mouse Call_Network_V2" network.

Statistical Analysis

Model performances were compared using the Mann-Whitney U test due to the non-normality of the variables. The analysis was performed using SPSS version 27 software.

Our Model

Model Architecture and Syllable Isolation

As mentioned in the methods (see section Manual Labeling), we labeled the recorded data by manually extracting the start and end time points of each USV and not their frequency boundaries. Therefore, the model output contained only temporal data (each USV's start and end time points) and no spectral data (such as USV frequency and structure). However, our goal was to create a model capable of not only identifying the time steps containing USVs but also of estimating their frequencies and segmenting them from the background noise. To achieve this, the model processes the 124 by 51 feature maps in two phases; CNN and BiLSTM. The output of the CNN phase is a 1×124 vector estimating the probability of each frequency to be a part of a USV. Concatenating these vectors along with the temporal domain into a 2D map allowed us to use it as a mask to clean the raw spectrograms (**Figure 2B**, Classification and mask). In the second phase, the BiLSTM layer classifies each vector (representing each time point) and estimates whether this vector is part of a USV or not.

CNN Layers

Each time bin is fed into the model along with its neighboring 25-time steps. We experimentally found this number of neighboring steps to be the minimum number that yields an optimal performance of the model. While models with fewer neighbors

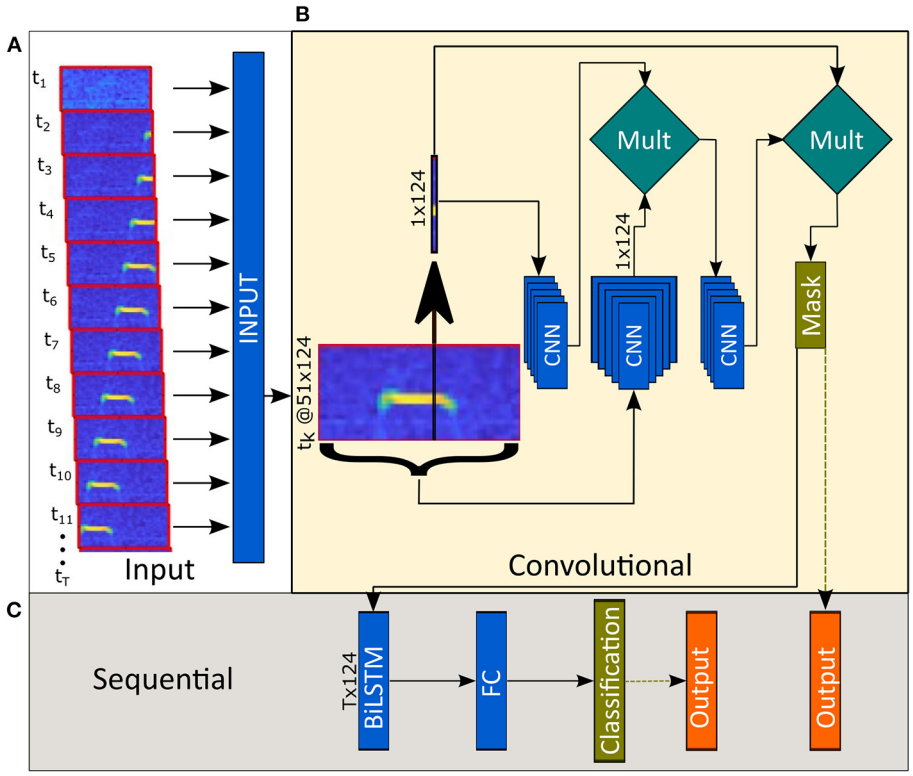


FIGURE 1 | Model architecture. **(A)** Sequential data (124 by 51 Mel-spectrograms per time step) are fed into the model. **(B)** The first phase consists of convolutional layers and outputs a mask defining the estimated frequency (or the lack of) of USVs. **(C)** The second half of the model combines data from all time points and classifies each time step.

resulted in inadequate predictive capabilities, models with a higher number of neighbors did not improve the model enough to justify increasing complexity. Inputting each time bin with neighboring 25-time steps creates a form of self-attention mechanism (Vaswani et al., 2017; Bello et al., 2019); a 1D convolution is performed on the mid-time bin to accentuate potential USVs, and a 2D convolution is performed on the entire feature map to suppress global noises (**Figure 1**). By multiplying the outputs of the two branches, we got a clean 1D representation of each time-bin. Finally, we multiplied the input by the clean 1D vectors before passing it to the sequential layers to force it to act as a filter. The output of this phase is used again in the post-processing stage to produce clean syllables **Figure 2C** (see section Pre-processing and Augmentation).

Sequential Layers

The extracted vectors from the CNN layers are first reshaped, i.e., the vector is transformed from a 124×1×1 vector (Height x Width x Channel) into a 1×124 (1 × Channel) vector. They are then fed into the sequential layers, comprised of BiLSTM layers followed by fully connected (FC) layers. Finally, a weighted softmax-cross entropy algorithm was implemented in the classification layer to output a categorical 1D vector where each time step was labeled either "0" or "1" for "background" and "USV," respectively. It is also possible to output a numeric 2D probability vector with values between 0-1 for each label ("background" or "USV") at each time step. The output is the model's estimated probability of each time step to be part of a USV (**Figure 1C**). The model was trained using 1-s long audio segments (332-time steps). However, it is possible to feed longer or shorter longer segments to train the model or to extract USVs.

Training Data

Training data was acquired by recording vocalizations emitted during dyadic interactions between mice of various strains, ages, and during various social contexts. In total, we recorded over 200 h of mouse vocal communications representing over 1,200 h of interactions. We then meticulously combed the audio recordings and extracted over 40k USVs to be used as training data (See methods for details).

The training data included two types of audio clips; (1) files containing "calls" (a series of USV syllables) and (2) files containing only background with no calls. Most of the background samples were extracted from the recorded training data, but some samples were extracted from recordings made by other teams and laboratories in order to diversify the noise parameters and make the model suitable for general recording conditions.

It is important to note that the model was tested and compared to DeepSqueak on an entirely new set of files downloaded from

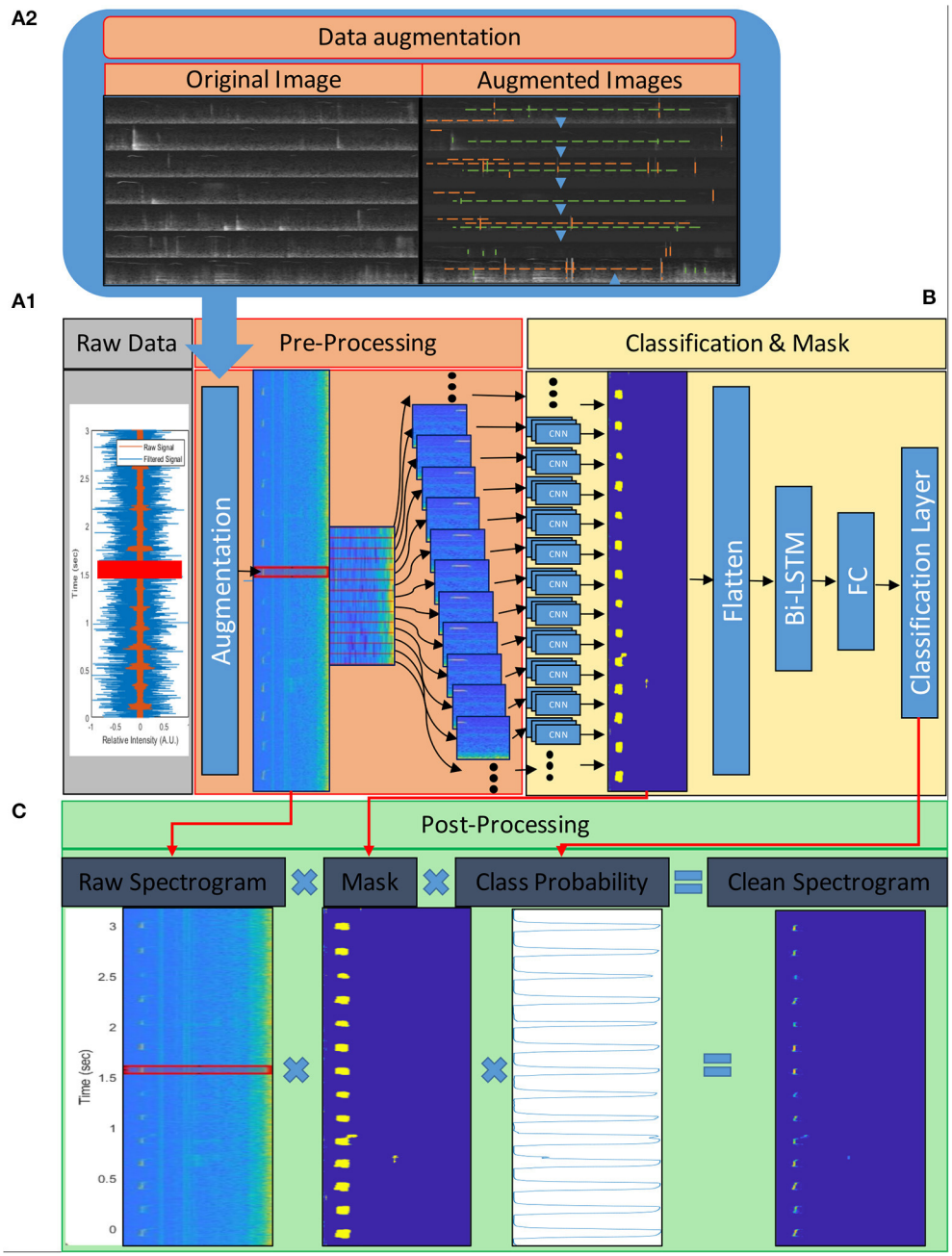


FIGURE 2 | USV processing flow chart. The vocalizations are recorded at a high sampling rate *(>250 samples per sec)* and, after scaling, undergo processing, which can be divided into three phases; **(A)** pre-processing, **(B)** syllable detection, and **(C)** post-processing. **(A)** First, the signal is converted to the frequency domain by extracting Mel-frequency cepstrum (MFC) features; the spectrograms indicate frequency contours and relative intensities of the frequency components. The extracted frequencies are binned in 124 frequency bands (range 1–125 kHz) and 332-time steps per second (∼3 ms per time step); **(A.1)** These frequency maps are augmented during training to create a more generalized model; the insert **(A2)** shows spectrograms of training images before (Original) and after (Augmented) augmentation. Blue arrows indicate the vertical translation direction of the images; green dashed lines indicate existing noises' location; orange dashed lines indicate randomly added lines to simulate humming (horizontal lines) and impulse (vertical lines) noises. We also scaled the values of the images in the training data but did not do so in this example to make it easier to plot. Each time step is fed into the classification model along with its preceding and following 25-time bins. **(B)** The 124 by 51 feature maps are processed by the model in two phases; CNN and BiLSTM. The CNN layers filter the feature maps and produce 124 by 1 column masks for each time step. These masks indicate the presence (or lack of) and frequency of USVs in their respective time steps. The masks are flattened and fed into the BiLSTM layers and to fully connected layers (FC) to classify each time step as a "1" or "0," indicating whether they contain/are part of a USV or not, respectively. The output of these layers is also used in **(C)** the post-processing phase to produce a cleaned signal. The cleaning is done by multiplying the original spectrogram with the mask and the probability vector produced by the classification model.

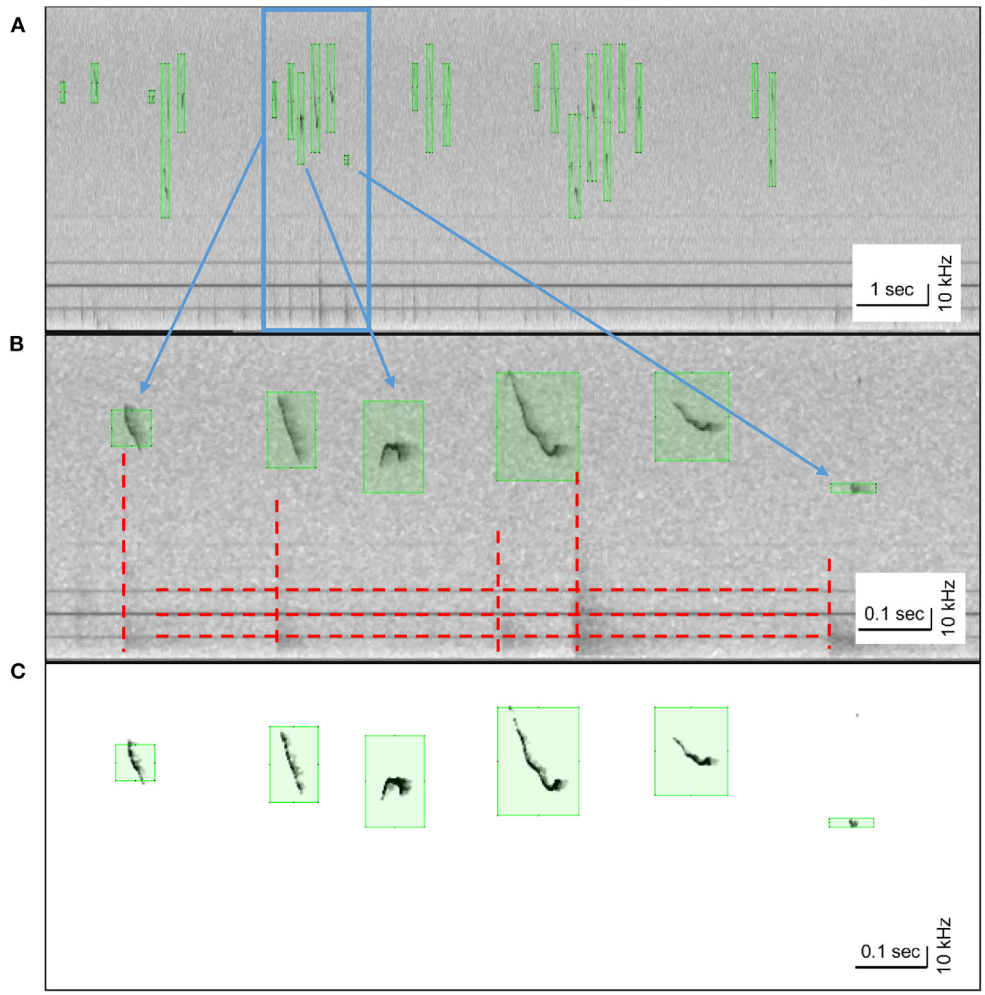


FIGURE 3 | USV detection and denoising. Spectrograms (linear) of example USVs detected by HM model. **(A)** Detected USVs over a 10-s audio file; green squares indicate the estimated frequency and time boundaries of each USV. **(B)** A close-up of the highlighted blue area in **(A)** containing both types of noises, humming, and impulse noises, denoted by vertical and horizontal red dashed lines (respectively). **(C)** The denoised version of the USVs shown in **(B)**, processed by HybrideMouse; all background noises were removed, and the USVs were isolated.

mouseTube. The recordings that were used to train the model, both the ones recorded at our lab and those collected from others, were not used to test the model. The files used for testing were recorded under different conditions at different laboratories.

Retraining and Fine-Tuning

We added the option to retrain the model. This can be performed by manually labeling or automatically labeling (and manually adjusting) samples of recordings and retraining the model to create new, customized models.

Pre-processing and Augmentation

For pre-processing of audio clips (**Figure 2A**), the signal was converted to the frequency domain by extracting Mel-frequency cepstrum (MFC) features. The extracted frequencies were binned into 124 frequency bands (range 1–125 kHz) and 332-time steps per second (3 ms per time step). Each time step was fed into the classification model along with its preceding and following 25-time steps (bins) as 124 by 51 images for each time step (**Figure 2A**). We also converted the labels' timetables into the corresponding one-dimensional categorical vectors. Each label vector had the number of time steps as the corresponding feature map, and each time step that contained a part of a USVs was labeled "1," and the rest that did not contain parts of a USV were labeled"0."

We have used a similar pre-processing of audio clips for training and testing, aside from the following augmentations. During training, the images were augmented in three ways. USVs vary markedly, depending on the emitter's strain, genotype, age, sex, socio-emotional state, etc. Although our training data included a great variety, it did not include all types of mouse USVs. Therefore, we randomly shifted the feature maps along the feature-axis (frequency axis) in order to train the model to recognize new vocalizations emitted at different frequency bands

TABLE 1 | Comparing temporal-spectral estimations.

	HybridMouse	DeepSqueak balanced recall	DeepSqueak high recall
Total number of USVs	10,327	9,175	10,091
Duration (Mdn, sec)	0.052	0.072	0.0512
FrLow (Mdn, kHz)	50.6	47.7	50.1
FrHigh (Mdn, kHz)	73	78.8	71.4
FrLow (<20 kHz) (%)	0.3	8.8	10.7
FrHigh (<20 kHz) (%)	0	1.2	4.3

Duration—the estimated median length of the USVs in seconds, FrLow and FrHigh—the estimated median lower and upper (respectively) frequency boundary of the USV in kHz, FrLow (<20 kHz), and FrHigh (<20 kHz)—the percent of USVs that were estimated to have lower or upper (respectively) frequency boundaries lower than 20 kHz.

while preserving their basic structure. In addition, the scale of the signal depends on the acquisition system and method and, ideally, should be normalized (zero-centered and with a range between −1 and 1). Signal normalization can be performed by subtracting the mean from the signal and dividing it by the maximum value. However, since the audio signal is fed in short segments, it was challenging to shift and scale it that way. In addition, normalizing the feature maps was also not ideal due to the non-linear nature of the Mel-spectrogram transformation. To overcome these difficulties, we randomly scaled (added values [−3, +3 a.u]) and shifted (vertical shift [−50,+10 a.u]) the feature maps during training and trained the model to be less sensitive to the input scale and center. Lastly, horizontal and vertical lines were randomly added to simulate continuous humming and impulse noises, respectively (**Figure 2A.1**). We also tried other, more traditional augmentation methods, such as adding Gaussian or pepper-and-salt noise to the images. In these cases, however, when the augmentation levels were high, the model accuracy declined, and when the levels were low, the augmentations didn't have an effect; therefore, we did not use these augmentations.

Post-processing

The last step of USVs extraction took place after the model prediction. The output of the convolutional layer was used as a mask and multiplied with the raw spectrogram and the probability vector to extract clean syllables without background (**Figure 2B**).

The HybridMouse Software

We integrated the HybridMouse model in a user-friendly Matlab app (**Supplementary Figure 1**). This app was designed to facilitate comparisons between recordings from up to two microphones. Our original paradigm includes recording from one main microphone (Main Mic) and an additional microphone (Mini Mic).

RESULTS

Syllable Isolation and Denoising

HybridMouse model extracts the timestamps of each USV, its frequency, and a clean denoised representation. These capabilities were tested on 58 audio files (180–300 s total length each) downloaded from the Mousetube database and recorded under different conditions and settings. We found that the model excels in estimating the onsets and endings of the syllables (**Figures 3A,B**). We also found that the model manages to produce denoised representations of the USVs (**Figure 3C**), which can be used to extract additional features such as USVs' frequency boundaries and more.

Comparing Performance of HybridMouse to DeepSqueak

We compared the performance of HybridMouse to DeepSqueak, a benchmark model for rodent USV detection (Coffey et al., 2019). To that end, we analyzed the same files using DeepSqueak with two different settings; balanced recall (DS_B) and high recall (DS_H) (**Table 1**). We found that most of the USVs were detected by both models with high accuracy, and, naturally, the detection capabilities declined in noisier sections. Yet, in most cases, the detection capabilities of HybridMouse were higher than DeepSqueak in both settings (**Figure 4A**).

Using all models, we estimated the detected USVs duration as well as the lower (LowFr) and higher (HighFr)-frequency boundaries. We found that the estimations of frequency boundaries (low and high) and USV lengths were similar between all models. However, the frequencies estimated by DeepSqueak contained more USVs with frequencies lower than 20 kHz (**Table 1**; **Figure 4B**; **Supplementary Figure 2**). These USVs could have been USVs that DeepSqueak was able to detect but could not accurately estimate their frequencies or low-frequency noise segments that it mislabeled as USVs. These results suggest that DeepSqueak has lower precision due to a larger number of false-positive events, especially in high-recall settings, compared to HybrideMouse.

On the other hand, using a lower threshold enabled DeepSqueak, when set on high-recall, to capture the duration of the USVs more reliably. The median value estimated by this setting is more similar to the value that HybridMouse estimated. This was in contrast to when DeepSqueak was set to a balanced-recall setting, which only registered the longer USVs and ignored the shorter ones. Notably, HybridMouse suffers from low accuracy when estimating the higher frequency boundaries due to the spectral map's representation as a Mel-spectrogram with higher resolution in the lower frequencies than the higher ones.

To quantify the detection capabilities of HybridMouse and statistically compare it to DeepSqueak, we manually labeled 13 arbitrarily selected audio files and extracted each model's precision, recall, and *F*1 score. Of these 13 files, four contained very few USVs (∼2) and were excluded from the analysis (**Table 2**; **Supplementary Tables 1–3**). As expected, running DeepSqueak on a high-recall setting missed fewer USVs, thus achieving higher recall than a balanced-recall setting; however,

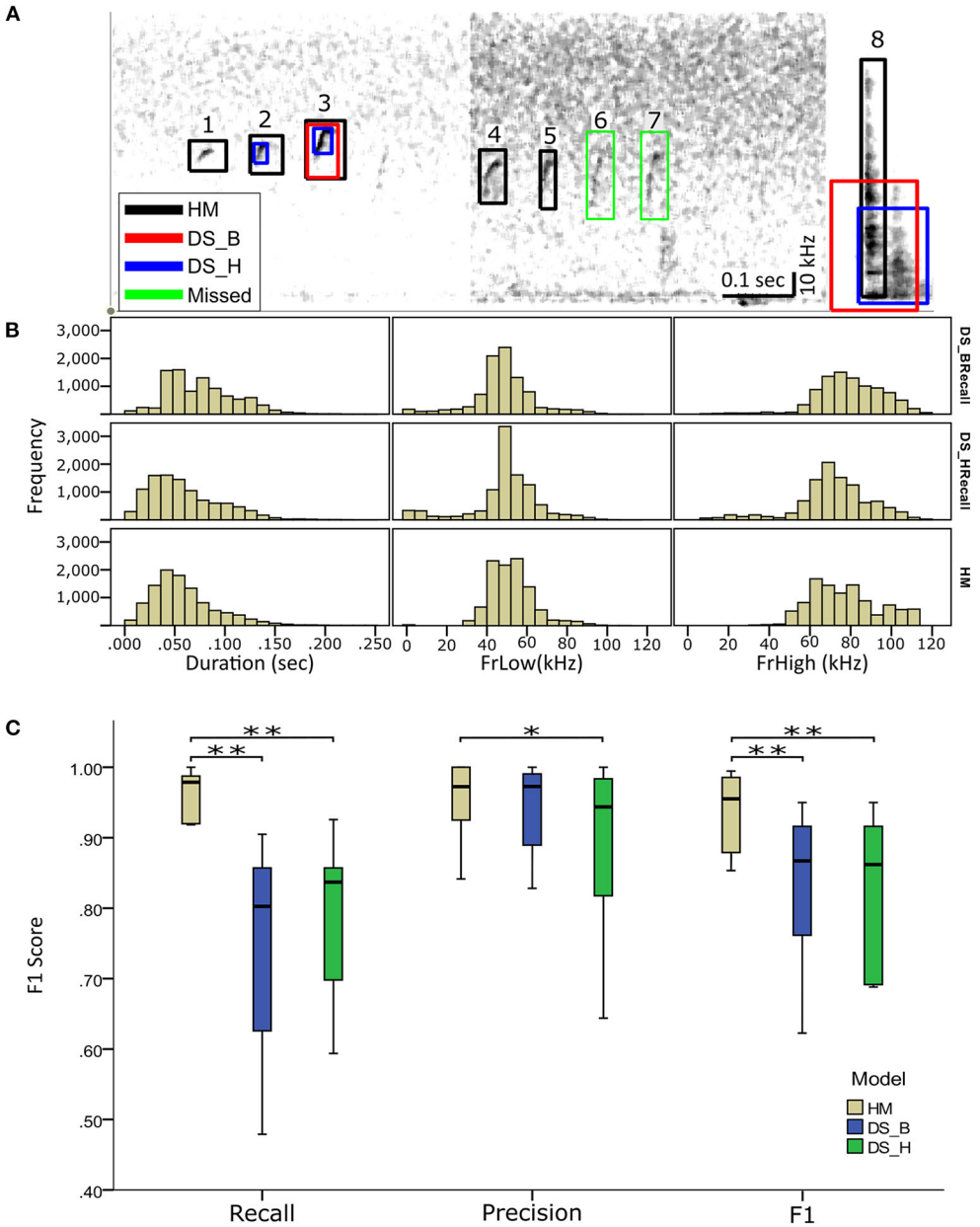


FIGURE 4 | Comparing detection of USVs by HybridMouse to DeepSqueak. **(A)** Visual comparison between the models and examples of their detection performance, showing spectrograms (linear) of audio segments with the detected USV markings by all models; HM (black), DS_H (blue), and DS_B (red) and the missed USVs (green). Shown are USVs detected by all models and settings (3), USVs detected only by HM and DS in high-recall settings (2), USVs detected only by HM (1, 4, and 5). We also see USVs missed by all models (6 and 7) and finally, a noise segment that all models mislabeled as a USV (8). **(B)** Comparing temporal-spectral estimations; Audio files were analyzed using HM and DeepSqueak with balanced recall (DS_B) and high recall (DS_H) settings. Each detected USV's duration, lower-frequency (FrLow), and upper-frequency (FrHigh) boundaries were estimated and compared. **(C)** Comparing the performance of the HybridMouse model (HM–blue) to the DeepSqueak model with two different settings; balanced recall (DS_B–green) and high recall (DS_H–yellow). Wilcoxon signed-rank $^{*}p < 0.05$, $^{**}p < 0.01$. HybridMouse scored significantly higher in recall (Mdn = 0.979) than DeepSqeak on balanced recall setting (Mdn = 0.802) (Wilcoxon signed-rank $z = -2.666$, $p < 0.01$), and on high recall setting (Mdn = 0.837) (Wilcoxon signed-rank $z = -2.666$, $p < 0.01$). All models performed rather similarly in the precision, with DS_H receiving the lowest score. HybridMouse achieved significantly higher $F1$ score (Mdn = 0.955) than DS_B (Mdn = 0.876) (Wilcoxon signed-rank $z = -2.66$, $p < 0.01$), and DS_H (Mdn = 0.862) (Wilcoxon signed-rank $z = -2.66$, $p < 0.01$).

this comes with the price of mislabeling more noise fragments as USVs and performing poorly on the precision metric. On the other hand, HybridMouse missed fewer USVs and mislabeled fewer noise fragments as USVs than both DeepSqueak settings and in all the tested files. HybridMouse scored significantly higher in recall (Mdn = 0.979) than DeepSqeak on balanced

TABLE 2 | Comparing detection of USVs by HybridMouse to DeepSqueak.

	HybridMouse	DeepSqueak balanced recall	DeepSqueak high recall
Recall	0.98	0.80	0.84
Precision	0.97	0.97	0.94
*F*1	0.96	0.87	0.86

Comparing median recall, precision, and F1 scores of the HybridMouse model to the DeepSqueak model with two different settings; balanced recall and high recall.

recall setting (Mdn = 0.802) (Wilcoxon signed-rank z = −2.666, $p < 0.01$), and on high recall setting (Mdn = 0.837) (Wilcoxon signed-rank z = −2.666, $p < 0.01$). All models performed similarly in the precision, with DS_H receiving the lowest score. However, the difference in recall was reflected in the final *F*1 score with HybridMouse achieving significantly higher score (Mdn = 0.955) than DS_B (Mdn = 0.876) (Wilcoxon signed-rank z = −2.66, $p < 0.01$), and DS_H (Mdn = 0.862) (Wilcoxon signed-rank z = −2.66, $p < 0.01$). Thus, HybridMouse outperformed DeepSqueak (**Figure 4C**).

Lastly, to show that our model is more robust to low SNR, we calculated the global SNR of each tested audio file and found that HybridMouse performs well even in low SNR (**Figures 5A–C**). This is further demonstrated by the example of one of the tested files (interaction between male and female, 300 s duration shown in **Figures 5D–I**), showing that below 0.1 local SNR, DeepSqueak struggles to detect the USVs, resulting in a high number of false-negative events. Though lowering the detection threshold (high recall) improves the detection (fewer false-negative events), it comes at the cost of raising the number of false-positive events. In addition, we compared the power spectral densities of the detected USVs and the missed USVs by HybridMouse (**Supplementary Figure 3**) and again showed that the missed USVs have low amplitude and low SNR.

DISCUSSION

Like many vertebrates, mice convey various information through vocal communication (Todt and Naguib, 2000; Wilkins et al., 2013; Chen and Wiens, 2020). For these reasons, analyzing social vocalizations can progress our understanding of animal social behaviors and the underlining mechanisms governing these behaviors. Studies in this field rely heavily on observations and subjective assessments based on visual cues, such as freezing and body pose. However, these studies tend to ignore the vocal cues, which can be paramount to understanding and accurately assessing the socio-emotional state of the animals.

One of the reasons this field is ignored is that most murine vocalizations are ultrasonic and inaudible to the human ear and can also be very short and sparse. Another reason is the shortage of tools that may be used for automatic high-throughput analysis of such vocalizations.

With the recent explosion of computerized algorithms and the rise of computational power, multiple tools have been developed to analyze animal vocal communication, such as VoICE (Burkett et al., 2015), MUPET (Van Segbroeck et al., 2017), which rely on classical signal and image processing methods. More recent models such as USVSEG (Tachibana et al., 2020) and Acoustilytix (Burkett et al., 2015) rely on machine learning and deep learning tools to extract useful information from the recording of animal vocalizations. One of the most popular and publically-available models for rodent USV analysis is DeepSqueak, a benchmark in this field that outperforms its predecessors (Coffey et al., 2019). This model relies on deep neural networks to extract USVs from audio recordings. However, DeepSqueak relies only on convolutional neural networks. These CNN-based object detection models such as YOLO (Redmon et al., 2016) and Faster R-CNN (on which DeepSquak is based) perform well under ideal recording conditions; however, they are not suitable for analyzing recordings with a low signal-to-noise ratio. This is because they are made explicitly for image analysis and are less optimized for auditory data, which is represented by a sequence of images rather than a single image; thus, these models do not allow capturing the syllables within their context. Other models based on recurrent neural network-based models are more suitable for analyzing data that have sequential nature.

In this study, we presented HybridMouse, a hybrid CNN-BiLSTM based model for mouse USV extraction. This hybrid model incorporates both CNN and recurrent neural networks (BiLSTM), thus exploiting both the auditory signal's structure and its context.

We decided to compare HybridMouse to DeepSqueak, which is considered a benchmark model for analyzing murine USVs; it is well-known and free, unlike other available tools. DeepSqueak comes with four different pre-trained models, and the model chosen for comparison was trained explicitly for mouse USVs, with two different settings: balanced recall (DS_B) and high recall (DS_H). All other parameters were kept at default settings.

The HybridMouse model was tested on novel data recorded in different labs on which it was not trained and was found to be a robust tool for automatically identifying and isolating USVs even in harsh recording conditions with low SNR. We found that HybridMouse manages to detect almost all of the manually identified vocalizations, resulting in high recall without increasing the rate of false-positive events, thus demonstrating high precision. We compared HybridMouse to DeepSqueak in a balanced setting. We found that although the precision of both models was similar, DS_B identifies fewer USVs, resulting in a significantly lower recall and an overall lower *F*1 score. We also compared HybridMouse to DeepSqueak in a high-recall setting. This comparison showed that HybridMouse still managed to identify a higher number of USVs and achieved a significantly higher recall score even in this setting. Lowering the detection threshold also increased the number of false-positive events, resulting in significantly lower precision and overall *F*1 score than HybridMouse.

We further investigated the performance of the models by calculating the global SNR of each tested file and showing that HybridMouse is more robust to lower SNR than DeepSqueak. While the latter performed well and was comparable to the former on files with high SNR, as the levels of SNR dropped, so did the performance of DeepSqueak. Furthermore, pup

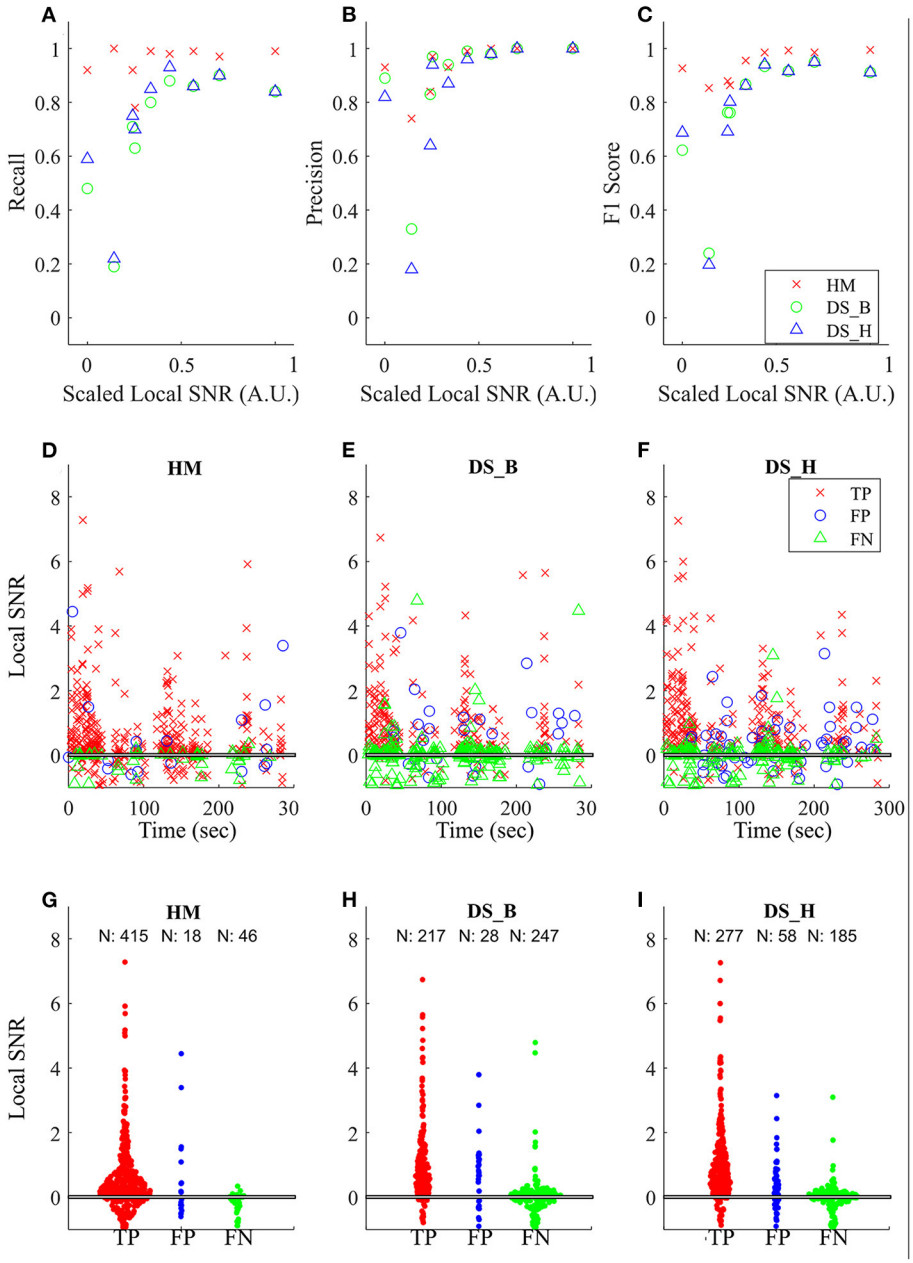


FIGURE 5 | Comparing the effects of SNR on detection. Comparing the effects of SNR on detection capabilities of the model using the three parameters **(A)** recall, **(B)** precision, and **(C)** *F*1 score. The relative SNRs were calculated as described in the methods and rescaled [0,1]. The performance of HybridMouse (HM) remains relatively high. In contrast, the performance of DeepSqueak in both settings (balanced and high recall) (DS_B and DS_H, respectively) drops in all parameters when the relative SNR is low. The middle and lower panels show an example of one of the tested files (interaction between male and female, 300 s duration). The black horizontal line (overlapping with the x-axis) indicates the point where the magnitude of the signal is equal to that of its surroundings. The middle panels **(D–F)** show the detected USVs (TP), missed USVs (FN), and segments mislabelled as USVs (FN) by the models throughout the interaction. And the lower panels **(G–I)** show swarm plots of these parameters and the number of detections in each category.

vocalizations are recorded in a particular setting where the microphone is closer to the recorded pup, and there are very few background noises. Therefore, it was no surprise that these recordings had high SNR and high detection rates by all models.

The high time-resolution and robustness of HybridMouse were achieved by two means. The first one was through the data augmentation methods that we applied to the training data. We added random horizontal and vertical lines, simulating consistent humming and burst noises, respectively. The data augmentation

has significantly improved the model by lowering its false-positive rate. The second method was implemented within the model architecture by evaluating each timestep in the contexts of its neighboring timesteps in two levels. The first level was during the CNN layers, where each timestep was fed into the model with its 25 preceding and following time steps. The second layer was in the BiLSTM layers, where each timestep was evaluated in the context of its neighboring 330 timesteps. We also tested other model configurations with different lengths of neighboring timesteps and found that the abovementioned parameters are optimal for maximal accuracy and manageable complexity.

However, this meant that the spectral maps had to be compressed to reduce model complexity. This compression was made by extracting Mel-spectrogram with 124 filter banks, where the lower frequencies were represented with higher resolution than the higher frequencies. In addition, the model was trained on data that was only labeled in the time domain and not the frequency domain. The combined effect of these two properties resulted in a model with high temporal accuracy but lower spectral accuracy, specifically when estimating the upper boundaries of the USVs. This limitation should be taken into account when using HybridMouse.

Another limitation is that the software does not characterize the USV structure. Our model's main goal is to extract USVs even in harsh recording settings. Once these USVs are extracted, the users may proceed with the analysis independently, using any number of available methods for clustering, Classification, and syntax analysis on the extracted USVs.

In summary, HybridMouse is a powerful tool for automatic identification and extraction of mouse ultrasonic vocalizations in variable recording conditions. It may be used for high-throughput detection of murine social vocalizations.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found at: https://mousetube.pasteur.fr/; https://github.com/gutzcha/HybridMuse.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee, University of Haifa.

AUTHOR CONTRIBUTIONS

YG: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing, and visualization. KB: conceptualization and writing–review and editing. SN: conceptualization, resources, data curation, writing–review and editing, and project administration. LC: conceptualization and data curation. YH-O: conceptualization, supervision, and writing–review and editing. SW: conceptualization, writing–review and editing, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by ISF-NSFC joint research program (Grant No. 3459/20 to SW), the Israel Science Foundation (ISF Grants No. 1350/12, 1361/17 to SW, and 258/17 to LC), the Ministry of Science, Technology and Space of Israel (Grant No. 3-12068 to SW), and the United States-Israel Binational Science Foundation (BSF Grant No. 2019186 to SW).

ACKNOWLEDGMENTS

We thank Dr. Irina Kustanovich for her contribution in acquiring the training data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2021.810590/full#supplementary-material

REFERENCES

Bello, I., Zoph, B., Vaswani, A., Shlens Quoc, J., and Le Google Brain, V. (2019). "Attention augmented convolutional networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (Seoul).

Burkett, Z. D., Day, N. F., Peñagarikano, O., Geschwind, D. H., and White, S. A. (2015). VoICE: a semi-automated pipeline for standardizing vocal analysis across models. *Sci. Rep.* 5, 1–15. doi: 10.1038/srep10237

Chen, Z., and Wiens, J. J. (2020). The origins of acoustic communication in vertebrates. *Nat. Commun.* 11, 1–8. doi: 10.1038/s41467-020-14356-3

Coffey, K. R., Marx, R. G., and Neumaier, J. F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 859–868. doi: 10.1038/s41386-018-0303-6

Fonseca, A. H. O., Santana, G. M., Bosque Ortiz, G. M., Bampi, S., and Dietrich, M. O. (2021). Analysis of ultrasonic vocalizations from mice using computer vision and machine learning. *Elife* 10:e59161. doi: 10.7554/eLife.59161

Haack, B., Markl, H., and Ehret, G. (1983). *Sound Communication Between Parents and Offspring. Audit. Psychobiol. Mouse.* Ulm: Universität Ulm. Fakultät für Naturwissenschaften, 57–97.

Lahvis, G. P., Alleva, E., and Scattoni, M. L. (2011). Translating mouse vocalizations: prosody and frequency modulation. *Genes Brain Behav.* 10, 4–16. doi: 10.1111/j.1601-183X.2010.00603.x

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., and Roian Egnor, E. S. (2015). Female mice ultrasonically interact with males during courtship displays. *Elife* 4, 1–24. doi: 10.7554/eLife.06203

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. Available online at: http://pjreddie.com/yolo/ (accessed December 07, 2021).

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Tachibana, R. O., Kanno, K., Okabe, S., Kobayasi, K. I., and Okanoya, K. (2020). USVSEG: a robust method for segmentation of ultrasonic vocalizations in rodents. *PLoS ONE* 15:e0228907. doi: 10.1371/journal.pone.0228907

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. doi: 10.1016/S0065-3454(08)60107-2

Van Segbroeck, M., 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, 465–485.e5. doi: 10.1016/j.neuron.2017.04.005

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need." In *31st Conference on Neural Information Processing Systems (NIPS 2017)* (Long Beach, CA).

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

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 Behavioral Neuroscience

ORIGINAL RESEARCH
published: 14 February 2022
doi: 10.3389/fnbeh.2021.812939

Identification, Analysis and Characterization of Base Units of Bird Vocal Communication: The White Spectacled Bulbul (*Pycnonotus xanthopygos*) as a Case Study

Aya Marck[1], Yoni Vortman[2†], Oren Kolodny[1†] and Yizhar Lavner[3*†]*

[1] The Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, Jerusalem, Israel, [2] Department of Animal Sciences, Hula Research Center, Tel-Hai College, Tel-Hai, Israel, [2] Department of Computer Science, Tel-Hai College, Tel-Hai, Israel

OPEN ACCESS

Edited by:
Sara Anna Bonini, University of Brescia, Italy

Reviewed by:
Nihaad Paraouty, New York University, United States
Masumi Wakita, Teikyo University, Japan

***Correspondence:**
Aya Marck
aya.marck@mail.huji.ac.il
Yizhar Lavner
yizhar.lavner@gmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *10 November 2021*
Accepted: *28 December 2021*
Published: *14 February 2022*

Citation:
Marck A, Vortman Y, Kolodny O and Lavner Y (2022) Identification, Analysis and Characterization of Base Units of Bird Vocal Communication: The White Spectacled Bulbul (Pycnonotus xanthopygos) as a Case Study. Front. Behav. Neurosci. 15:812939. doi: 10.3389/fnbeh.2021.812939

Animal vocal communication is a broad and multi-disciplinary field of research. Studying various aspects of communication can provide key elements for understanding animal behavior, evolution, and cognition. Given the large amount of acoustic data accumulated from automated recorders, for which manual annotation and analysis is impractical, there is a growing need to develop algorithms and automatic methods for analyzing and identifying animal sounds. In this study we developed an automatic detection and analysis system based on audio signal processing algorithms and deep learning that is capable of processing and analyzing large volumes of data without human bias. We selected the White Spectacled Bulbul (*Pycnonotus xanthopygos*) as our bird model because it has a complex vocal communication system with a large repertoire which is used by both sexes, year-round. It is a common, widespread passerine in Israel, which is relatively easy to locate and record in a broad range of habitats. Like many passerines, the Bulbul's vocal communication consists of two primary hierarchies of utterances, *syllables* and *words*. To extract each of these units' characteristics, the fundamental frequency contour was modeled using a low degree Legendre polynomial, enabling it to capture the different patterns of variation from different vocalizations, so that each pattern could be effectively expressed using very few coefficients. In addition, a mel-spectrogram was computed for each unit, and several features were extracted both in the time-domain (e.g., zero-crossing rate and energy) and frequency-domain (e.g., spectral centroid and spectral flatness). We applied both linear and non-linear dimensionality reduction algorithms on feature vectors and validated the findings that were obtained manually, namely by listening and examining the spectrograms visually. Using these algorithms, we show that the Bulbul has a complex vocabulary of more than 30 words, that there are multiple syllables that are combined in different words, and that

a particular syllable can appear in several words. Using our system, researchers will be able to analyze hundreds of hours of audio recordings, to obtain objective evaluation of repertoires, and to identify different vocal units and distinguish between them, thus gaining a broad perspective on bird vocal communication.

Keywords: repertoire, bird call detection, bird vocalization, vocal units, White Spectacled Bulbul, *Pycnonotus xanthopygos*, deep learning, unsupervised learning

INTRODUCTION

Vocal communication is an essential tool for transferring information. It serves a diverse range of species and is a topic of multi-disciplinary interest. Studying the regularities and contexts of bird vocalizations may provide keys to understanding numerous aspects of bird behavior. While being an essential part of various species' biology, the study of vocal attributes and the inference of the signaling properties remains a major challenge. This is because the information conveyed by vocal communication includes many components and facets that include physical attributes such as amplitude, frequency, rhythm, and intensity, as well as more complex aspects such as syllables, words, phrases and more (Kershenbaum et al., 2016). In addition, audio recordings produce a vast amount of digital data per vocalization. Furthermore, these parameters may be expressed differently, which leads to different patterns and correlations between populations and individuals that can be difficult to identify and even predict. This raises intriguing questions about the meaning of animal sounds (Bruno and Tchernichovski, 2019).

According to a study of blue tits, for example, there is a correlation between the call length in males' courtship songs and extrapair paternity. In this case, the call length provides information about the quality of the singer (Kempenaers et al., 1997). Similar patterns were found with respect to rhythm in sparrows (*Passerculus sandwichensis*) and to variability of calls in warblers (*Sylvia communis*) (Balsby, 2000; Sung and Handford, 2020). From the calls' characteristics we can reveal information not only at the level of the individual, but also at the level of the species. For example, studies have shown that species with a large repertoire typically have plastic and non-permanent songs, indicative of learning abilities throughout their lifetime in the vocal domain. These species are called *open-ended learners* (Robinson et al., 2019). Such large repertoires introduce multiple challenges when aiming to unravel the signaling properties behind the vocalizations. For instance, large repertoires have been found to indicate a high reproductive success in some species (Robinson and Creanza, 2019), yet, in Botero et al. (2009), it was demonstrated in tropical mockingbird (*Mimus gilvus*) that the variation between vocal expressions decreased as the bird aged, and the expressions became more consistent. In this study system, individuals with more consistent performance tended to achieve higher dominance status and greater reproductive success (Botero et al., 2009). These unexpected patterns demonstrate the diversity and complexity of vocal communication systems.

Manual annotation and analysis of bird song is laborious, time-consuming, and prone to subjective bias. Deep learning and algorithms for extracting audio parameters have the potential to overcome these limitations and challenges of reproducibility and of scaling up to large datasets. In recent years, analyzing digital recordings has benefited from the development of reliable automatic algorithms and deep learning, such as available software for syllable recognition and clustering (DeepSqueak, Coffey et al., 2019), an online tool for bird species identification (BirdNET, Kahl et al., 2021) and robust software for animal call detection in large and noisy databases (ORCA-SPOT, Bergler et al., 2019). Still, in many cases researchers rely on subjective naming of calls and on manual division of vocal units. In addition, in many studies the manual analysis is based on a limited amount of data and may miss out patterns which may be revealed only if enough data is automatically processed and analyzed. In a broader scope, the development of advanced automated tools for bio-acoustic analysis can support large-scale research and reveal organisms' vocal communication patterns, may facilitate monitoring of populations, and can be leveraged for management and conservation efforts in natural environments (Righini and Pavan, 2020; Kahl et al., 2021).

In this study, using automatic signal processing algorithms and deep learning, we analyzed White Spectacled Bulbul (Pycnonotus xanthopygos) vocalizations. This species is a common, widespread passerine, and was selected as our model since it is characterized by tight social bonds between individuals and a wide repertoire of vocalizations (Shirihai and Svensson, 2018), used year-round by both sexes. We analyzed and characterized 660 base units of the White Spectacled Bulbul from recordings of 14.5 h, to investigate its repertoire and its use of different vocal units. Our analyses show that Bulbul calls are complex vocalizations—*words*, most of them composed of more than one base unit—*syllable*. The complexity of the Bulbuls' vocal communication can be revealed by intuitive hearing as well as by inspecting spectrograms, or by a more elaborate analysis. However, here we present a set of quantitative automatic methods that make up a pipeline of automatic detection of Bulbul calls, and an analysis of these vocal units that allows classification into different groups, both by supervised and unsupervised learning. These methods (1) allow objective validation of the robustness of words' and syllables' classifications; (2) carry out automatic identification and classification to pre-defined classes; and (3) provide the basis for a fully automated process of defining the word and syllable repertoires of a species or an individual.

Our analyses show that the same syllables are used in different words and in distinct geographic populations. This pattern is very likely to indicate a complex hierarchical structure (Kershenbaum et al., 2016) and that the White Spectacled Bulbul is an open-ended learner vocalizing species. Furthermore,

this pattern can imply the existence of a more complicated form of communication. The hierarchy of syllables and words provides a basis for investigating syntax questions that are today the focus of widespread interest (Menyhart et al., 2015; Suzuki et al., 2016; Bruno and Tchernichovski, 2019; Searcy and Nowicki, 2019).

MATERIALS AND METHODS

A block diagram of the processing and analysis stages of different base units of Bulbul's vocalization is depicted in **Figure 1**. Each of the procedures used in each block is detailed below.

Data Set

Our dataset was collected from eight SM4 automatic recorders of Wildlife Acoustics (Wildlife Acoustics, 2018) placed at four different locations (**Figure 2**), two recorders at each. The two recorders were placed several hundreds of meters apart to ensure that more than one individual was recorded in each location. The recordings were taken for a period of 6 months to a year, at dawn, noon, and dusk, for a total of 4 h per day, with each recording lasting from 30 to 60 min. Overall, more than 7,000 h were recorded. Six of the recorders were located in northern Israel—four in the Hula valley (She'ar Yashuv and Agmon Hula) and two in a nearby location on the Naftali Mountain range (Yiftah) that is characterized by different habitat and weather. The last two recorders were installed in the bird observatory in Jerusalem, a distinct population for comparison. Since all the recordings were carried out in natural habitats, they contain many types of background noise including other birds and animals, weather sounds and artificial sounds. We used several methods (bandpass filtering, median clipping, and small object removal) described in section "Word Analysis" to filter out the different noises (**Figure 3**).

Pre-processing

The acoustic signal was sampled at 44,100 KHz and filtered using a Band-Pass filter between 1 KHz and 3.5 KHz to eliminate background noise and preserve frequencies relevant to the Bulbul's vocalization. The signal was divided into short segments, either consecutive segments of equal size (0.5–1 s each, similar to the typical bird vocalization duration) with 50% overlap for automatic detection of acoustic events, or of variable size for extracted words or syllables. In both cases, for each segment the discrete short-time Fourier transform (STFT) is calculated using:

$$X(n, \omega_k) = \sum_m x(m) \cdot w(n-m) \cdot e^{-i\frac{2\pi}{N}km}$$

where $x(n)$ is the acoustic signal, $w(n)$ is a han window used to multiply each frame, where frames of 512 samples (∼12 ms) with a hop size of 128 samples are regularly used. Consequently, a mel spectrogram was computed for each segment. Mel scale is a logarithmic-like scale based on the human auditory system that represents the sound frequencies in a similar way to how we and other animals perceive.

Acoustic Feature Analysis for Analyzing Syllables

In order to compare between vocalization units, we extract several parameters from each signal. Features were extracted both in the time-domain (e.g., zero-crossing rate and energy) and frequency-domain (e.g., spectral centroid and spectral flatness, MFC coefficients). The following features were used to characterize the spectrum:

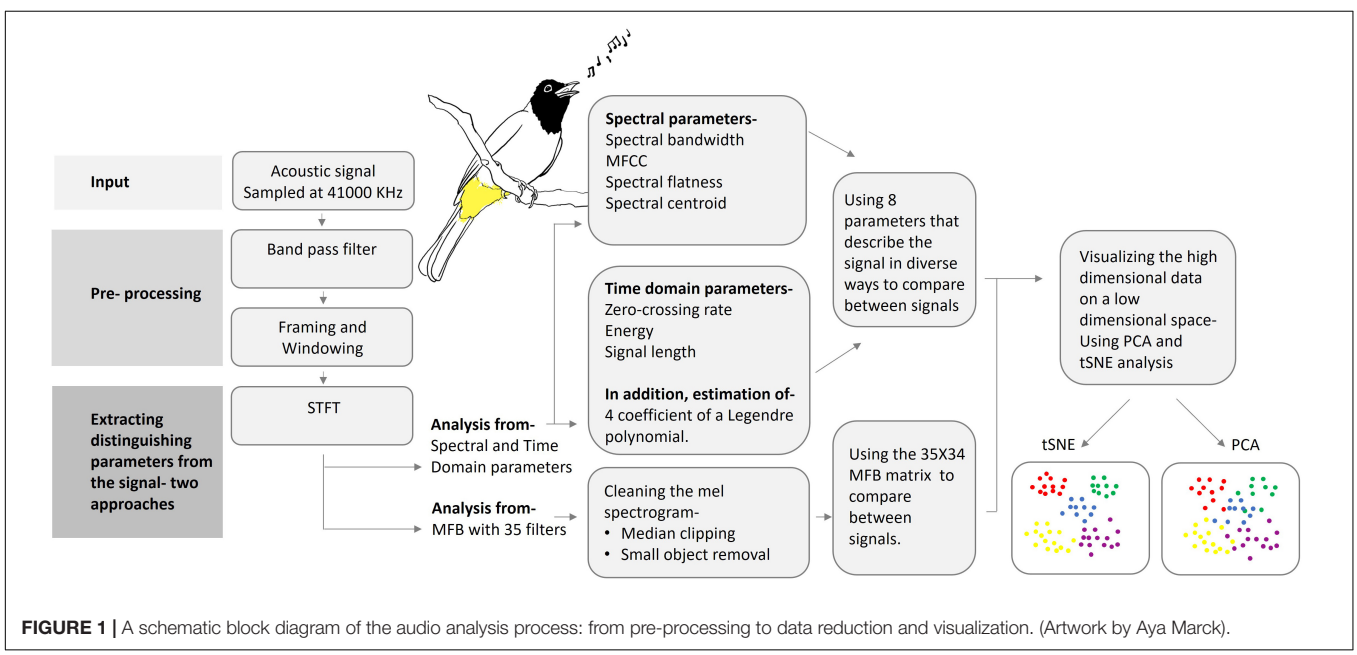


FIGURE 1 | A schematic block diagram of the audio analysis process: from pre-processing to data reduction and visualization. (Artwork by Aya Marck).

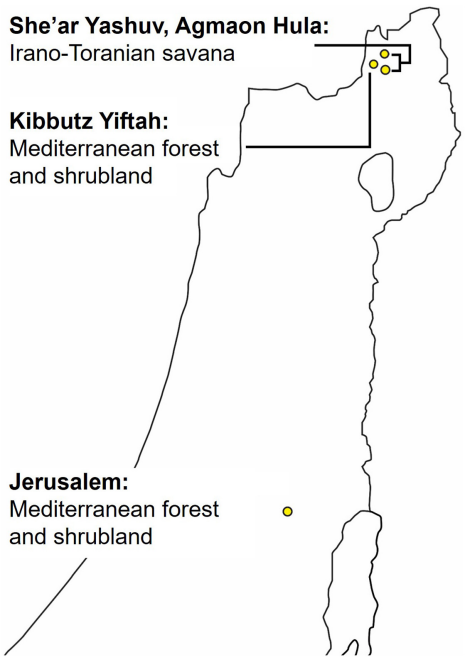


FIGURE 2 | Map of recorders' locations: three locations in the Hula valley and one in Jerusalem.

(a) **Spectral Centroid**(S_c) measures the center of mass of the spectrum, and is calculated as:

$$S_c = \frac{\sum_{k=0}^{N-1} f(k)\,|X(k)|}{\sum_{k=0}^{N-1} |X(k)|}$$

where | X(k)| is the magnitude of the kth' bin of DFT, and f(k) is its center frequency.

(b) **Spectral Flatness** (S_f) measures how tone-like ($S_f \approx 0$) or noise-like ($S_f \approx 1$)) is the sound and is computed as the ratio of the geometric mean to the arithmetic mean of the energy spectrum:

$$S_f = \frac{\sqrt[N]{\pi_{k=0}^{N-1}|X(k)|}}{\frac{1}{N}\sum_{k=0}^{N-1}|X(k)|}$$

(c) **Spectral Bandwidth**, as defined in Klapuri and Davy (2007), which is a weighted standard deviation of the spectrum in a given audio segment:

$$B = \left(\sum_{k=0}^{N-1} |X(k)|\left(f(k) - f_c\right)^2\right)^{1/2}$$

(d) **MFB** (log mel filter bank) are a set of filters arranged according to a mel-scale, a perceptually based frequency scale that aims to mimic the frequency perception of the human auditory system (Davis and Mermelstein, 1980). MFB is widely used in audio signal processing including bird analysis and music signal processing. It is calculated by using Discrete Fourier transform of each frame and applying overlapping triangular filter banks, where each filter output is a weighted sum of magnitudes of frequency bins within its support. Data reduction is also a benefit of this computation.

(e) **Mel Frequency Cepstral Coefficients (MFCC)** The MFCCs, derived from the MFB by applying a discrete cosine transform are very common features in audio analysis. A total of 13 coefficients are computed for each frame, where the first four are used for the analysis.

In Addition, two time-domain parameters were computed:

(f) **Zero Crossing Rate (ZCR)** is defined as the number of times an audio waveform changes its sign within the duration of the signal, and is calculated as:

$$ZCR = \frac{1}{2} \cdot \sum_{n=1}^{K-1} \left|sign(x(n)) - sign(x(n-1))\right|$$

Where K is the signal length.

(g) **Fundamental frequency** f_0 – which is evaluated using the YIN (De Cheveigné and Kawahara, 2002) or the PYIN (Mauch and Dixon, 2014) algorithms.

Legendre Polynomials

In many passerine species, most of the spectral energy of the vocalization is concentrated around the fundamental frequency (Nowicki, 1987; Podos, 2001) since the avian vocal tract attenuates a greater part of the energy of higher harmonics. It is therefore reasonable to assume that a considerable portion of the information conveyed by bird vocalization may be attributed to the intonation, i.e., the fundamental frequency contour.

In order to extract this information quantitatively, we modeled the fundamental frequency contour using a low degree Legendre polynomial, enabling it to capture the different patterns of variation from different vocalizations, so that each pattern could be effectively expressed using only 3–4 coefficients.

This analysis may help us characterize and visualize the fundamental frequency patterns of various syllables which were subjectively divided to different groups.

The usage of the Legendre polynomials for modeling the fundamental frequency was employed in different applications of speech and audio. For example, it has been used to model pitch contour for synthesizing intonation (Zhou et al., 1984), to describe mathematically the nuclear accents found in English in the English isles and to use it for intonation labeling (Grabe et al., 2007). It was also used for automatic language identification (Lin and Wang, 2005), as well as to detect sarcasm in speech and for analyzing prosody (Rakov and Rosenberg, 2013; Rakov, 2019).

The Legendre polynomials is a system of orthogonal polynomials defined as:

$P_n(t) = \frac{1}{2^n n!}\frac{d}{dx^n}(t^2-1)^n$ (Rodrigues formula), where $P_n(t)$ is the n-th order term. It can also be expanded with the polynomials $\{1,\ t,\ t^2 \ldots\}$ using Gram-Schmidt process.

According to this definition the first four terms are:

$$P_0 = 1, P_1 = t, P_2 = \frac{1}{2}\left(3t^2 - 1\right), P_3 = \frac{1}{2}\left(5t^3 - 3t\right)$$

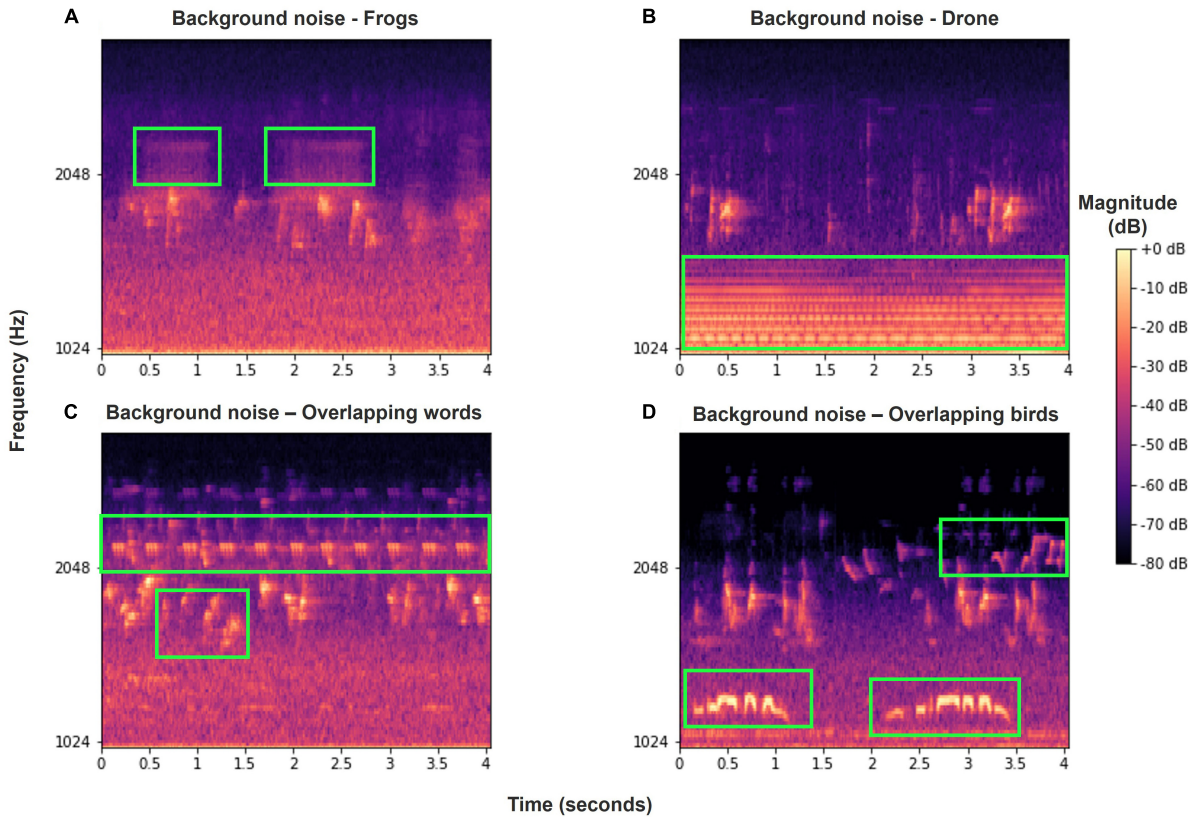


FIGURE 3 | Various examples of background noise or overlapping Bulbul calls in our recordings, marked in green frames. **(A)** Frog vocalizations; **(B)** drone sound; **(C)** overlapping words of other Bulbuls and other birds; **(D)** overlapping calls of other birds. Most of these noise events are filtered out using our pre-processing methods.

Following Grabe et al. (2007), we used the first four polynomials, L_0, L_1, L_2 and L_3 which represent the average of the signal, its slope, quadratic trend, and wavelike shape, respectively.

The following steps were carried out to fit the Legendre series $p\,(t)$ to $\hat{F}_0(t)$:

a. A single syllable or vocalization unit is demarcated and excerpted from the acoustic recording. This was carried out using Audacity (Audacity, 2021).
b. The sampled signal $s\,(t_n)$, $n = 0, 1, \ldots, M-1$ of length M, where $t_n = nT_s$ are the time samples, is filtered using a bandpass filter between 700 and 3,900 Hz, based on the range of frequencies characteristic of the White Spectacled Bulbul.
c. For each sampled syllable or vocalization unit the fundamental frequency contour $F_0(t_n)$ is estimated with either PYIN (Mauch and Dixon, 2014) or the YIN (De Cheveigné and Kawahara, 2002) algorithms, or by using a simple zero-crossing rate analysis signal $z(t_n)$. In many cases the latter is preferred, since the pitch detector algorithms (YIN and PYIN) which were developed mainly for speech and music signals, may not be robust enough for noisy bioacoustic data. Furthermore, the ZCR computation yields a good estimation of $F_0(t_n)$.
d. A polynomial fit is used after scaling the time axis to be between -1 and 1. The estimated contour, $\hat{F}_0(t_n)$, is modeled using an m-th degree Legendre series defined as:

$$p\,(t) = a_0 + a_1 L_1\,(t) + a_2 L_2\,(t) + \ldots + a_m L_m\,(t)$$

where $L_j\,(t)$ is a Legendre polynomial and a_j is its corresponding coefficient. The polynomial series is a least square fit to the data $\hat{F}_0(t_n)$, where the fitting process is carried out by solving an overdetermined set of linear equations of the form:

$$V\,(t) \cdot a = w \cdot \hat{F}_0(t_n)$$

where $V\,(t)$ is the pseudo Vandermonde matrix of t, a is the vector of coefficients.

Figures 4, **5** present spectrograms of a word (**Figure 4**) and of a single syllable (**Figure 5**), with the fitted Legendre series superimposed.

Word Analysis

Bulbuls have complex vocalizations which are described as words that consist of several base units—syllables, and intervals in between. Extraction of features from these complex units is complicated due to their various number and type of syllables,

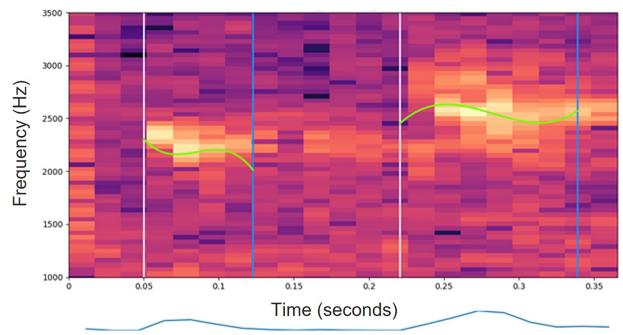


FIGURE 4 | A spectrogram of two syllables of a White-Spectacled Bulbul (*Pycnonotus xanthopygos*). The fundamental frequency contour modeled by a 3rd degree Legendre polynomial is superimposed on the spectrogram. The lower graph depicts the short-time energy of the syllables.

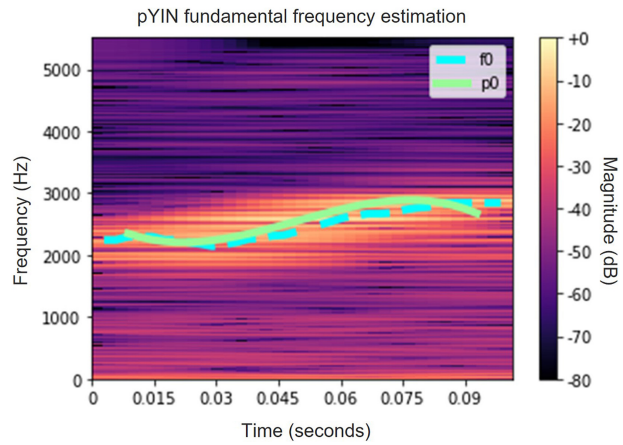


FIGURE 5 | A spectrogram of a single syllable of a White-Spectacled Bulbul (*Pycnonotus xanthopygos*). The fundamental frequency contour is estimated by PYIN (cyan) and by ZCR (light green).

as well as their varying intervals. Therefore, we aimed to create one feature vector that describe the entire vocalization. For this, we used the mel-spectrogram, applied to the raw isolated word signal, with 35 mel filters between a low frequency f_L and a high frequency f_H. We set f_L to 700 Hz and f_H to 3,900 Hz, according to the range of frequencies for most Bulbul vocalizations.

Consequently, a variation of median clipping (Lasseck, 2014; Fukuzawa et al., 2016) following by a small object removal is applied. These two simple image processing techniques are applied to increase the SNR, since in most of the recordings a high background noise is present:

(1) **Median clipping**—in this technique a binary mask is generated for masking background noise, where for each time-frequency point (i, j), its corresponding spectrogram value $S(i, j)$ is compared to a threshold value which is based on the median of the corresponding row and column of that point. Thus, the median clipped spectrogram $S_{mc}(i, j)$ is obtained by:

$$S_{mc}(i,j) = \begin{cases} S(i,j) & if\ S(i,j) > F \cdot \max(med(S_i),\ med(S_j) \\ S_L & otherwise \end{cases}$$

where F is a multiplication factor set here to be 3.5 and S_L is the lowest value in the spectrogram which is set to −80 dB.

(2) **Small object removal**—used to remove small blobs which are probably irrelevant to bird vocalizations and may stem from background noise. This is carried out by converting the median clipped mel-spectrogram to a binary matrix, and for each non-zero entry calculating its immediate non-zero neighbors. Non-zero entries whose number of neighbors is below a pre-defined threshold are zeroed, and a binary mask is obtained. The final spectrogram is then obtained by:

$$S_{sor}(i,j) = S_{mc} \cdot M_{neigh}$$

Alternatively, the usage of a white top-hat transform (Sonka et al., 2014) was examined with no significant difference. An example of a mel-spectrogram before and after these processing operations is depicted in **Figure 6**.

Finally, to compare between feature vectors that represent different words with variable length, we transform the vectors' dimensions to a fixed size by zero padding. Alternatively, the spectrogram was calculated using a fixed number of frames with fixed duration and variable hop length.

Synthesis

To demonstrate that Legendre polynomial coefficients can extract most of the vocal information from Bulbul calls (at least as perceived by a human), and to use another method to validate the parameters we use, we generated a synthetic vocalization based solely on these coefficients. The synthesis is carried out using the following steps:

a. For each input signal with one syllable—a pre-processing is applied, which includes downsampling to a sampling rate of 11,025 Hz, bandpass filtering of the signal using cut-off frequencies (700, 3,900) and demarcation of the syllable boundaries.
b. Fundamental frequency contour of the syllable is evaluated, using either pitch detection algorithm or applying a ZCR analysis on the bandpassed signal. The evaluation is carried out using a frame length of 64 samples (5.8 ms) and a step size of 32 samples (2.9 ms). The result of this stage is a vector of consecutive fundamental frequency values.
c. A three-degree Legendre polynomial is fitted to the fundamental frequency contour, and four Legendre coefficients are obtained.
d. Using the coefficients, a Legendre series is fitted for the time points for which the fundamental frequency contour was evaluated.
e. For each frame m, a short sinusoid is produced, with frequency $p_0\ (t_m)$ using:

$$x(n) = Re\left\{A\ \cdot e^{j(2\cdot\pi\cdot p_0(t_m)\cdot n+\varphi)}\right\}$$

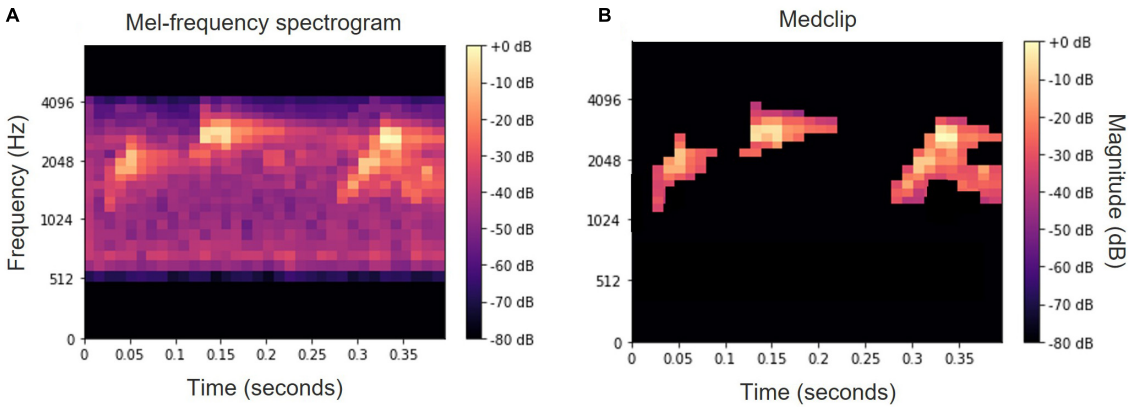


FIGURE 6 | Mel-spectrogram of the word "*tu ti tuyu*" before **(A)** and after **(B)** median clipping and small object removal.

where A is the signal amplitude, n is the sample index, and φ is the phase shift which is corrected for each frame to avoid phase discontinuities. The initial phase is set to $\varphi_{t_m} = 0$ and then for next frames it is set according to the final phase of the former sinusoid:

$$\varphi_{t_{m+1}}(0) = \begin{cases} acos\left(x_{t_m}(N-1)\right) + 2\cdot\pi\cdot\frac{T_s}{T_0(t_m)} \\ \quad if\ 0 \leq \varphi_{t_m}(N-1) < \pi \\ 2\cdot\pi - acos\left(x_{t_m}(N-1)\right) + 2\cdot\pi\cdot\frac{T_s}{T_0(t_m)} \\ \quad if\ \pi \leq \varphi_{t_m}(N-1) < 2\pi \end{cases}$$

where N is the frame size, T_0 (t_m) is the fundamental period evaluated for the m_{th} frame and T_s is the sampling interval.

The concatenation of all the frames yields a chirp-like signal, with a fundamental frequency contour according to the evaluated Legendre series.

Visualizing by Data Reduction

To validate the division into different syllables or words made by listening and observation, which may be subjective, we first tagged 660 vocalizations, containing 1,004 syllables, all collected from nine separate audio files that were recorded at the same location with the same device, each lasting an hour. Further, an algorithm for dimension reduction was utilized, based on the spectral analysis data, to objectively examine the proposed grouping, both for words and syllables. We used PCA (Principal Component Analysis) and t-SNE algorithm (t-distribution Stochastic Neighbor Embedding, Van der Maaten and Hinton, 2008) for dimension reduction and visualization. Both methods perform a mapping from a high dimension to a low dimension of 2–3, so that the proximity or distance between points in the high dimension is maintained in the low dimension. Syllables feature vectors were reduced from 8 dimensions to 2, and words feature vectors from 1,190 dimensions to 2. We expect that syllables or words that were divided by listening into one group should be in the same cluster, whereas feature vectors from sounds classified subjectively as belonging to different groups would be divided by the algorithm into different clusters and would be far apart. While methods based on linear algorithms such as PCA may not yield clear results, the usage of a nonlinear method such as t-SNE, may show the clustering and separation of sounds in a manner similar to their definition on an auditory basis.

For analysis and computations, we used Python 3.8 and suitable packages; Librosa (McFee et al., 2015) for audio signal processing, and Scikit-learn (Kramer, 2016) for data analysis and data reduction. Dataset for this analysis and codes are available on GitHub (see Data Availability Statement).

Detection of Bulbul Calls

The analysis of words and syllables receives as input an audio signal where the relevant vocalization is located. It is therefore necessary to identify and extract the desired call events from long and noisy recordings. This can be done manually; however, the number of call events that can be derived in this way is limited. A machine learning approach should be applied in order to extract thousands of vocalization units for further analysis. We used several Deep Neural Networks (DNN), and in particular Convolutional Neural Networks (CNN) (Goodfellow et al., 2016) to automatically detect the Bulbuls' calls in the recordings. Most of the recordings are between half an hour to 1 h long and contain intense background noise as well as other birds' and other animals' vocalizations (including human speech). Several models were tested for the detection:

(a) A CNN with 5 blocks of convolution and max pooling layers, connected to a 90 hidden units fully connected (FCN) layer and an output layer with a total of 1,117,781 trainable parameters;
(b) A resnet architecture with 14 convolution layers in resnet blocks, connected to a FCN layer of 90 hidden units with a total of 625,397 trainable parameters;
(c) A mini-Xception model (Chollet, 2017, 2021) with 7 convolution layers and a total of 717,249 trainable parameters.

The input for all the models was obtained by dividing the acoustic signal into consecutive segments of 1 s each, with

an overlap of 50%. For each segment, a log mel-spectrogram was calculated by using frames of 2,048 samples (∼48 ms for fs = 44,100) and hop size of 700 samples. The mel-spectrogram is a matrix of 50 × 60 (number of mel filters x number of time bins), which was pre-processed by median clipping and small object removal for noise reduction.

The CNN model (**Figure 7**) is composed of 5 blocks of convolution: a first block of a convolution layer with 32 3 × 3 kernels following by a max-pooling layer (2 × 2) and a Relu activation function. The following convolution blocks are the same, where the number of kernels doubles at each block. After flattening the output feature map of the final convolution layer, a fully connected layer of 90 units is applied with dropout of 0.5. Finally, the output layer with one unit and a Sigmoid activation function and threshold value of 0.5 yields a binary output of (Bulbul = 1/non-Bulbul = 0). Consecutive segments predicted as Bulbul (1), were merged into one call event for further processing.

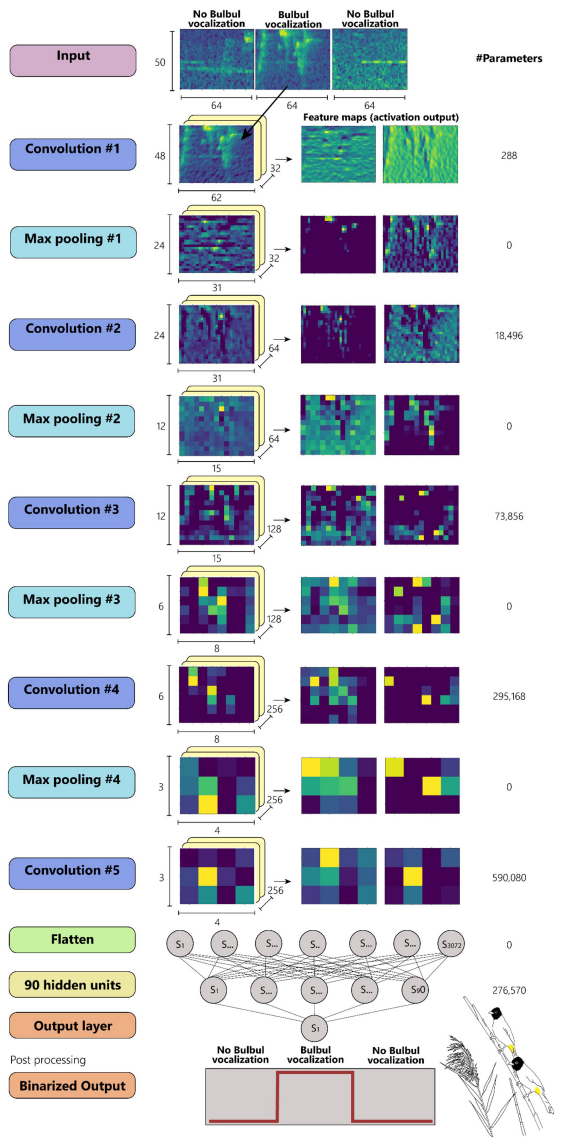


FIGURE 7 | A CNN model configuration with 5 convolutional layers. (Artwork by Aya Marck).

The training set for the detection included 57 recordings of variable durations, with a total duration of more than 8 h, which were annotated manually. The annotations were made by examining the spectrograms and listening to the corresponding sounds, and the start and end times of each identified vocalization were listed ("strong labeling," Mesaros et al., 2021). This dataset contains several thousands of Bulbul calls, along with other birds, human activity, and many other sounds from various sources. We used 70% of this dataset for training and the remainder for testing. Out of the training data, 10% was randomly selected and served for validation. A segment-based evaluation is applied, and each segment is considered a bird call if at least 30% of it is overlapped with an annotated Bulbul vocalization event.

For data augmentation we used five different methods to increase variability of the data, thus improving the robustness of the networks: (a) Adding white noise to each Mel-spectrogram. (b) Applying horizontal random flip to the mel- spectrogram. (c) Applying a random zoom-in with factors between (−0.2 to −0.01). (d) Time stretching or compressing in the acoustic signal in the range (0.7–1.3). (e) Pitch shift of the signal in the range (0.94–1.06). This process tripled the size of the input data. Dataset for this network and codes are available on GitHub (see Data Availability Statement).

MAIN RESULTS

Convnet Network Performance—Bulbul Event Detection

We measured the performance of the DNNs in detecting Bulbul's vocalizations using a test dataset of 3 h with several thousand individual calls, which also contained high background noise from other birds and animals, as well as anthropophonic and geophonic sounds (Righini and Pavan, 2020). The test dataset was pre-processed with the same procedure used for the training dataset, which included MFB calculation, median clipping, and small object removal. The test set was randomly selected from the recordings dataset, and a segment-based evaluation was carried out using a 1 s segment.

A correct identification rate of 75% (True Positive Rate, or recall, i.e., the ratio between the number of Bulbul vocalization segments correctly identified, to the total number of segments with Bulbul vocalization in the test recording set) was yielded by the CNN described in section "Detection of Bulbul Calls," with a relatively low False Positive occurrences of less than one third (27%) of the True detections. In a manual examination of the results, the non-identified calls (false negatives) were usually further from the microphone or very noisy. The Resnet and the mini-exception models yielded similar results.

A Wide Repertoire of Distinct Words That Repeat Themselves

The White Spectacled Bulbul demonstrated a broad vocabulary of more than 30 distinct words. Over 660 calls were tagged, named,

and analyzed, and were manually categorized as 13 different words (see examples in **Supplementary Figure 1**). Each word was represented by a mel spectrogram of 35 × 34 which was cleaned and filtered as described in section "Word Analysis." Two computational analyses were performed to visualize the 1,190 dimensions of the mel-spectrogram as a two-dimensional map—PCA and t-SNE. **Figure 8A** shows the PCA result where each dot represents a word, and each color represents one unique naming tag. As shown, the vocalizations that were perceived and categorized as belonging to the same word by a human expert were also mapped to the same region on the 2-D plane of the unsupervised PCA (groups of similar colors). This grouping is further demonstrated using a second unsupervised method: the tSNE analysis (**Figure 8B**). The tSNE plot places most of the words (different colors) in well-defined, separate clusters. These results suggest that Bulbuls use distinct words that appear non-random as they repeat themselves across different recordings throughout the year. Our manual process of naming words and categorization aligns well with these unsupervised dimensionality reduction analyses.

Different Words Are Composed From the Same Shared Base Units

A total of 1004 audio signals containing 22 different syllables were excerpted from the words and manually categorized and tagged with a number by listening and examining the spectrogram (see examples in **Supplementary Figure 2**). Syllables were represented by an eight-parameter feature vector, which includes—syllable length, spectral flatness, spectral centroid, bandwidth, and four Legendre polynomials coefficients—based on the fundamental frequency contour. The results of both PCA and tSNE are provided in **Figures 9**, **10A**, respectively. **Figure 10B** shows that using only the Legendre coefficients as parameters is sufficient to describe the variance of the acoustic signal. In **Figure 10A**, words (denoted with capital letters) are composed of different syllables. The same syllables often appear in different words. This analysis can serve as an effective test or validation for manual assessments; for example, we found that two syllables from different words that clustered together were initially misidentified as different syllables. Later listening and visual inspection of the spectrograms confirmed that they represent a single syllable in different contexts. These results show that there is a collection of distinct syllables that repeat themselves and appear in different words, indicating that different words are constructed from the same shared units of similar and non-random syllables.

Classification of Words From a New Dataset

The final stage in the automatic pipeline presented above is to classify the segments detected as Bulbul vocalizations by the deep CNN into their corresponding classes. For this purpose, we first applied the trained CNN model presented in section "Detection of Bulbul Calls" to a 3-h long recording dataset. A group of 800 segments were detected as Bulbul calls and demarcated using the model. These were used to construct two test datasets: 1. A dataset of 126 segments consisting only of words recognized as belonging to the predefined repertoire, 2. A dataset containing 200 segments selected randomly from all detected segments. Later examination found that these included both known words (101 segments) and unknown segments that cannot be classified into existing word-categories. These segments include words that the researcher has yet to annotate, a mix of words (when more than one bird sings in unison), fragments of words (initial or final syllables), as well as a few false positives (i.e., not a Bulbul vocalization).

Using the dimensionality reduced PCA representation described in section "A Wide Repertoire of Distinct Words That Repeat Themselves" for training, the high-dimensional mel-spectrograms of the test segments were projected into the reduced PCA space to produce a low-dimensional representation of the test data. Consequently, three simple classifiers were used to classify the test segments: A k-nearest neighbor (KNN) classifier with K = 3, a nearest centroid classifier, where the prediction of the test word is set according to the label of the closest centroid among all centroids of the training groups, and a Support Vector Machine (SVM) with a radial basis function kernel. The classification results of applying the classifiers on the first fully annotated test set are summarized in **Table 1**. As can be seen, using a 10-dimensional representation a very high classification accuracy was obtained, of 95.2, 94.4, and 96.8%, for the nearest centroid, KNN and SVM, respectively. Even better scores were achieved using a 100 dimensional representation.

The same pre-processing was used in the second dataset, in which the detected words were selected randomly. However, to reject the unrelated segments detected erroneously as Bulbul, a threshold value was set, based upon the distances of all training word samples from their respective closest centroid. For each test segment, whenever the nearest centroid distance is higher than the threshold, this instance is discarded. Using this procedure, most of the non-Bulbul segments were rejected, as well as some Bulbul vocalizations. A classification accuracy of 77% was achieved for this dataset.

Evidently, when the CNN is used to identify words that were included in the training repertoire, this classification tool can guarantee a fully automated process, with very high recognition rates. When unknown vocalizations are also considered, recognition rates are lower. These can be improved in a number of ways; the researcher may inspect the detected words before classification to remove the irrelevant vocalizations, and high accuracy results could be also achieved by applying simple classification tools.

DISCUSSION

The field of bio-acoustic research is rapidly expanding, with technological advances facilitating new approaches to fundamental biological questions and new applications in conservation. This includes utilization of deep neural networks in ecological studies for monitoring and processing large datasets of field recordings (Bergler et al., 2019; Dufourq et al., 2021). As in our framework, most of these studies use a convolutional neural network, with an augmentation approach similar to ours.

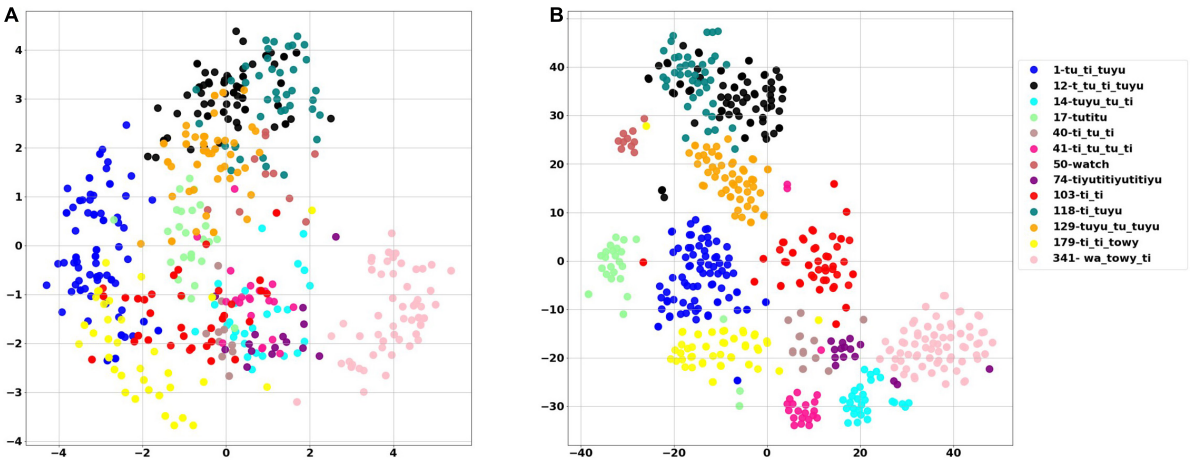


FIGURE 8 | The PCA score plot of the first two PCs of the word analysis **(A)**, and tSNE analysis **(B)** where each color represents one unique naming tag. Instances of each word are grouped together by both unsupervised algorithms. For example, all of the pale pink points are grouped together, as they all represent word number 314—"*wa towy ti*."

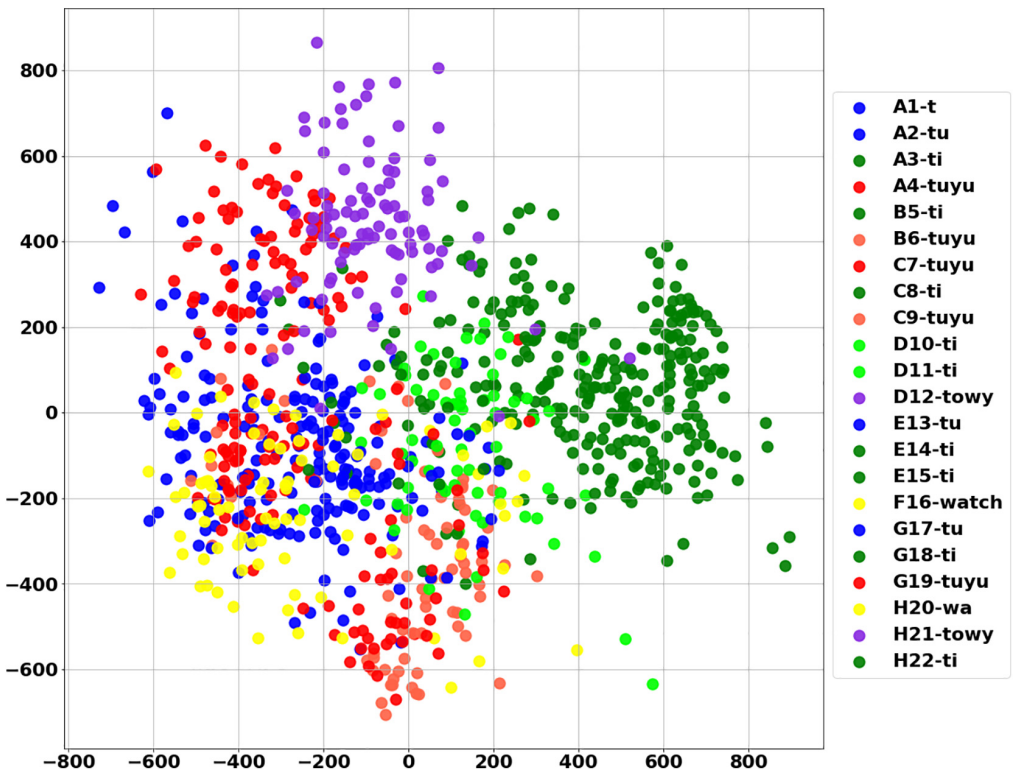


FIGURE 9 | Visualization of syllables using PCA. Explained variance ratio (first two components): (0.43169461 0.2822671).

These are typically complemented by pre-processing and post-processing stages, which are tailored to the specific species, environment and future use of the data. Several frameworks that automate the analysis of animal communication and quantify vocal behavior have been developed. These include studies and software packages such as Sound Analysis Pro (Tchernichovski and Mitra, 2004), which includes automatic segmentation and calculation of acoustic features as well as clustering of vocal units, and DeepSqueak (Coffey et al., 2019), in which a regional CNN (Faster rCNN) and k-means are used to detect and cluster mouse ultrasonic vocalizations. Similar to our syllable analysis approach, these programs extract different acoustic features to characterize and differentiate between vocal units, and use unsupervised methods to visualize and classify the data. In our

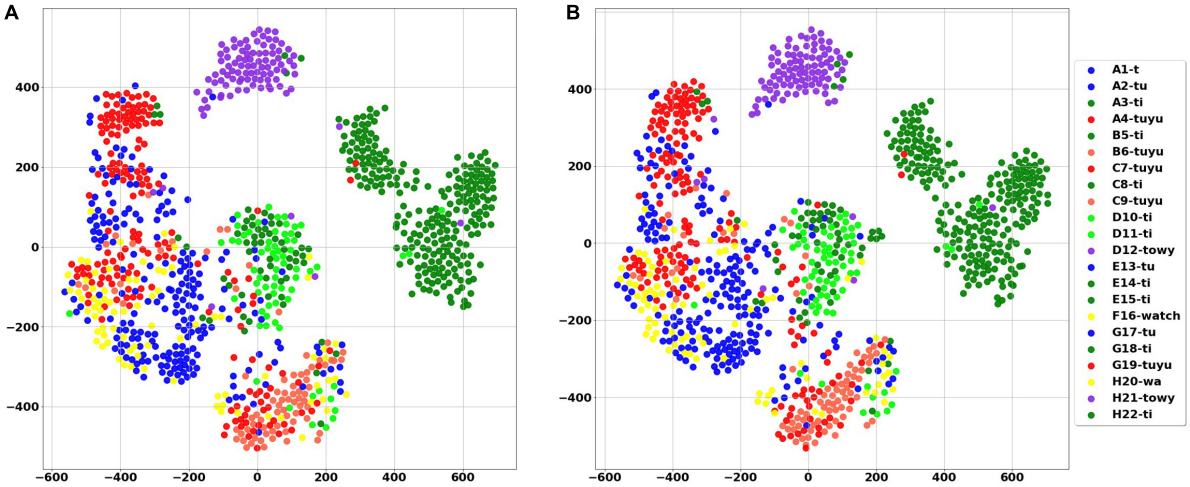


FIGURE 10 | Visualization of syllables using tSNE. Each syllable is represented by a single point. **(A)** The analysis includes all parameters. **(B)** The analysis includes only 4 Legendre coefficients. In both figures same syllables appears in different words. For example, the purple dots that represent the syllable "*towy*," were excerpted from two different words (denoted by the letters D and H) and are clustered together. Similarly, the syllable "*ti*" (dark green), derived from six different words, resides in the same region both in PCA and tSNE analyses.

study, however, we used solely Legendre Polynomials to capture the shape of the fundamental frequency contour, demonstrating that very few coefficients are sufficient to effectively express the different patterns of syllable variation. Goffinet et al. (2021) used a variational autoencoder (VAE) to extract features in reduced latent spaces, employed on mouse USV and zebra finch vocalizations, and demonstrated the effectiveness of a latent space representation when compared with handpicked selected features in different vocal analyses. This kind of analysis shares the data-driven approach applied in our word analysis process and seems effective for recognition of patterns and characterization of units from the complex and high-dimensional data of vocal communication. However, most of these studies used recordings in artificial environments (did not contain high background noise) or were designed for a specific species, and it is challenging to apply them to non-model passerine in the field.

Several significant cross-disciplinary challenges in the study of vocal communication still exist (Prat et al., 2017; Teixeira et al., 2019; Mercado and Perazio, 2021). These methodological challenges arise due to the vast amount of digital data produced, the large number of parameters that can be potentially extracted (e.g., frequency, duration, pitch, etc.) and the lack of clear hypotheses regarding the parameters and the signals they convey (Suzuki et al., 2019).

A basic conceptual challenge is the categorization of vocal units, and more generally, the definition of the repertoire. Most often, the construction of a repertoire dictionary is an expression of guidelines defined by the researcher. Vocal units can be categorized, for example, by a "hard" division (only the exact call is considered the same word), or a "soft" division (a variety of calls that sound alike are considered the same word). This scheme may or may not express the way that animals perceive or use their repertoire (Kershenbaum et al., 2016; Mercado and Perazio, 2021). Moreover, while human perceptual properties may be fit for such a task, this may cause additional methodological challenges, may add room for inconsistencies and reduce reproducibility. Thus, quantitative validation of vocal categorization may aid in overcoming these challenges.

By taking advantage of the benefits of automatic analysis we overcome these challenges in two ways:

1. **Processing large amounts of data**—Our CNN model which is used for identifying Bulbul sounds is highly efficient since it reduces manual work and processes big datasets. Further, deep learning has the ability to recognize discriminative patterns in a non-trivial way and can consider combinations of multiple variables that the human auditory system may miss. With the right adjustments, this network can be used with noisy recordings taken in the wild, to identify various bird species and perhaps other animals with wide vocal repertoire. As part of a post-processing, the call events can be analyzed—both for audio analysis (e.g., clustering and comparing

TABLE 1 | A summary of the classification results with three different classifiers and with a different amount of dimensionality reduction components.

Accuracy % (# errors)			n_components
SVM	**KNN**	**Nearest centroid**	
98.4 (2)	97.6 (3)	96.8 (4)	100
96.8 (4)	94.4 (7)	95.2 (6)	10
92.1 (11)	92.8 (9)	92.9 (10)	5
84.9 (19)	85.7 (17)	82.5 (22)	3
70.6 (37)	71.4 (36)	70.6 (37)	2

The classification accuracy is the ratio between the number of correctly classified words and the total number of words in the test dataset (of 126 words).

between calls) and for statistical analysis (e.g., call events per day and daily or seasonal fluctuation of calls).

2. **Validation and avoidance of biases**—Our automatic analysis for syllables and words can validate our subjective classification and makes it possible to significantly reduce biases that may occur in manual analysis. The categorization process can be done manually by characterizing vocal units in a parametric analysis (like frequency, etc.) into distinct words. But, it is a long and tedious process, made more difficult by large amounts of audio files. Linear and non-linear dimensionality reduction and visualization techniques as well as supervised classification schemes can demonstrate that the manual choices which have been made in early stages are to an extent consistent. Further, it can highlight mistakes and reveal new insight about the categorization. As seen in our results, by using PCA, two syllables from different words which were cataloged differently were found to be the same syllable.

In addition to validating the manual work, the use of our syllable analysis tool enables us to compare similar syllables in different geographic regions in order to identify minor differences. Furthermore, the word analysis tool allows to compare dialect differences between populations, identify which words are unique and which are shared, and investigate if these differences are correlated with geographical distances or genetic differences. During our fieldwork, cameras were placed next to some of the automatic recorders. Further behavioral research can be conducted by analyzing these thousands of short video clips using other deep learning models. Additionally, since females and males are morphologically similar, sex and age of the recorded individuals are unknowns. This information, and possible sex differences in vocalizations can be obtained by combining audio recordings, corresponding video clips and DNA samples (research in progress).

Our study revealed a particularly large vocal repertoire produced by various Bulbul populations. We confirmed through our analysis that the repertoire is derived from a combination of the same basic units which are used to generate new or unique words. This result may be explained in at least two ways. First, is that the Bulbul uses an efficient hierarchical repertoire by maximizing a limited stock of syllables for composing a variety of different words. These signal combinations can be cultural or socially transmitted. Another explanation is that syllables are innate and there may be genetic constraints upon the neural control or physiological mechanisms, limiting the production of syllables, while words, which are made up of syllables, can be invented or learned. The combination of the large repertoire together with a vocal structure of words comprised from syllables may suggest that the Bulbul is an open-ended vocal learning species (Cornez et al., 2017).

Our pipeline provides a robust framework that enables us to process large amounts of data with very little manual intervention, and to classify and validate our findings in an unsupervised analysis. Using the pipeline, raw noisy recording can be processed down to the level of a single word or syllable. This facilitates further analysis and research on Bulbul vocal communication, opening the door to investigation of question such as whether the emergence of novel words characterizes isolated populations, whether different Bulbul calls convey a specific message or information and whether the syllable arrangement into words has certain rules that operate over them. The framework's code and documentation are available on GitHub in the following link: https://github.com/BulbulNET?tab=repositories. The code can be utilized for study of Bulbul vocalizations as is and can easily be adapted to analysis of vocalizations of other passerines that share similarities with the Bulbuls' vocalization structure. We would be happy to assist in the incorporation of the code or parts of it into new pipelines that are being developed for such studies with the goal of generating new insights into the complex world of animal acoustic communication.

DATA AVAILABILITY STATEMENT

The analysis methods and datasets for this study can be found in GitHub, in the following link https://github.com/BulbulNET?tab=repositories.

AUTHOR CONTRIBUTIONS

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

FUNDING

This study was supported in part by the US–Israel Binational Science Foundation (BSF), the Israel Science Foundation (ISF) (grant number 1826/20), and the Hula Research Center which is a joint venture of Tel-Hai and the KKL - Jewish National Fund (JNF).

ACKNOWLEDGMENTS

We thank Anat Levi, Dima Ruinskiy, Moshe Leibovich, Rami Cohen, Re'a Shaish, Ronen Melamed, Sivan Margalit, Uria Sade, and Yael Lenard for data collection and processing. We also thank Arik Kershenbaum and the members of the Kolodny, Vortman, and Lavner labs for insightful comments. The study was facilitated by the staff and infrastructure of the Nili and David Jerusalem Bird Observatory of the JBO—Amir Balaban and Gideon Perlman, and of the Agmon Hula of the JNF-Shai Agmon.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2021.812939/full#supplementary-material

REFERENCES

Audacity (2021). *Audacity Team*. Available online at: https://audacityteam.org/ (accessed October 19, 2020).

Balsby, T. J. (2000). Song activity and variability in relation to male quality and female choice in whitethroats *Sylvia communis*. *J. Avian Biol.* 31, 56–62.

Bergler, C., Schröter, H., Cheng, R. X., Barth, V., Weber, M., Nöth, E., et al. (2019). ORCA-SPOT: an automatic killer whale sound detection toolkit using deep learning. *Sci. Rep.* 9:10997. doi: 10.1038/s41598-019-47335-w

Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., De Kort, S. R., et al. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Anim. Behav.* 77, 701–706. doi: 10.1016/j.anbehav.2008.11.020

Bruno, J. H., and Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behav. Process.* 163, 53–59. doi: 10.1016/j.beproc.2017.11.001

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, 1251–1258. doi: 10.1109/CVPR.2017.195

Chollet, F. (2021). *Deep Learning With Python*, 2nd Edn, Shelter Island, NY: Manning.

Coffey, K. R., Marx, R. G., and Neumaier, J. F. (2019). DeepSqueak: a deep learning–based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 859–868. doi: 10.1038/s41386-018-0303-6

Cornez, G., Madison, F. N., Van der Linden, A., Cornil, C., Yoder, K. M., Ball, G. F., et al. (2017). Perineuronal nets and vocal plasticity in songbirds: a proposed mechanism to explain the difference between closed-ended and open-ended learning. *Dev. Neurobiol.* 77, 975–994. doi: 10.1002/dneu.22485

Davis, S., and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Signal Process.* 28, 357–366. doi: 10.1109/TASSP.1980.1163420

De Cheveigné, A., and Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* 111, 1917–1930. doi: 10.1121/1.1458024

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 Sen. Ecol. Conserv.* 7, 475–487. doi: 10.1002/rse2.201

Fukuzawa, Y., Marsland, S., Pawley, M., and Gilman, A. (2016). "Segmentation of harmonic syllables in noisy recordings of bird vocalisations", in *Proceedings of the 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. Palmerston North, 1–6. doi: 10.1109/IVCNZ.2016.7804445

Goffinet, J., Brudner, S., Mooney, R., and Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife* 10:e67855. doi: 10.7554/eLife.67855.sa2

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Grabe, E., Kochanski, G., and Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Lang. Speech* 50, 281–310. doi: 10.1177/00238309070500030101

Kahl, S., Wood, C. M., Eibl, M., and Klinck, H. (2021). BirdNET: a deep learning solution for avian diversity monitoring. *Ecol. Inform.* 61, 101236. doi: 10.1016/j.ecoinf.2021.101236

Kempenaers, B., Verheyen, G. R., and Dhondi, A. A. (1997). Extrapair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. *Behav. Ecol.* 8, 481–492. doi: 10.1093/beheco/8.5.481

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., et al. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol. Rev.* 91, 13–52. doi: 10.1111/brv.12160

Klapuri, A., and Davy, M. (eds.) (2007). *Signal Processing Methods for Music Transcription*. Berlin: Springer doi: 10.1007/0-387-32845-9

Kramer, O. (2016). "Scikit–learn," in *Machine Learning for Evolution Strategies*. Berlin: Springer. 45–53. doi: 10.1007/978-3-319-33383-0_5

Lasseck, M. (2014). "Large–scale identification of birds in audio recordings," in *Proceedings of the CLEF Working Notes*. Berlin: Springer. 643–653.

Lin, C. Y., and Wang, H. C. (2005). "Language identification using pitch contour information," in *Proceedings of the (ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*. Philadelphia, PA, Vol. 1, I–601.

Mauch, M., and Dixon, S. (2014). "pYIN: a fundamental frequency estimator using probabilistic threshold distributions," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP)*. Florence, 659–663. doi: 10.1109/ICASSP.2014.6853678

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., et al. (2015). "Librosa: audio and music signal analysis in python," in *Proceedings of the 14th Python in Science Conference,* Vol. 8 (Austin, TX), 18–25. doi: 10.25080/Majora-7b98e3ed-003

Menyhart, O., Kolodny, O., Goldstein, M. H., DeVoogd, T. J., and Edelman, S. (2015). Juvenile zebra finches learn the underlying structural regularities of their fathers' song. *Front. Psychol.* 6:571. doi: 10.3389/fpsyg.2015.00571

Mercado, E., and Perazio, C. E. (2021). All units are equal in humpback whale songs, but some are more equal than others. *Anim. Cogn.* doi: 10.1007/s10071-021-01539-8 [Epub ahead of print].

Mesaros, A., Heittola, T., Virtanen, T., and Plumbley, M. D. (2021). Sound event detection: a tutorial. *IEEE Signal Process. Magaz.* 38, 67–83. doi: 10.1109/MSP.2021.3090678

Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* 325, 53–55. doi: 10.1038/325053a0

Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* 409, 185–188. doi: 10.1038/35051570

Prat, Y., Azoulay, L., Dor, R., and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biol.* 15:e2002556. doi: 10.1371/journal.pbio.2002556

Rakov, R., and Rosenberg, A. (2013). *Sure, I Did the Right Thing: A System for Sarcasm Detection in Speech*. Shanghai: Interspeech, 842–846. doi: 10.21437/Interspeech.2013-239

Rakov, R. (2019). *Analyzing Prosody With Legendre Polynomial Coefficients*, Ph.D. thesis. New York, NY: City University of New–York.

Righini, R., and Pavan, G. (2020). A soundscape assessment of the Sasso Fratino integral nature reserve in the Central Apennines, Italy. *Biodiversity* 21, 4–14. doi: 10.1080/14888386.2019.1696229

Robinson, C. M., and Creanza, N. (2019). Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success. *Ecol. Evol.* 9, 8362–8377. doi: 10.1002/ece3.5418

Robinson, C. M., Snyder, K. T., and Creanza, N. (2019). Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open–ended learning. *eLife* 8, e44454. doi: 10.7554/eLife.44454

Searcy, W. A., and Nowicki, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Anim. Behav.* 151, 217–227. doi: 10.1016/j.anbehav.2019.01.015

Shirihai, H., and Svensson, L. (2018). *Handbook of Western Palearctic Birds, Volume 1: Passerines: Larks to Warblers*. London: Bloomsbury Publishing.

Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image Processing, Analysis, and Machine Vision*. Boston, MA: Cengage Learning.

Sung, H. C., and Handford, P. (2020). Song characters as reliable indicators of male reproductive quality in the Savannah Sparrow (*Passerculus sandwichensis*). *Can. J. Zool.* 98, 32–38. doi: 10.1139/cjz-2019-0018

Suzuki, T. N., Wheatcroft, D., and Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7:10986. doi: 10.1038/ncomms10986

Suzuki, T. N., Griesser, M., and Wheatcroft, D. (2019). Syntactic rules in avian vocal sequences as a window into the evolution of compositionality. *Anim. Behav.* 151, 267–274. doi: 10.1016/j.anbehav.2019.01.009

Tchernichovski, O., and Mitra, P. P. (2004). *Sound Analysis Pro User Manual*. New York, NY: CCNY, 482.

Teixeira, D., Maron, M., and van Rensburg, B. J. (2019). Bioacoustic monitoring of animal vocal behavior for conservation. *Conserv. Sci. Pract.* 1:e72. doi: 10.1111/csp2.72

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t– SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Wildlife Acoustics (2018). *Song Meter SM4 Acoustic Recorder*. Concord, MA: Wildlife Acoustics Inc. Available online at: https://www.wildlifeacoustics.com/products/song-meter-sm4 (accessed December 2, 2020).

Zhou, S. M., Suen, C. Y., and Bui, T. D. (1984). Simulation of speech intonation by Legendre orthogonal polynomials. *Simulation* 42, 215–219. doi: 10.1177/003754978404200502

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fnbeh.2022.833383

Inhibition of miR-128 Enhances Vocal Sequence Organization in Juvenile Songbirds

Caitlin M. Aamodt[1,2]* *and Stephanie A. White*[1]*

[1] Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, CA, United States,
[2] Department of Pediatrics, University of California San Diego School of Medicine, La Jolla, CA, United States

The molecular mechanisms underlying learned vocal communication are not well characterized. This is a major barrier for developing treatments for conditions affecting social communication, such as autism spectrum disorder (ASD). Our group previously generated an activity-dependent gene expression network in the striatopallidal song control nucleus, Area X, in adult zebra finches to identify master regulators of learned vocal behavior. This dataset revealed that the two host genes for microRNA-128, ARPP21 and R3HDM1, are among the top genes whose expression correlates to how much birds sing. Here we examined whether miR-128 itself is behaviorally regulated in Area X and found that its levels decline with singing. We hypothesized that reducing miR-128 during the critical period for vocal plasticity would enhance vocal learning. To test this, we bilaterally injected an antisense miR-128 construct (AS miR-128) or a control scrambled sequence into Area X at post-hatch day 30 (30 d) using sibling-matched experimental and control pupils. The juveniles were then returned to their home cage and raised with their tutors. Strikingly, inhibition of miR-128 in young birds enhanced the organization of learned vocal sequences. Tutor and pupil stereotypy scores were positively correlated, though the correlation was stronger between tutors and control pupils compared to tutors and AS miR-128 pupils. This difference was driven by AS miR-128 pupils achieving higher stereotypy scores despite their tutors' lower syntax scores. AS miR-128 birds with tutors on the higher end of the stereotypy spectrum were more likely to produce songs with faster tempos relative to sibling controls. Our results suggest that low levels of miR-128 facilitate vocal sequence stereotypy. By analogy, reducing miR-128 could enhance the capacity to learn to speak in patients with non-verbal ASD. To our knowledge, this study is the first to directly link miR-128 to learned vocal communication and provides support for miR-128 as a potential therapeutic target for ASD.

Keywords: microRNA, songbird, neurodevelopment, learning, memory, stereotypy

OPEN ACCESS

Edited by:
Sara Anna Bonini,
University of Brescia, Italy

Reviewed by:
Li Wang,
Children's National Hospital, United States
Jasmine L. Loveland,
Max Planck Institute for Ornithology, Germany
Manfred Gahr,
Max Planck Society, Germany

***Correspondence:**
Caitlin M. Aamodt
caamodt@health.ucsd.edu
Stephanie A. White
sawhite@ucla.edu

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *11 December 2021*
Accepted: *02 February 2022*
Published: *25 February 2022*

Citation:
Aamodt CM and White SA (2022) Inhibition of miR-128 Enhances Vocal Sequence Organization in Juvenile Songbirds. Front. Behav. Neurosci. 16:833383. doi: 10.3389/fnbeh.2022.833383

INTRODUCTION

The zebra finch songbird is an important model for understanding the cellular and molecular basis of learned vocal communication. Songbirds and humans share cortico-striato-thalamic circuits with convergently specialized gene expression patterns that underlie learned vocal communication (Pfenning et al., 2014). Young male zebra finches learn to communicate a unique courtship song

during a developmental critical period, and after this plasticity window closes it is much more difficult for a bird to modify his song (Aamodt et al., 2020). During this period, young males learn to sing sequences of individual syllables that are gradually organized into patterns called "motifs." Females of this species do not sing but may respond to a courtship song with innate calls. Female zebra finches strongly prefer stereotyped motifs compared to song with more variable organization (Woolley and Doupe, 2008), and young birds put forth effort to hone their songs.

Since the sequencing of the zebra finch genome (Warren et al., 2010) it has become apparent that many genes that affect social communication in humans are also dynamically regulated when a bird sings (Hilliard et al., 2012; Burkett et al., 2018). Previous work from our group identified modules of genes that are co-regulated by singing (Hilliard et al., 2012). Within this dataset Activity-Regulated Phosphoprotein 21 (ARPP21) and R3H Domain Containing 1 (R3HDM1) are two of the top genes most strongly correlated to singing. ARPP21 and R3HMD1 are host genes for microRNA-128. This led us to hypothesize that miR-128 could be a key regulator of activity-dependent gene expression in learned vocalization.

microRNAs are small (21-22 nucleotide) sequences that typically regulate gene transcripts. One of the most highly expressed microRNAs in the brain, miR-128, is elevated in postmortem cerebellar tissue from autism spectrum disorder (ASD) patients and is among the top ten microRNAs in a co-regulated module abberantly upregulated in postmortem ASD cortex (Wu et al., 2016), suggesting that it plays a role in vocal communication. This microRNA regulates both neural proliferation and neuronal migration (Bruno et al., 2011; Franzoni et al., 2015). Loss of Fragile-X mental retardation protein (FMRP) leads to elevation of miR-128 (Men et al., 2020), indicating that increased miR-128 may be a convergent mechanism shared between idiopathic and syndromic forms of ASD. Prior studies linking miR-128 to cognition and behavior further support a role for miR-128 in human and songbird vocal learning. In the mouse infralimbic prefrontal cortex miR-128 inhibition decreased fear memory extinction, whereas miR-128 overexpression promoted the extinction of the memory (Lin et al., 2011). miR-128 levels in the striatum also regulate motor behavior (Tan et al., 2013). Lower miR-128 levels were associated with greater neuronal excitability and increased motor activity, whereas higher levels suppressed neuronal activity and associated motor behavior. The authors attributed these findings to miR-128-driven changes in ion channels as well as in the Extracellular Regulated Kinase 2 (ERK2) pathway, which is also important for birdsong (Burkett et al., 2018) and ASD (Gazestani et al., 2019).

Additional support for the role of miR-128 in social vocalization comes from rodent studies of ultrasonic vocalization (USV), a behavior commonly studied in rodent pups who have been isolated from their mother. Although to our knowledge USVs have not been studied in mice with miR-128 directly modified, there are several rodent studies where reduction of miR-128 target genes linked to ASD led to deficits in vocal communication. One such gene is Reelin (RELN), a secreted glycoprotein that plays an important role in neurodevelopment. RELN is a direct target of miR-128 [Evangelisti et al., 2009; Lin et al., 2011; Rev in Ching and Ahmad-Annuar (2015)]. In mice, reduction or loss of RELN leads to delays in both motor and vocal development (Ognibene et al., 2007; Romano et al., 2013). Loss of RELN's target receptors very low density lipoprotein receptor (VLDLR) and disabled 1 (DAB1) also lead to deficits in neonate USVs (Fraley et al., 2016), suggesting that this pathway plays an important role in vocal communication.

Another interesting line of evidence comes from *in utero* exposure models of developmental disabilities. These include prenatal exposure to agents that trigger inflammation to mimic maternal infection, such as LPS [Rev in Premoli et al. (2021)], and pharmaceuticals such as valproate (Gzielo et al., 2020). Exposure to these agents *in utero* leads to reduced USVs in offspring. Although LPS-induced levels of miR-128 in rodents have not been directly measured, studies in multiple cell lines suggest that LPS dramatically increases miR-128 levels (Shyamasundar et al., 2018; Xie et al., 2021). Valproate has also been shown to increase miR-128 in cultured human neurons (Kidnapillai et al., 2020), suggesting that this molecular mechanism could be common to ASD emerging from diverse etiologies. However, the role of miR-128 in learned vocal communication has yet to be explored.

Here we use a small interfering RNA (siRNA) antisense construct to inhibit miR-128 in the striatopallidum of developing zebra finches. We show that inhibition of miR-128 leads to enhanced motif sequence stereotypy 45 days after surgery relative to controls that received a scramble sequence. We then characterize additional effects of miR-128 inhibition on spectral and temporal features of song. Our results indicate that miR-128 plays an important role in learned vocal communication and may thus be a viable therapeutic target for disorders of social communication where miR-128 is aberrantly elevated, including ASD.

MATERIALS AND METHODS

Enrichment Map

1,215 predicted miR-128 targets from TargetScan (Agarwal et al., 2015) were analyzed for pathway enrichment using g:Profiler (Raudvere et al., 2019), including the biological processes of Gene Ontology (Ashburner et al., 2000; The Gene Ontology Consortium, 2021) and molecular pathways of Reactome (Jassal et al., 2020). The resulting Generic Enrichment Map file was visualized using the Enrichment Map application (Merico et al., 2010) in Cytoscape (Shannon et al., 2003).

Hypergeometric Overlap Gene Enrichment Test

Gene enrichment analysis (**Figure 1A**) was performed to identify lists of genes enriched in predicted miR-128 targets using the GeneOverlap package in R.[1] The Simons Foundation Autism Research Initiative (SFARI Gene) database provided a comprehensive list of ASD candidate genes

[1]https://github.com/shenlab-sinai/geneoverlap

(Abrahams et al., 2013).[2] The analysis included all available genes from the SFARI database and SFARI genes found in adult songbird Area X (Hilliard et al., 2012; Burkett et al., 2018). We also looked for overlap between genes affected by high effect size mutations that cause ASD and miR-128 targets. These included chromodomain-helicase-DNA-binding protein 8 targets at enhancers [CHD8- enhancers] and promoters [CHD8- promoters] in human fetal cortex (Cotney et al., 2015), and Fragile-X mental retardation protein target genes [FMRP] from P11–P25 mouse whole brain extracts (Darnell et al., 2011). We also included genes differentially expressed in Down Syndrome [Down Syndrome] postmortem brain using samples from multiple regions spanning mid-fetal development to adulthood (Olmos-Serrano et al., 2016) as a negative control. As a second control we included all genes expressed in adult songbird Area X. In these datasets we tested for enrichment in predicted miR-128 target genes, mRNA binding protein (and miR-128 host gene), ARPP21 target genes, and genes targeted by both miR-128 and ARPP21 (Rehfeld et al., 2018). The background was limited to 14,313 songbird brain-expressed genes. Statistically significant gene overlap was calculated using Fisher's Exact Test followed by FDR correction.

microRNA Assay

Upon extraction whole brains were immediately flash frozen using liquid nitrogen and stored at −80°C until processing. Brains were then allowed to raise to −20°C in the cryostat, then mounted whole to a chuck using TissueTek (Sakura) mounting medium. Brains were sectioned coronally until reaching the beginning of Area X. Tissue punches were collected from Area X using a 20 gauge Luer adapter inserted 1 mm deep. Total RNA (including small RNA) was isolated using a TRIzol Reagent kit (Zymo Research). Taqman MicroRNA Assays specific for miR-128 (assay ID: 002216) and U6 snRNA (assay ID: 001973) as a loading control were used to reverse- transcribe microRNA using and the TaqMan microRNA reverse transcription kit (Applied Biosystems, Waltham, MA, United states). Five ng of RNA were added to each 15 uL reverse transcription reaction. Expression was analyzed using the TaqMan microRNA assay kit (Applied Biosystems, Waltham, MA, United states) with TaqMan Fast Advanced Master Mix according to the manufacturer's instructions. 1.33 uL cDNA template, or nuclease free water for no template controls, was added to each well. All reactions were run in triplicate. The thermal protocol was set to one cycle at 50°C for 2 min for UNG activation, one cycle at 95°C for 20 s for enzyme activation, then 40 cycles at 95°C for 1 s to denature followed by 20 s at 60°C for annealing/extension. miR-128 levels were normalized to U6 as a loading control. Liver small RNAs were used as a negative control, since miR-128 is expressed at very low levels (Luo et al., 2012).

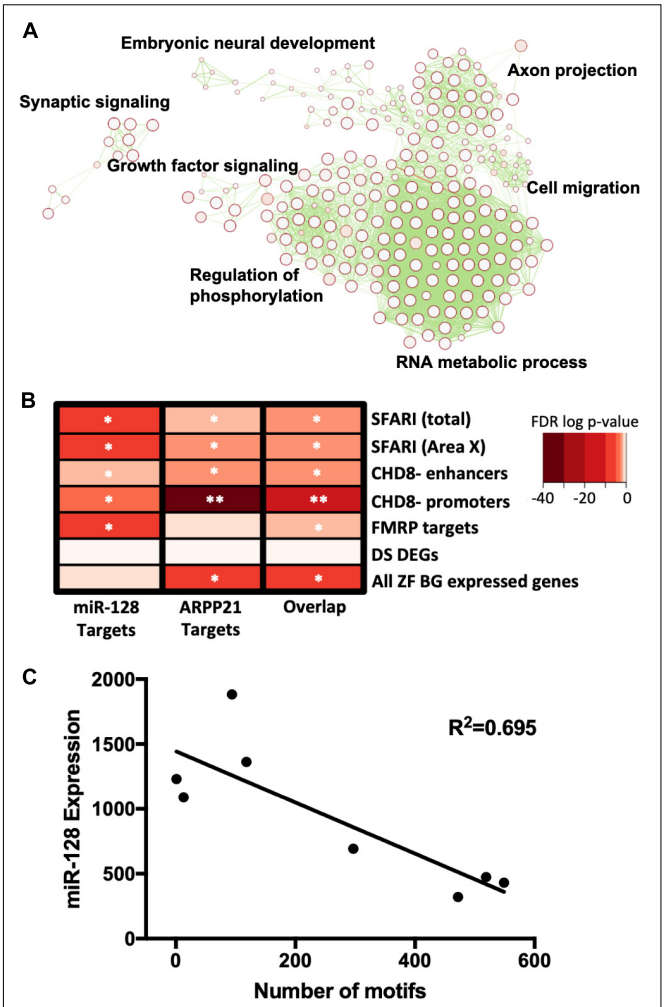


FIGURE 1 | (A) Predicted miR-128 targets were analyzed for pathway enrichment and then clustered by similarity to generate an enrichment map. Cluster annotations are summarized with adjacent labels in black. **(B)** miR-128 targets are enriched in ASD-related gene datasets. Heat map shows gene enrichment analysis for miR-128 target genes, ARPP21 target genes, or both [Overlap] in autism-related gene lists: all autism risk genes from the Simons Foundation Autism Research Initiative (SFARI) database [SFARI (total)], SFARI genes expressed in Area X [SFARI (Area X)], CHD8 targets at enhancers [CHD8- enhancers] and promoters [CHD8- promoters], Fragile-X mental retardation protein target genes [FMRP targets]. Neural genes differentially expressed in Down syndrome [DS DEGs] and all genes expressed in the adult zebra finch basal ganglia [All ZF BG expressed genes] are included as negative controls. FDR corrected *p*-value indicated by $*p < 5.00E{-}02$, $**p < 5.00E{-}12$. **(C)** Activity-dependent regulation of miR-128 in adult zebra finch Area X as measured by qPCR. miR-128 undergoes activity-dependent regulation in the vocal communication region of the basal ganglia in adult songbirds ($R^2 = 0.695$, $p = 0.01$). miR-128 levels were normalized to loading control U6 snoRNA.

Subjects

Subjects were male zebra finches (*Taeniopygia guttata*). To measure activity-dependent regulation of gene expression adult birds (>125 d) were allowed to sing in a sound attenuation chamber for 2 h prior to tissue collection by cervical dislocation and brain extraction. A total of eight birds were used. For the miR-128 inhibition experiments juvenile birds beginning at 30 days post-hatch (30 d) were selected for surgery and subsequently raised to 75 d. A total of 20 birds underwent stereotaxic neurosurgeries targeting Area X bilaterally (see

[2]gene.sfari.org

below). Ten birds from seven breeding pairs were treated by focal injection of an AAV bearing a miR-128 sponge sequence and 10 siblings were treated with a scrambled sequence as a control. Birds were primarily housed in home cages with parents and siblings, unless being recorded individually in a sound attenuation chamber. The vivarium and recording chambers are humidity- and temperature-controlled (22°C) and on a 12 h light: dark cycle with half hour "dusks" and "dawns." Birdseed, water, millet, cuttlebone, and grit were provided *ad libitum*. Baths, hardboiled egg, and vegetables were provided weekly. Animal use was in accordance with the Institutional Animal Care and Use Committee at the University of California, Los Angeles and complied with the American Veterinary Medical Association Guidelines.

Experimental Timeline

Experiments were performed as schematized in **Figure 2A**. Subjects were housed in their home cage with both parents except during short periods to record song individually. At 30 d surgeries were performed and the birds were allowed to recover in their home. At ∼70 d birds were returned to recording chambers for behavioral experiments. Birds were given a day or two to habituate to the chamber, recorded for undirected singing the next day, prevented from singing for 2 h and then allowed to sing and recorded for 2 h, allowed to sing normally for 1 day, and then the final recording was collected and the bird was sacrificed 2 h after lights on.

siRNA Constructs and Viral Vectors

An siRNA sponge previously shown to inhibit miR128 function (Lin et al., 2011) and a scrambled control construct were adapted for use in zebra finches and cloned into an AAV1 viral vector (Virovek, Hayward, CA, United states). The sponge consisted of three "bulged" miR-128 binding site sequences separated by 4nt spacers. The scramble construct differed only in that the control sequence contained a standard scramble sponge provided by Virovek. Both sponge and scramble sequences were driven by the U6 promoter, as previously described (Haesler et al., 2007). Both viral titers were 10^13 vg/ml. 303.6 nl of virus were delivered to each hemisphere in six injections of 50.6 nl spaced 5 um apart.

Stereotaxic Surgery

At 30 d, neurosurgeries were performed as described (Heston and White, 2015; Burkett et al., 2018). Briefly, subjects were administered preoperative meloxicam and then anesthetized using 2–4.5% isoflurane in oxygen. A custom-built avian stereotaxic apparatus was used to target Area X at the following coordinates: 45° head angle, 5.15 mm rostral of the bifurcation of the midsagittal sinus, 1.60 mm lateral of the midline, at a depth of 3.25 mm relative to the surface of the exposed brain. Virus was injected bilaterally using a Drummond Nanoject II *via* a glass micropipet (∼40 um inner diameter). Six 50.6 nL injections were performed in Area X at a depth of 3.25, 3.20, 3.15, 3.10, 3.05, and 3.0 mm to maximize the area transfected. Injections were performed with a 15 s pause to allow virus to disperse, with a final 5-min wait before removing the micropipet. The scalp incision was closed using Vetbond (3M, St. Paul, MN, United States), then covered with dental cement. Birds remained on oxygen for ∼2 min until alert, then returned to their home cages.

Audio Recording

Recordings were collected in sound attenuation chambers. Songs were digitally recorded (Sound Analysis Pro; SAP; Tchernichovski et al., 2000) using a PreSonus FirePod or Audiobox (44.1 kHz sampling rate; 24-bit depth) and stored uncompressed.

Sequence Stereotypy

Sequence stereotypy for juveniles and tutors was measured by calculating syllable transition probability using Vocal Inventory Clustering Engine [VoICE; (Burkett et al., 2015)].[3] Syllables from the first 20 motifs were hand segmented and a SAP Feature Batch was generated, producing a table of the quantified values for all spectral features for each syllable (Tchernichovski et al., 2000). Motifs were defined as sequences of recurrently produced syllables comprised of at least three syllables, though all birds in this study produced greater than three syllables per motif. Syllables are defined as discrete acoustic elements (Yu and Margoliash, 1996). We segmented syllables at the milliseconds of silence that separate each syllable. In cases where two syllables were conjoined, we chose to lump into a single syllable or split into two syllables based on how they most frequently appeared in the bird's song."

Then VoICE was used to cluster syllables by similarity. Once syllable types were defined, VoICE was used to calculate syllable transition probability across all motifs in the order they were produced and generate a stereotypy score.

NS-UD Paradigm

"NS-UD" experiments were performed as follows and as described in **Figure 3B**, as well as prior publications (Miller et al., 2010; Chen et al., 2013; Heston and White, 2015; Burkett et al., 2018). Briefly, on one morning, individual birds were distracted if they appeared to initiate singing during the first 2 h after lights on, then recorded to capture song after the non-singing (NS) state. No birds sang >10 motifs in this condition. Prior work shows that the gene expression profiles of NS birds are similar to those of birds that voluntarily did not sing in enclosed chambers, suggesting that this paradigm does not induce a stress response (Miller et al., 2010). NS songs were compared to songs from the same birds that were recorded on a different day following 2 h of singing in a single-housed chamber (undirected singing: UD). Syllables from the first 20 motifs sung in these two conditions were hand segmented and acoustic features were quantified in a SAP Feature Batch. Syllables were clustered using VoICE. To calculate the effect size, we used the formula (NS-UD)/NS + UD).

Song Tempo

Tempo was defined as the number of syllables per motif divided by the duration of the motif in seconds. Tempo scores were averaged across the first 20 motifs for each bird. Tempo data was divided into two bins based on the syntax score each bird's

[3]https://github.com/zburkett/VoICE

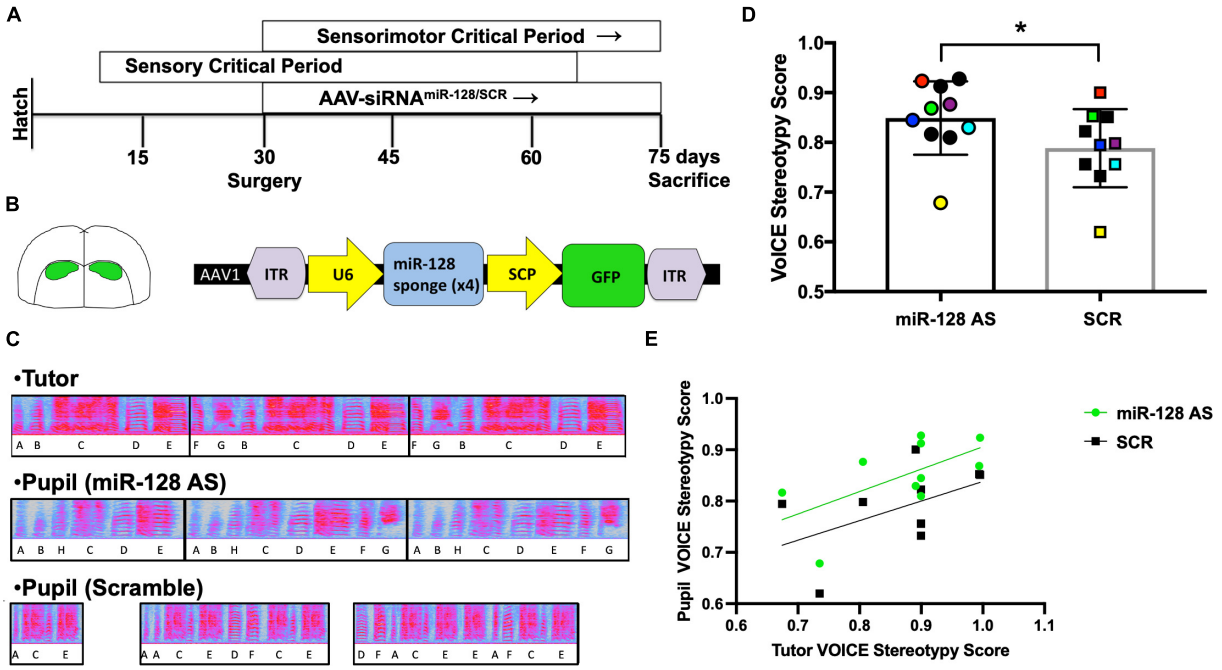


FIGURE 2 | (A) Timeline of experimental procedures. Developmental critical periods are shown in top two boxes. **(B)** Region of interest (Area X) and miR-128 antisense viral construct. (Left) Diagram showing viral targeting of Area X. (Right) Schematic characterizing the antisense construct used to inhibit miR-128. Adeno-associated virus 1 vector (AAV1) with inverted terminal repeats to maintain construct integrity (ITR). The type III RNA polymerase III promoter U6 was used to drive expression of the miR-128 siRNA sponge. A proprietary supercore promotor designed by Rachael Neve was used to drive GFP. **(C)** Exemplar motifs from an adult male (Tutor), a juvenile son who received the miR-128 antisense construct [Pupil (miR-128 AS)], and another juvenile son from the same clutch who received the scramble control construct [Pupil (Scramble)]. **(D)** Analysis of song sequence organization in 75 d sibling-matched miR-128 AS and scramble control birds (N = 10/group), $^*p$ = 0.014, Wilcoxon signed-rank test. Symbol colors denote pupils that shared the same tutor. **(E)** Tutor and control pupil syntax stereotypy scores are positively correlated (R^2 = 0.49, $^*p$ = 0.025), whereas the correlation between tutors and AS miR-128 pupils does not reach significance (R^2 = 0.37, p = 0.061).

tutor. Since there was an odd number of tutors, the tutor with the median score had two pupils in each of the four bins based on their own scores. The final N included five birds in each bin, 20 birds total.

RESULTS

We began our investigation by broadly determining what functional roles miR-128 targets share. Since miR-128 and many of its targets are highly conserved in vertebrates (Bruno et al., 2011), we wanted to better understand what processes miR-128 regulates. Using predicted targets (Rehfeld et al., 2018) we generated an enrichment map to identify clusters of genes with shared functional roles (**Figure 1A** and **Supplementary Table 1**). Several key processes were implicated, including embryonic neural development, axon projection, cell migration, and growth factor signaling. Regulation of phosphorylation was also strongly implicated, suggesting converging effects on signaling enzyme cascades central to birdsong (Burkett et al., 2018) and autism (Gazestani et al., 2019).

We next investigated whether miR-128 targets would be enriched in gene lists relevant for human ASD using a hypergeometric gene overlap test (**Figure 1B** and **Supplementary Table 2**). miR-128 targets were enriched in the Simons Foundation Autism Research Initiative's (SFARI) list of high confidence syndromic ASD genes (with a confidence score of 3 or greater), using both the entire list or restricting the list to those genes expressed in Area X. We also found that miR-128 targets were enriched in genes regulated by syndromic ASD gene chromodomain helicase DNA binding protein 8 (CHD8) at either its enhancers or promoters. Similarly, we found an enrichment of miR-128 targets among genes regulated by syndromic ASD gene and RNA binding protein Fragile X Mental Retardation Protein (FMRP). As a negative control, we checked for enrichment of miR-128 targets in Down Syndrome differentially expressed genes and found that they were not enriched, as expected. As an additional negative control, we checked for enrichment of miR-128 targets across all genes expressed in the adult zebra finch striatopallidum and did not find an enrichment. We interpret these results to suggest that autism severity is mediated by multiple hits that converge on shared pathways, and that one of these hits that influences other changes that predispose ASD is the level of miR-128 expression.

Next, we checked to see how the amount of singing, measured by total number of motifs produced during the first 2 h of lights-on, affects miR-128 expression in Area X (**Figure 1C**). Using tissue from eight adult (>120 d) males, we found that total number of motifs was negatively correlated to miR-128

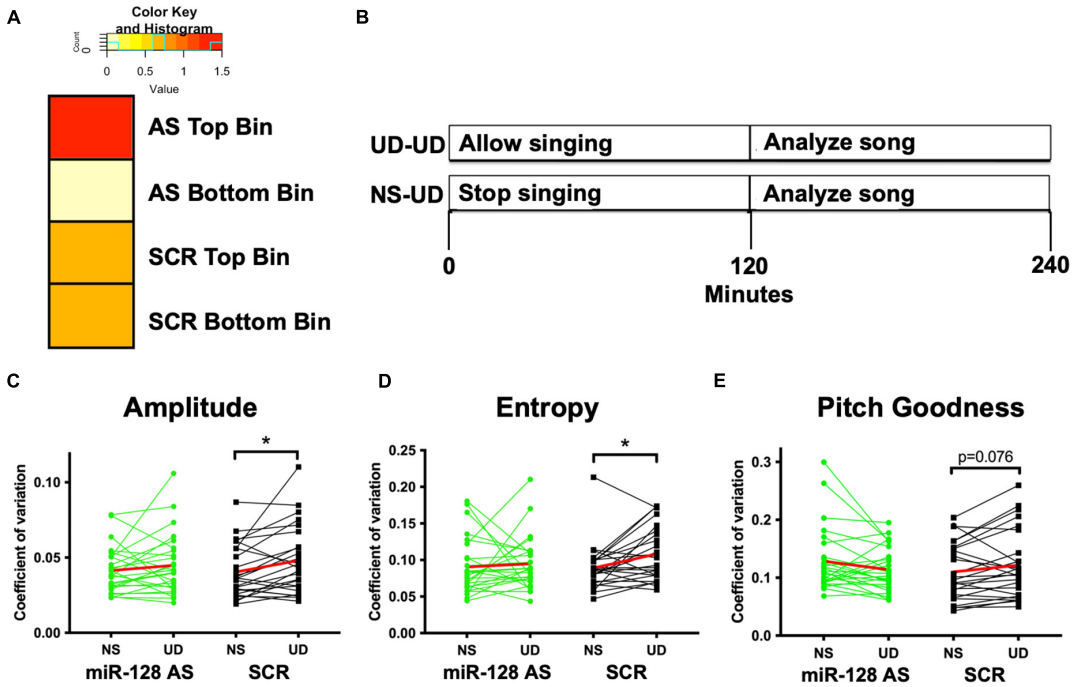


FIGURE 3 | (A) Tempo data divided into two bins based on the syntax score each bird's tutor. AS miR-128 birds from tutors with higher syntax scores had faster tempos (tempo score < 0.13, Odds Ratio 1.5) relative to AS miR-128 birds from tutors with lower syntax scores (OR 0), or SCR birds from both bins (OR 0.67). **(B)** Schematic illustrating the non-singing vs. singing (NS-UD) paradigm. On separate days a test subject is either allowed to sing by himself for 4 h (UD-UD) or is prevented from singing for 2 h and then allowed to sing for the subsequent 2 h (NS-UD). Our prior work revealed that song variability is positively correlated to the amount of time singing (Heston and White, 2015; Burkett et al., 2018). **(C–E)** Coefficient of variation (CV) for spectral features [**(C)** Amplitude, **(D)** Entropy, **(E)** Pitch Goodness] of song following 2 h of singing (UD) or non-singing (NS). The experimental AS miR-128 group is shown in green, and the control SCR group is in black. The average trendline is shown in red. Syllables from five AS miR-128 birds (N = 26) and five sibling-matched control SCR birds(N = 25) were analyzed, $^*p < 0.05$ bootstrap t-test.

levels (R^2 = 0.695, p = 0.01). This suggests that miR-128 is downregulated in Area X in an activity-dependent manner.

To better understand how miR-128 influences learned vocal communication we injected an antisense miR-128 (AS miR-128) siRNA construct or a scramble control sequence into Area X (**Figure 2A**, N = 10 per group) of sibling matched pairs at 30 d, then raised the birds with both parents in their home cage. At 75 d we assessed how song quality was influenced (**Figure 2B**). Sequence stereotypy was measured by calculating syllable transition probability using Vocal Inventory Clustering Engine [VoICE; (Burkett et al., 2015); See text footnote 3]. Syllables were hand segmented and a SAP Feature Batch was generated, producing a table of the quantified values for all spectral features for each syllable (Tchernichovski et al., 2000). Then VoICE was used to cluster syllables by similarity. Once syllable types were defined, VoICE was used to calculate syllable transition probability and generate a stereotypy score. Strikingly, syllable sequence organization was significantly improved in juvenile birds who received the miR-128 antisense construct relative to sibling-matched controls raised under normal conditions (**Figures 2C,D**).

AS pupil sequence stereotypy scores correlated with their tutors' scores (**Figure 2E**), although this correlation did not reach significance in AS miR-128 pupils due to higher syntax scores at the lower end of the score range. To better understand the ceiling effect for pupils at the higher end of the spectrum we calculated tempo for each pupil as previously described (Mets and Brainard, 2018). Overall, we found no differences in tempo between groups using a t-test with resampling or Wilcoxen rank sum test. However, when the birds were binned by tutor syntax score, birds in the higher bin for AS miR-128 were more likely to have faster tempos (OR = 1.5, **Figure 3A**). We also counted the total number of motifs produced over the course of 2 h in the 75 d miR-128 AS and SCR sibling pairs (**Supplementary Figure 1**). Although not significant, there was an interesting trend of miR-128 AS birds with siblings who sang a low number of motifs singing more than their counterparts, except for the birds whose SCR siblings sang the highest number of motifs. This suggests that there may be a limit to how much reducing miR-128 levels promotes song production. Future research should focus on birds with impoverished song to better characterize the potential therapeutic value of miR-128 inhibition.

To further characterize the effects of the miR-128 sponge on learned vocal behavior we employed the "NS-UD paradigm" as follows (**Figure 3B**; Miller et al., 2010; Chen et al., 2013; Heston and White, 2015; Burkett et al., 2018). When a bird sings by himself for 2 h in the morning (undirected) the spectral features of birdsong become more variable in correlation to the amount of time the bird has been singing (Miller et al., 2010). On 1 day a bird was recorded after 2 h of singing, while on a separate day

the same bird was recorded after 2 h of non-singing. Analysis of the coefficient of variation for the spectral features on both days showed that birds who received the scrambled control had more variable song in the UD vs. NS condition, in keeping with previous findings. However, the miR-128 antisense birds did not show this same increase in spectral feature variability. Results reached significance for amplitude and entropy and approached significance for pitch goodness using a *t*-test with resampling (**Figures 3C–E** and **Supplementary Figure 2**).

Motif level analysis of syllable sequence stereotypy did not uncover differences between the NS and UD conditions; however, it did replicate the results in **Figure 2D** with independent data sets, showing that the result was consistent across days and robust even after 2 h of singing when song is typically less stereotyped (**Supplementary Figure 3**).

DISCUSSION

Here we show that inhibition of miR-128 during juvenile vocal development enhances song stereotypy. We did this by injecting a miR-128 antisense construct delivered *via* an AAV1 vector in Area X of 30 d birds and analyzing song at 75 d relative to sibling matched controls who received a construct with a scrambled sequence. Overall stereotypy scores for all pupils were correlated to their tutors' scores, as expected. Strikingly, inhibition of miR-128 consistently led to enhancements in sequence stereotypy relative to sibling-matched controls. AS miR-128 birds from tutors with high stereotypy scores were more likely to produce songs with faster tempos. Although SCR controls showed more variability in spectral features after 2 h of singing in solitude, as is typical in this species, AS miR-128 birds did not show a typical loss of stereotypy in this time frame.

Zebra finches are the key model system for studying how activity-dependent gene regulation underlies developmental vocal communication learning (e.g., Haesler et al., 2007; Heston and White, 2015). Mechanisms that limit or constrain plasticity for vocal learning are of particular interest, because premature closure could be a major contributor to communication deficits in humans. Consistent with previous results, we found that miR-128 was highly expressed in zebra finch brain (**Figure 1C**; Luo et al., 2012), similarly to humans (Bruno et al., 2011). We show that reducing miR-128, which is aberrantly elevated in ASD patients, improves complex vocal sequences in 75 d zebra finches (**Figure 2**). These results show that miR-128 plays a role in constraining learned vocal communication, and thus miR-128 may be a promising pharmacotherapeutic target for treating ASD.

Our work is in keeping with previous studies showing that microRNAs regulate transcription in learned vocal communication. Shi and colleagues found that miR-9 and miR-140-5p are upregulated when a bird practices his song by himself and regulate key targets important for learned vocalization, including the speech and language-related ForkheadboxP2 transcription factor (FoxP2; Shi et al., 2013; Shi et al., 2018). Similarly, Larson et al. (2015) characterized changes in microRNAs associated with seasonal song plasticity in Gambel's white-crowned sparrows. They found that miR-212 and miR-132 were key regulators of seasonal plasticity, which is interesting given that these microRNAs, like miR-128, are also aberrantly elevated in ASD (Wu et al., 2016). Within the auditory forebrain, microRNAs are regulated by passive listening to conspecific song (Gunaratne et al., 2011). Several studies have also identified sex differences in songbird microRNA regulation (Luo et al., 2012; Lin et al., 2014).

In rodents and human cells, miR-128 has been linked to both neuroplasticity and neurodevelopment. That leads us to think that here miR-128 may be regulating ongoing postnatal neurogenesis in Area X. Interestingly, songbirds and humans share ongoing basal ganglia neurogenesis in adulthood (Vellema et al., 2010; Thompson et al., 2013; Ernst et al., 2014), so this adaptation may be common to vocal learning species.

Interestingly, in the birds that received the AS miR-128 construct we did not see changes in the NS-UD paradigm that are typically associated with prolonged singing (**Figure 3**). This raises interesting questions about the evolution of learned vocal communication. Perhaps the evolution of stereotyped vocal sequence production starts with a smaller window when an organism can perform this behavior, and then evolution selects for individuals with greater endurance for producing learned vocalizations in a precise manner. Although some variability is necessary for any form of motor learning (Fee, 2014), it may be that the required changes can be more subtle and decrease as the bird learns, similarly to how juvenile birds produce more variable song than adults. It was recently discovered that miR-128 has undergone positive selection in the human lineage (Wang et al., 2020). In humans, multiple single nucleotide polymorphisms that reduce the expression of ARPP21 correlate with higher intelligence scores (Savage et al., 2018). Altogether, this suggests that miR-128 may be a key for understanding the evolution of language.

CONCLUSION

In conclusion, these results reveal that expression levels of miR-128 in the basal ganglia directly affect sequence stereotypy in learned vocal communication. This study provides evidence that therapeutically targeting miR-128 during juvenile development has beneficial effects on vocal learning in an animal model of speech and language. Inhibition of miR-128 may be a safe approach for future therapeutic design for children with ASD and other developmental disabilities that impair speech and language development.

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 the Institutional Animal Care and Use Committee at the University of California,

Los Angeles and complied with the American Veterinary Medical Association Guidelines.

AUTHOR CONTRIBUTIONS

CA and SW conceived and designed the study. CA coordinated and carried out all behavioral components and the molecular work, performed the statistical analysis, and drafted the manuscript. SW helped draft the manuscript. Both authors gave final approval for publication.

FUNDING

This work was supported by a Predoctoral (T32MH073526) and a Postdoctoral (T32HD087978) Ruth L. Kirschstein Institutional National Research Service Award, a Ruth L. Kirschstein Predoctoral Individual National Research Service Award F31MH110209 (CA), and NIH R01 MH070712 and UCLA Faculty Research Grants (SW).

ACKNOWLEDGMENTS

We thank rotation student Jacquie Weiss, and undergrads Carolina Fuentes and Claude Andres for aiding the efforts of these experiments. We also thank Nancy Day and Zach Burkett for helpful discussions and committee members Dan Geschwind, Pam Kennedy, and Marcelo Wood for valuable feedback on this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2022.833383/full#supplementary-material

Supplementary Figure 1 | (A–H) Coefficient of variation (CV) for spectral features [**(A)** Duration, **(B)** Amplitude, **(C)** Pitch, **(D)** Frequency Modulation, **(E)** Amplitude Modulation, **(F)** Entropy, **(G)** Pitch Goodness, **(H)** Mean Frequency] of song following 2 h of singing (UD) or non-singing (NS). The experimental AS miR-128 group is shown in green, and the control SCR group is in black. The average trendline is shown in red. Syllables from five AS miR-128 birds (N = 26) and five sibling-matched control SCR birds (N = 25) were analyzed.

Supplementary Figure 2 | Motif level analysis of syllable sequence stereotypy in NS and UD conditions.

Supplementary Figure 3 | Number of motifs produced during 2 h of singing in 75 d miR-128 AS and SCR controls. Symbol colors denote pupils that shared the same tutor.

Supplementary Table 1 | Gene set enrichment analysis results used to generate **Figure 1A**.

Supplementary Table 2 | Gene lists used to generate **Figure 1B**.

REFERENCES

Aamodt, C. M., Farias-Virgens, M., and White, S. A. (2020). Birdsong as a window into language origins and evolutionary neuroscience. *Philos. Trans. R. Soc.Lond. B. Biol. Sci.* 375:20190060. doi: 10.1098/rstb.2019.0748

Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. A., et al. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* 4:36. doi: 10.1186/2040-2392-4-36

Agarwal, V., Bell, G. W., Nam, J., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005. doi: 10.7554/eLife.05005

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bruno, I. G., Karam, R., Huang, L., Bhardwaj, A., Lou, C. H., Shum, E. Y., et al. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell* 42, 500–510. doi: 10.1016/j.molcel.2011.04.018

Burkett, Z. D., Day, N. F., Kimball, T. H., Aamodt, C. M., Heston, J. B., Hilliard, A. T., et al. (2018). FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. *eLife* 7:e30649. doi: 10.7554/eLife.30649

Burkett, Z. D., Day, N. F., Penþagarikano, O., Geschwind, D. H., and White, S. A. (2015). VoICE: a semi- automated pipeline for standardizing vocal analysis across models. *Sci. Rep.* 5:10237. doi: 10.1038/srep10237

Chen, Q., Heston, J. B., Burkett, Z. D., and White, S. A. (2013). Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. *J. Exp. Biol.* 216, 3682–3692. doi: 10.1242/jeb.085886

Ching, A. S., and Ahmad-Annuar, A. (2015). A Perspective on the role of microRNA-128 regulation in mental and behavioral disorders. *Front. Cell. Neurosci.* 9:465. doi: 10.3389/fncel.2015.00465

Cotney, J., Muhle, R. A., Sanders, S. J., Liu, L., Willsey, A. J., Niu, W., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* 6:6404. doi: 10.1038/ncomms7404

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y., Mele, A., Fraser, C. E., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261. doi: 10.1016/j.cell.2011.06.013

Ernst, A., Alkass, K., Bernard, S., Salehpour, M., Perl, S., Tisdale, J., et al. (2014). Neurogenesis in the striatum of the adult human brain. *Cell* 156, 1072–1083. doi: 10.1016/j.cell.2014.01.044

Evangelisti, C., Florian, M. C., Massimi, I., Dominici, C., Giannini, G., Galardi, S., et al. (2009). MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness. *FASEB J.* 23, 4276–4287. doi: 10.1096/fj.09-134965

Fee, M. S. (2014). The role of efference copy in striatal learning. *Curr. Opin. Neurobiol.* 25:194. doi: 10.1016/j.conb.2014.01.012

Fraley, E. R., Burkett, Z. D., Day, N. F., Schwartz, B. A., Phelps, P. E., and White, S. A. (2016). Mice with Dab1 or Vldlr insufficiency exhibit abnormal neonatal vocalization patterns. *Sci. Rep.* 6:25807. doi: 10.1038/srep25807

Franzoni, E., Booker, S. A., Parthasarathy, S., Rehfeld, F., Grosser, S., Srivatsa, S., et al. (2015). miR-128 regulates neuronal migration, outgrowth, and intrinsic excitability via the intellectual disability gene Phf6. *eLife* 4:e04263. doi: 10.7554/eLife.04263

Gazestani, V. H., Pramparo, T., Nalabolu, S., Kellman, B. P., Murray, S., Lopez, L., et al. (2019). A perturbed gene network containing PI3K-AKT, RAS-ERK and WNT-β- catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat. Neurosci.* 22, 1624–1634. doi: 10.1038/s41593-019-0489-x

Gunaratne, P. H., Lin, Y. C., Benham, A. L., Drnevich, J., Coarfa, C., Tennakoon, J. B., et al. (2011). Song exposure regulates known and novel microRNAs in the zebra finch auditory forebrain. *BMC Genomics* 12:277. doi: 10.1186/1471-2164-12-277

Gzielo, K., Potasiewicz, A., Hołuj, M., Litwa, E., Popik, P., and Nikiforuk, A. (2020). Valproic acid exposure impairs ultrasonic communication in infant, adolescent and adult rats. *Eur. Neuropsychopharmacol.* 41, 52–62. doi: 10.1016/j.euroneuro.2020.09.006

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in

songbird basal ganglia nucleus Area X. *PLoS Biol.* 5:e321. doi: 10.1371/journal.pbio.0050321

Heston, J. B., and White, S. A. (2015). Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *J. Neurosci.* 35, 2885–2894. doi: 10.1523/JNEUROSCI.3715-14.2015

Hilliard, A. T., Miller, J. E., Fraley, E. R., Horvath, S., and White, S. A. (2012). Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron* 73, 537–552. doi: 10.1016/j.neuron.2012.01.005

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503. doi: 10.1093/nar/gkz1031

Kidnapillai, S., Wade, B., Bortolasci, C. C., Panizzutti, B., Spolding, B., Connor, T., et al. (2020). Drugs used to treat bipolar disorder act via microRNAs to regulate expression of genes involved in neurite outgrowth. *J. Psychopharmacol.* 34, 370–379. doi: 10.1177/0269881119895534

Larson, T. A., Lent, K. L., Bammler, T. K., MacDonald, J. W., Wood, W. E., Caras, M. L., et al. (2015). Network analysis of microRNA and mRNA seasonal dynamics in a highly plastic sensorimotor neural circuit. *BMC Genomics* 16:905. doi: 10.1186/s12864-015-2175-z

Lin, Q., Wei, W., Coelho, C. M., Li, X., Baker-Andresen, D., Dudley, K., et al. (2011). The brain-specific microRNA miR-128b regulates the formation of fear-extinction memory. *Nat. Neurosci.* 14, 1115–1117. doi: 10.1038/nn.2891

Lin, Y. C., Balakrishnan, C. N., and Clayton, D. F. (2014). Functional genomic analysis and neuroanatomical localization of miR-2954, a song-responsive sex-linked microRNA in the zebra finch. *Front. Neurosci.* 8:409. doi: 10.3389/fnins.2014.00409

Luo, G. Z., Hafner, M., Shi, Z., Brown, M., Feng, G. H., Tuschl, T., et al. (2012). Genome-wide annotation and analysis of zebra finch microRNA repertoire reveal sex-biased expression. *BMC Genomics* 13:727. doi: 10.1186/1471-2164-13-727

Men, Y., Ye, L., Risgaard, R. D., Promes, V., Zhao, X., Paukert, M., et al. (2020). Astroglial FMRP deficiency cell-autonomously up-regulates miR-128 and disrupts developmental astroglial mGluR5 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 117, 25092–25103. doi: 10.1073/pnas.2014080117

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G. D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5:e1398. doi: 10.1371/journal.pone.0013984

Mets, D. G., and Brainard, M. S. (2018). Genetic variation interacts with experience to determine interindividual differences in learned song. *Proc. Natl. Acad. Sci. U.S.A.* 115, 421–426. doi: 10.1073/pnas.1713031115

Miller, J. E., Hilliard, A. T., and White, S. A. (2010). Song practice promotes acute vocal variability at a key stage of sensorimotor learning. *PLoS One* 5:e8592. doi: 10.1371/journal.pone.0008592

Ognibene, E., Adriani, W., Macrì, S., and Laviola, G. (2007). Neurobehavioural disorders in the infant reeler mouse model: interaction of genetic vulnerability and consequences of maternal separation. *Behav. Brain Res.* 177, 142–149. doi: 10.1016/j.bbr.2006.10.027

Olmos-Serrano, J. L., Kang, H. J., Tyler, W. A., Silbereis, J. C., Cheng, F., Zhu, Y., et al. (2016). Down syndrome developmental brain transcriptome reveals defective oligodendrocyte differentiation and myelination. *Neuron* 89, 1208–1222. doi: 10.1016/j.neuron.2016.01.042

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346:1256846. doi: 10.1126/science.1256846

Premoli, M., Memo, M., and Bonini, S. A. (2021). Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen. Res.* 16, 1158–1167. doi: 10.4103/1673-5374.300340

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. doi: 10.1093/nar/gkz369

Rehfeld, F., Maticzka, D., Grosser, S., Knauff, P., Eravci, M., Vida, I., et al. (2018). The RNA-binding protein ARPP21 controls dendritic branching by functionally opposing the miRNA it hosts. *Nat. Commun.* 9:1235. doi: 10.1038/s41467-018-03681-3

Romano, E., Michetti, C., Caruso, A., Laviola, G., and Scattoni, M. L. (2013). Characterization of neonatal vocal and motor repertoire of reelin mutant mice. *PLoS One* 8:e64407. doi: 10.1371/journal.pone.0064407

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* 50, 912–919. doi: 10.1038/s41588-018-0152-6

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shi, Z., Luo, G., Fu, L., Fang, Z., Wang, X. J., and Li, X. C. (2013). miR-9 and miR-140-5p target FoxP2 and are regulated as a function of the social context of singing behavior in zebra finches. *J. Neurosci.* 33, 16510–16521. doi: 10.1523/JNEUROSCI.0838-13.2013

Shi, Z., Piccus, Z., Zhang, X., Yang, H., Jarrell, H., Ding, Y., et al. (2018). miR-9 regulates basal ganglia-dependent developmental vocal learning and adult vocal performance in songbirds. *eLife* 7:e29087. doi: 10.7554/eLife.29087

Shyamasundar, S., Ong, C., Yung, L. L., Dheen, S. T., and Bay, B. H. (2018). miR-128 Regulates Genes Associated with Inflammation and Fibrosis of Rat Kidney Cells In Vitro. *Anat. Rec. (Hoboken)* 301, 913–921. doi: 10.1002/ar.23763

Tan, C. L, Plotkin, J. L., Venø, M. T., von Schimmelmann, M., Feinberg, P., Mann, S., et al. (2013). MicroRNA-128 governs neuronal excitability and motor behavior in mice. *Science* 342, 1254–1258. doi: 10.1126/science.1244193

Tchernichovski, O., Nottebohm, F., Ho, C. E., Bijan, P., and Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Anim. Behav.* 59, 1167–1176. doi: 10.1006/anbe.1999.1416

The Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49, D325–D334. doi: 10.1093/nar/gkaa1113

Thompson, C. K., Schwabe, F., Schoof, A., Mendoza, E., Gampe, J., Rochefort, C., et al. (2013). Young and intense: FoxP2 immunoreactivity in Area X varies with age, song stereotypy, and singing in male zebra finches. *Front. Neural Circuits* 7:24. doi: 10.3389/fncir.2013.00024

Vellema, M., van der Linden, A., and Gahr, M. (2010). Area-specific migration and recruitment of new neurons in the adult songbird brain. *J. Comp. Neurol.* 518, 1442–1459. doi: 10.1002/cne.22281

Wang, L., Sinnott-Armstrong, N., Wagschal, A., Wark, A. R., Camporez, J. P., Perry, R. J., et al. (2020). A MicroRNA Linking Human Positive Selection and Metabolic Disorders. *Cell* 183, 684–701.e14. doi: 10.1016/j.cell.2020.09.017

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., et al. (2010). The genome of a songbird. *Nature* 464, 757–762. doi: 10.1038/nature08819

Woolley, S. C., and Doupe, A. J. (2008). Social context-induced song variation affects female behavior and gene expression. *PLoS Biol.* 6:e62. doi: 10.1371/journal.pbio.0060062

Wu, Y. E., Parikshak, N. N., Belgard, T. G., and Geschwind, D. H. (2016). Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat. Neurosci.* 19, 1463–1476. doi: 10.1038/nn.4373

Xie, H., Chai, H., Du, X., Cui, R., and Dong, Y. (2021). Overexpressing long non-coding RNA OIP5-AS1 ameliorates sepsis-induced lung injury in a rat model via regulating the miR-128-3p/Sirtuin-1 pathway. *Bioengineered* 12, 9723–9738. doi: 10.1080/21655979.2021.1987132

Yu, A. C., and Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science* 273, 1871–1875. doi: 10.1126/science.273.5283.1871

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 Behavioral Neuroscience

METHODS
published: 04 March 2022
doi: 10.3389/fnbeh.2022.826550

Semi-Automated Training of Rat Ultrasonic Vocalizations

Aaron M. Johnson[1], Charles Lenell[2], Elizabeth Severa[1], Denis Michael Rudisch[3], Robert A. Morrison[4,5] and Adrianna C. Shembel[4,5]*

[1] Department of Otolaryngology-Head and Neck Surgery, New York University Grossman School of Medicine, New York, NY, United States, [2] Department of Surgery, University of Wisconsin-Madison, Madison, WI, United States, [3] Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States, [4] Department of Speech Language and Hearing, School of Behavioral and Brain Sciences, UT Dallas, Dallas, TX, United States, [5] Department of Otolaryngology-Head and Neck, Voice Center, UT Southwestern, Dallas, TX, United States

Rats produce ultrasonic vocalizations (USVs) for conspecific communication. These USVs are valuable biomarkers for studying behavioral and mechanistic changes in a variety of diseases and disorders. Previous work has demonstrated operant conditioning can progressively increase the number of USVs produced by rats over multiple weeks. This operant conditioning paradigm is a useful model for investigating the effects of increased laryngeal muscle use on USV acoustic characteristics and underlying central and peripheral laryngeal sensorimotor mechanisms. Previous USV operant conditioning studies relied on manual training to elicit USV productions, which is both time and labor intensive and can introduce human variability. This manuscript introduces a semi-automated method for training rats to increase their rate of USV production by pairing commercially available operant conditioning equipment with an ultrasonic detection system. USV training requires three basic components: elicitation cue, detection of the behavior, and a reward to reinforce the desired behavior. With the semi-automated training paradigm, indirect exposure to the opposite sex or an olfactory cue can be used to elicit USV production. The elicited USV is then automatically detected by the ultrasonic acoustic system, which consequently triggers the release of a sucrose pellet reward. Our results demonstrate this semi-automated procedure produces a similar increase in USV production as the manual training method. Through automation of USV detection and reward administration, staffing requirements, human error, and subject behavioral variability may be minimized while scalability and reproducibility are increased. This automation may also result in greater experimental flexibility, allowing USV training paradigms to become more customizable for a wider array of applications. This semi-automated USV behavioral training paradigm improves upon manual training techniques by increasing the ease, speed, and quality of data collection.

Keywords: ultrasonic vocalizations, voice, automated training, operant conditioning, rat

OPEN ACCESS

Edited by:
Sara Anna Bonini, University of Brescia, Italy

Reviewed by:
Nicola Simola, University of Cagliari, Italy
Markus Fendt, University Hospital Magdeburg, Germany

****Correspondence:***
Aaron M. Johnson
aaron.johnson@nyulangone.org

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *01 December 2021*
Accepted: *24 January 2022*
Published: *04 March 2022*

Citation:
Johnson AM, Lenell C, Severa E, Rudisch DM, Morrison RA and Shembel AC (2022) Semi-Automated Training of Rat Ultrasonic Vocalizations. Front. Behav. Neurosci. 16:826550. doi: 10.3389/fnbeh.2022.826550

INTRODUCTION

Rats produce ultrasonic vocalizations (USVs) for conspecific communication of affective states (Brudzynski, 2009). USVs are produced within the larynx and require fine motor control of the intrinsic laryngeal muscles (Kelm-Nelson et al., 2018). These USVs are valuable biomarkers for studying behavioral and mechanistic changes in a variety of diseases and disorders, such as Parkinson's disease, autism, and age-related voice disorders (Johnson et al., 2015; Caruso et al., 2020; Krasko et al., 2021).

Previous work has demonstrated operant conditioning can progressively increase the number of USVs produced by rats over multiple weeks (Johnson et al., 2011). This operant conditioning paradigm is a useful model for investigating the effects of increased laryngeal muscle use on USV acoustic characteristics and underlying central and peripheral laryngeal sensorimotor mechanisms (Johnson et al., 2013; Lenell et al., 2019; Krasko et al., 2021; Shembel et al., 2021). However, previous work relied on manual training by hand to elicit USV productions (Johnson et al., 2011). While this method is effective in training USVs, it is time and labor intensive and can introduce human variability. Rewarding by hand can only be done one rat at a time and requires at least one human tester to monitor USVs. One solution to these shortcomings is to automate USV training, which would allow for simultaneous monitoring of multiple animals while keeping temporal and stimuli-response criteria consistent. The goals of this study were to develop an automated operant conditioning procedure for training rats to increase their rate of USVs and to determine whether outcomes of the automated training paradigm were equivalent to manual hand training.

MATERIALS AND EQUIPMENT

The semi-automated system consists of operant conditioning hardware and software for reinforcing USV production and an ultrasonic acoustic monitoring system for detecting USVs. **Table 1** contains a detailed list of all equipment and **Figure 1** presents a schematic of the primary equipment.

Operant Conditioning Hardware and Software

The automated operant conditioning system consists of commercially available hardware and software from the HABITEST Modular Behavioral Test System (Coulbourn Instruments, Holliston, MA, United States). A single setup is comprised of a modular test cage housed within a soundproof acoustic chamber to allow for simultaneous training of multiple animals in the same room. The modular test cage contains a pellet feeder and trough to deliver a nutritional pellet reward (F0021, Bio-Serv, Flemington, NJ, United States). Pellet reward retrieval is monitored using a photocell sensor on the trough. To elicit USVs, rats are either indirectly exposed to a conspecific of the opposite sex or an olfactory stimulus controller delivers an air puff with the scent of an animal of the opposite sex via a tube connected to an outlet located immediately above the trough. All equipment is connected to and controlled by an environmental control board that, in turn, is connected to a Habitest Linc. The Linc is connected to a computer and allows the user to design and carry out experiment protocols through the Graphic State and Graphic StateRT software programs, respectively.

TABLE 1 | Detailed list of equipment and materials.

Name	*Product number*	
Operant conditioning system (Coulbourn instruments)		
HABITEST Linc	H02-08	
Environment control board	H03-04	
Modular test cage - rat	H10-11R-TC	
Non-shock floor for rat test cage	H10-11R-TC-NSF	
Assorted wall panel set – rat	H90-00R-M-KT01	
House light rat (Led)	H11-01-R-LED	
Pellet trough led - rat	H14-01R-LED	
Single photocell sensor	H20-94	
Pellet feeder, 45 mg - Rat	H14-23R	
Olfactory stimulus controller	H15-03	
Olfactory evaporation chamber	H15-20	
Fan module- rat	H29-05R	
Tygon 2375 1/8" Id -1/16" Wt	724120	
5 Volt TTL - 24 Volt converter 110v	H91-24	
Spare to switch input adapter	H91-16-SPARE-IN	
Olfactory stim inj panel - rat	H15-01R	
H10-24 with acoustic liner	H10-24A	
Graphic state notation 4 software	GS4.0	
Name	***Product number***	
Ultrasonic recording and detection system (Avisoft bioacoustics)		
Avisoft RECORDER	10301	
UltraSoundGate 816H	34171	
Condenser ultrasound microphone CM16/CMPA	40011	
Calibrated 40 kHz reference signal generator	60105	
Name	***Product number***	
Air supply for automated scent delivery (Airgas)		
Regulator analytical first stage 0-50 PSI delivery CGA346 3500 PSI Inlet 1/4" Male NPT	Y11244B346-AG	
Flowmeter standard 65 mm 20 HF Flow 60 PSI teflon frame stainless steel float 1/8" female NPT	Y21T654-AG	
Air breathing Gr D Size 125 CGA 346	AI B125	
Adapters	Y99593HCB4F-AL, Y99480094-AG, Y994801220-AG, Y9926115-AG	
Name	***Manufacturer***	***Product Number***
Additional equipment		
Dustless precision pellets, 45 mg, rodent purified diet	Bio-Serv	F0021
24-Channel, 8.5 mA, digital I/O device	National instruments	NI USB-6501
BNC cables and adapters to transmit TTL signal from NI USB-6501 to Coulbourn H91-24	Digi-Key	assorted

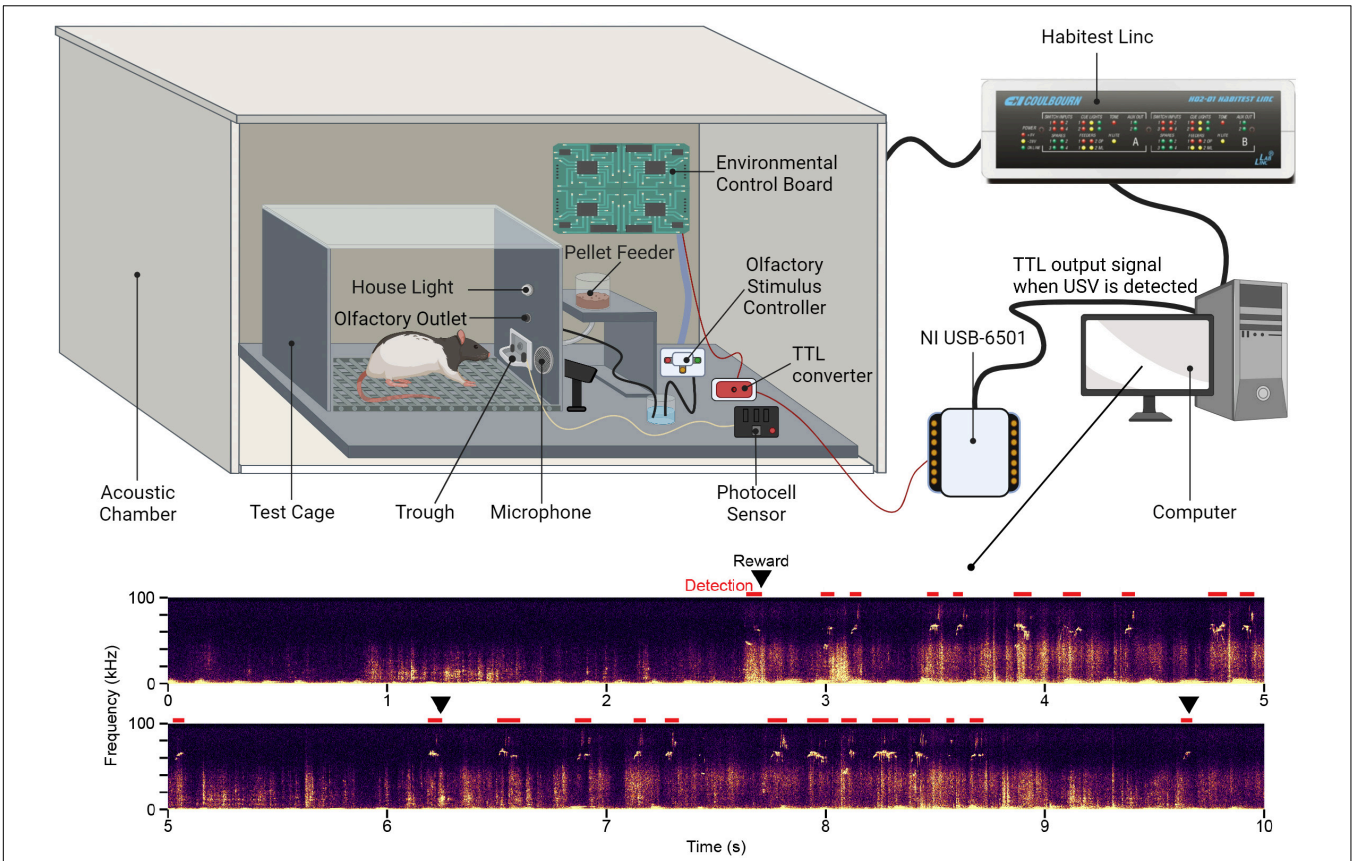


FIGURE 1 | Schematic of the equipment used for semi-automated ultrasonic vocalization training Figure created with BioRender.com.

Ultrasonic Acoustic Monitoring

Test cages are individually monitored using an ultrasonic acoustic system consisting of a condenser ultrasound microphone (CM16/CMPA, Avisoft Bioacoustics, Glienicke, Germany) connected to a USB recording interface (UltraSoundGate 816H, Avisoft Bioacoustics) which powers the microphone and performs the analog-digital conversion. Each cage is independently monitored for USV events using RECORDER software (Avisoft Bioacoustics).

When a USV is detected by the Avisoft program, a TTL signal is sent through a USB digital I/O interface (NI USB-6501, National Instruments, Austin, TX, United States) and routed to the environmental control board within the testing apparatus via a 5V to 24V converter, thereby providing a trigger to dispense the pellet reward. Additionally, the Avisoft system records acoustic files from each training session for later offline acoustic analysis. Commonly used acoustic analysis programs in our laboratory include Avisoft SASLab Pro and DeepSqueak, a MATLAB-based software for USV detection and acoustic analysis (Coffey et al., 2019).

Reward Pellet

Because the pellet reward plays a large role in reinforcing vocalization behavior in our paradigm, it is essential that the appeal of the pellet does not diminish over multiple weeks of training. Sucrose-only pellets were initially used as a reward, as these types of pellets have previously demonstrated effectiveness in eliciting target behaviors (Leenaars et al., 2019). Nutritional sucrose pellets (Bio-Serv Dustless Precision Pellets, 45 mg unflavored) were also considered. Nutritional pellet rewards provide a more complex and nutritionally dense option that may be more appealing to feed-restricted animals and reduces reliance on the need for grain pellet diets between training sessions. Nutritional sucrose pellets were ultimately chosen over the sucrose-only pellets as animals showed preference for the latter. A mix of sucrose and nutritional pellets (e.g., 1:3) may also be used to further motivate rats to seek the reward.

Animals

All experiments were approved by the Institutional Animal Care and Use Committee at the New York University Grossman School of Medicine. Our manual USV training paradigms have been successful in Long Evans (Charles River, Wilmington, MA, United States) and Fischer 344 x Brown Norway rats (National Institute on Aging rodent colony). Therefore, we used Long Evans rats (12 weeks old; male) to test the automated training. All rats were kept on a 12/12 reverse light cycle and trained during their dark cycle when rats are most active. Water was provided *ad libitum*. Access to food was restricted, with grain

pellet feed removed from the home cage prior to the days training was conducted to motivate interest in the nutritional sucrose pellet reward. Rats in the experimental group were restricted to nutritional sucrose pellet rewards received during training. Rats in the control group received a small quantity of nutritional sucrose pellets in their home cage in addition to grain feed pellets at the end of each training day. On non-training days, rats were given unrestricted access to grain feed. Leftover grain feed was removed in the morning before the next training session was conducted. Rats were weighed three times a week and were supplemented with grain feed if their weight dropped below 80% of their baseline.

MATERIALS AND METHODS

Protocol Overview

Ultrasonic vocalizations training requires three basic components: elicitation cue, detection of the behavior, and reward (**Figure 2**). In the manual hand training, the elicitation cue was direct physical exposure to the opposite sex. This cue was quite salient – perhaps overwhelmingly so, as it often distracted the animal from the reward. With the automated training paradigm, indirect exposure to the opposite sex (with the animals separated by a wire mesh) or an olfactory cue can be used to elicit initial vocalizations. When a USV is produced in response to the elicitation cue, it is detected by the ultrasonic acoustic system and a sucrose pellet is delivered through the trough in the HABITEST Modular Behavioral Test System cage. An overview of the training procedures is shown in **Figure 3**.

Pre-training Considerations

Acoustic Environment

One general consideration when preparing the physical space for USV training is to check for extemporaneous ultrasonic noises that could inadvertently trigger USV detection. Some possible sources of ultrasonic noise include computer equipment (e.g., monitors), cell phones, motion detectors, and HVAC vents. Rodent claws on metal gratings, siding, and cage clamps can also trigger USV detections. If multiple rats are present in the training room, individual training apparatuses need to be acoustically isolated to ensure each animal is only rewarded for their individual USV productions. The acoustic isolation chamber described in this experiment provides sufficient acoustic isolation, as would acoustic foam positioned between cages.

Familiarization and Habituation

To control for novelty effects which could influence USV elicitation and training (Kelm-Nelson et al., 2016), several pre-training familiarization and habituation procedures are used.

Handling

Familiarization with training staff can be achieved by handling the animals daily and by transporting them from their home cage to the test cage.

Reward Pellets

To familiarize the animals with the reward pellets, a small dish with the pellets can be placed in the animals' home cage a few days before training begins.

Testing Environment

Placing animals in the test cage several times before formal USV training begins to familiarize animals with the new environment can increase initial USV rate.

Sexual Encounters

Introducing male rats to females in estrus increases the chance of the males expressing interest in the opposite sex pairing, which can also improve olfactory USV elicitations. In our previous studies and pilot testing, we found that young male rats (under the age of 12 months) typically mount with one or two brief interactions with a receptive female, whereas older male rats (over the age of 12 months) may require several consecutive days of exposure to estrus females before an attempt at copulation occurs. Additionally, we have found up to 20-50% of male rats do not express interest in female rats and do not vocalize with the female rat exposure. These rats are excluded from training procedures.

Magazine Training

Magazine training can be utilized to familiarize experimental animals with a novel environment (test cage) prior to the start of formal USV training. In magazine training, the rat is placed in the operant conditioning chamber and presented a reward at random intervals (average 1-3 pellet per minute). Magazine training typically takes place for 10-30 min per day for 2-5 days, depending on the response from each rat.

Transport

If animals need to be transported from their colony to the training site, allow the animals to remain in their home cage for 10-15 min to habituate to the environment prior to each training session.

Behavior Shaping

The goal of behavior shaping is to reward approximate target behaviors. During shaping, each rat is rewarded once for every bout of USVs. Because USV production is influenced by strain, age, and sex of the rat, each bout may consist of a single USV or a string of several USVs in rapid succession.

Training Schedule

Session Timing

Training sessions always occur during the active (dark portion) of the light cycle. In our experience, the specific time of day the training session is conducted within the active portion of the cycle may influence USV production. To minimize any potential influence of time of day, we conducted training during the first 4 h of the dark portion of the cycle.

Session Duration

In our previous manual studies, we have demonstrated rodents vocalize throughout sessions with a duration of 10 min but

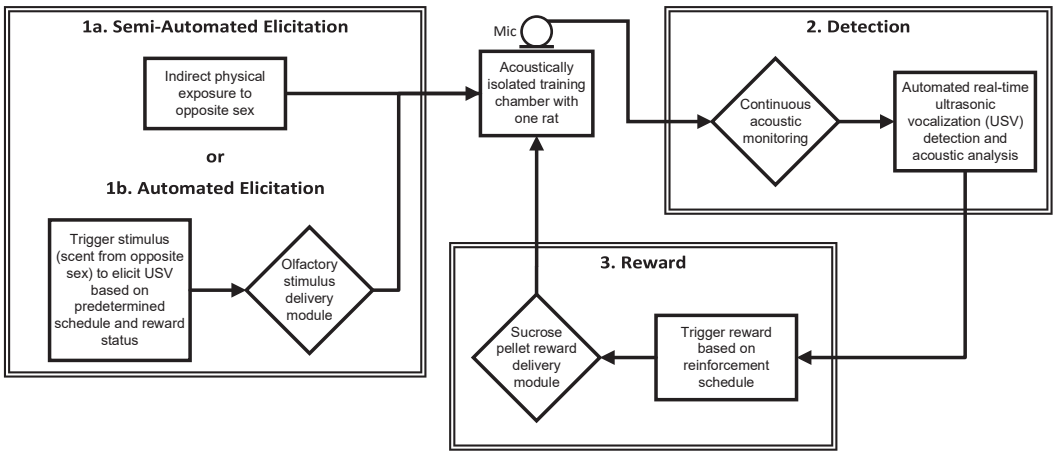


FIGURE 2 | Diagram of the semi-automated ultrasonic vocalization (USV) training procedure indicating the equipment (diamonds) and processes (rectangles) involved in the elicitation, detection, and reward of USVs.

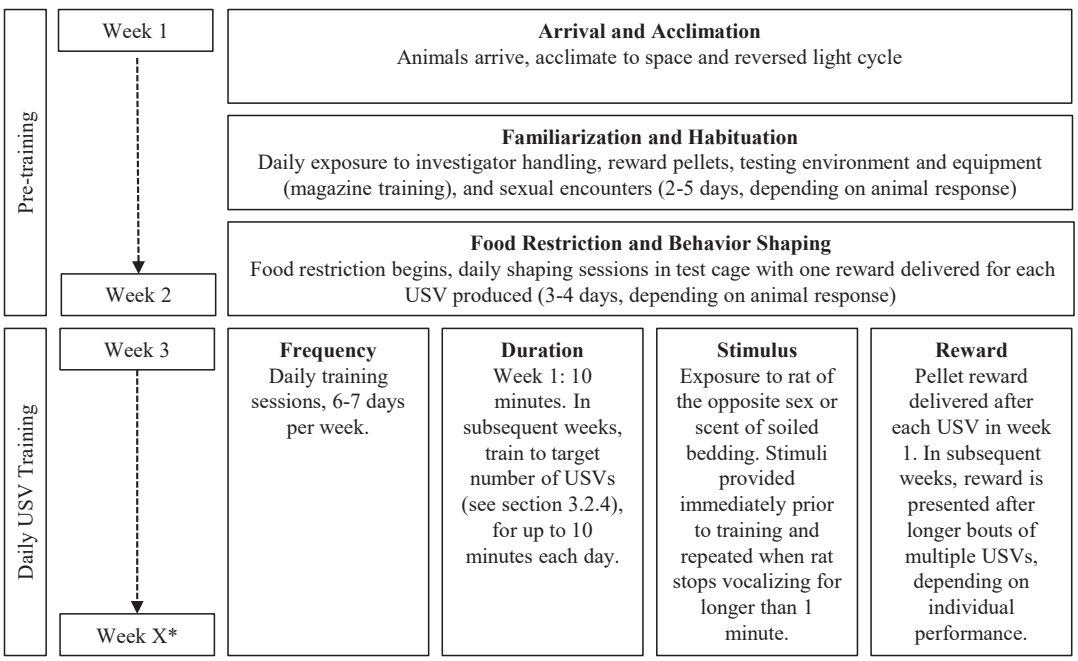


FIGURE 3 | Overview of pre-training and behavioral training timeline. *Total number of weeks of training is determined by the experimental goal.

do not perform consistently in longer sessions. For example, following initial introduction to male rats, most female rats produce 80% of their vocalizations within the first 6 min of a 10-min recording session (**Figure 4**; Lenell and Johnson, 2021). If longer duration of training is preferred with the semi-automated training, water *ad libitum* is recommended, as rats will get thirsty and lose motivation.

Number of Sessions per Week

Since weekend breaks from USV training can adversely affect consistency and rate of training responses, 6-day a week training schedules are recommended.

Daily Training Target

Because of variability in baseline USV production rates amongst rats, it is advisable to individualize weekly USV training targets based on each individual rat's baseline performance. For example, in a prior USV manual training study using Fisher 344 × Brown Norway male rats, the young (9-month-old) cohort produced a range of 33-150 USVs per session during their first week of 10-min USV training. In this example, a predetermined goal of 100 USVs per training session would be unattainable for one rat to produce while being an easy target for another rat.

Our training protocol has focused on the progression of USV production rate over a fixed amount of training duration. In

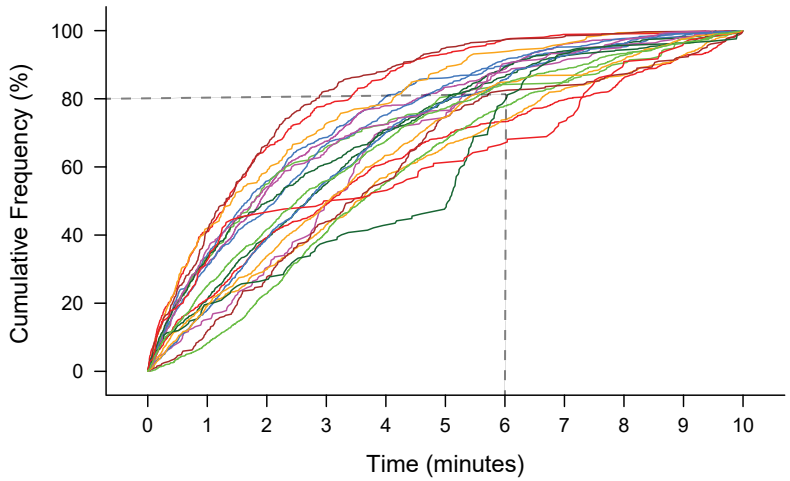


FIGURE 4 | The ultrasonic vocalization (USV) productions during 10 min of recording of 20 female Long-Evans rats elicited by an introduction to a male rat. Each line is the cumulative frequency of USVs for an individual rat. The majority (16/20) rats produced 80% of their total number of USVs within the first 6 min of recording. The total number of USVs produced ranged from 100 – 700.

manual training, production of USVs rapidly increased during the first four weeks of USV training and then steadily increased thereafter. USV training targets with automated training can be increased by 50% each week for the first 4 weeks and then by 20% in additional weeks of training. If rats are not able to achieve their individual goal for the week, the goal remains unchanged for the following week until the target percentage is achieved across at least 2 consecutive sessions.

Elicitation

USVs can be elicited using either manual presentation of a rat of the opposite sex or using soiled bedding as an olfactory stimulus. In both methods, prior to testing, animals should be familiarized with the scent and appearance of opposite sex rats to combat sexual naivety and increase valuation of the stimulus animals or olfactory stimulus. Once vocal training begins, the elicitation occurs within the first couple of weeks and is then phased out once the animals make the association between vocalizations and the sucrose reward. In manual USV training, USVs can be elicited by briefly exposing sexually experienced rats to the opposite sex. While this method robustly elicits USVs, it requires manual presentation and removal of stimulus animals. To bypass this, other labs have found olfactory exposure to the opposite sex can elicit USV production in sexually experienced male rats (Ciucci et al., 2008). Preliminary results from our lab also confirms that olfactory cues can elicit USVs, although the response is less robust than physical presentation of a stimulus animal.

For the olfactory stimulus, one week's worth of soiled bedding can be removed from female home cages and soaked in water. The water can then be used as an olfactory cue by placing the liquid through an olfactory module system (Coulbourn Instruments, Harvard Bioscience, Boston, MA, United States). The module propels air through the system using a compressed air tank with a low pressure (1-5 PSI) and low flow rate of (5 SCFH or less). This configuration delivers a one-second puff of air with the olfactory stimulus into the test chamber during each training session. To maintain novelty of the olfactory stimuli, soiled bedding should be presented from different animals. To prevent competing olfactory stimuli, bedding should be removed from the tray underneath the testing chamber after each session. One caveat of this procedure is that the brief olfactory cue may elicit USVs for some rats, however, low vocalizing rats may need physical presentation of the stimulus animal to elicit USVs.

Ultrasonic Vocalizations Detection Parameters

The Avisoft RECORDER software configuration allows for automatic USV detection based on several different USV acoustic parameters: signal amplitude, duration, frequency range, and amount of entropy. Additionally, a USV bout can be counted as a single event by adjusting the hold time parameter, which holds the current trigger as active for a defined duration between each detected USVs. The specific values for the detection parameters used in our preliminary studies are shown in **Figure 5**. These values may need to be adjusted to account for USV production variability across animals and extemporaneous environmental noise. The configuration can be tweaked and tested using the USV real-time monitoring tool and running the detection on pre-recorded sound files with the "run from.wav file" option. With multi-cage setups, each microphone should be calibrated to the same intensity level using an ultrasonic signal generator (Avisoft Bioacoustics) to maintain intensity consistency between recording hardware.

Reward

For initial learning effects, each vocalization should be paired with one sucrose pellet. After the animal is consistently making the association between USV production and reward, sucrose reward presentations can be presented after several bouts of vocalizations (variable reward).

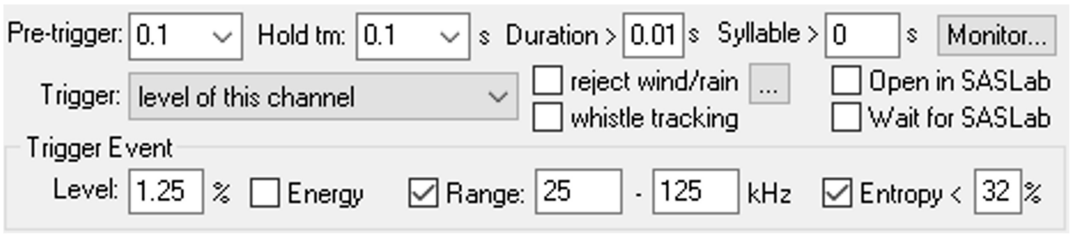


FIGURE 5 | Screenshot of the configuration settings in Avisoft RECORDER for automatic detection of ultrasonic vocalizations.

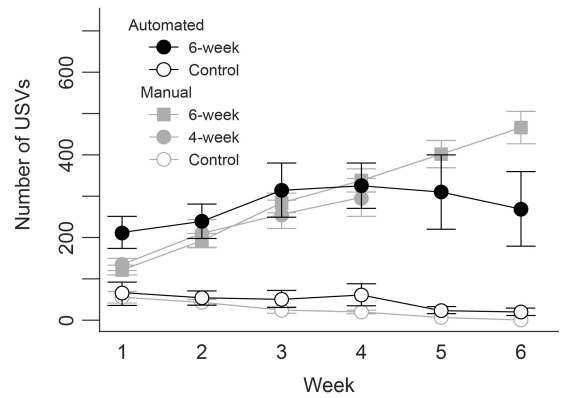


FIGURE 6 | Representative results of the weekly increase in ultrasonic vocalization production during semi-automated training compared to previous manual training.

Visual Cue for Reward

The preinstalled house light in the operant testing chamber can be illuminated as a means of providing a visual cue associated with the pellet reward and the start of the vocalization training session; however, we found that vocalizations occurred immediately upon placement of the animal into the testing chamber regardless of illumination status. Therefore, the house light can remain illuminated throughout the training to facilitate visual tracking of the animal's movement without concern it will influence USV production.

RESULTS

Figure 6 compares results from our previous work using manual USV training to animals trained using the semi-automated procedure. In the manual training, USV production steadily increased across 4 and 8 weeks of manual training in 16 male Long Evans rats. The semi-automated procedure resulted in a similar increase in the number of USVs produced over the first 4 weeks of training. However, increases in performance in rats trained using the semi-automated procedure began to flatten during the fifth week and then decreased in week 6. The untrained control groups in both the manual and semi-automated training paradigms produced a minimal number of USVs throughout the training periods.

DISCUSSION

The move to automate the USV training process brings several key advantages over commonly used manual training techniques. Automated training reduces staffing requirements while increasing the number of animals that can be trained simultaneously. Automated training also allows for multiple and longer training sessions throughout a day, providing an opportunity to increase the total number of vocalizations elicited compared to manual training. Furthermore, automating the reward process provides the animal with a more consistent behavioral experience and less human exposure, reducing inherent human variability with manual presentation of rewards. Together, these advantages position automated USV training to provide a higher level of scalability and consistency than traditional manual training, allowing for increased experimental bandwidth and ultimately faster, more reproducible data collection.

Automated training also allows for greater experimental flexibility. Here, we outline the use of olfactory and visual cues to enhance behavioral training. Other experimental designs could conceivably benefit from precisely timed administration of cues in different sensory modalities. For example, reward conditions during training could be tied to specific sound cues, or operant conditioning paradigms could automatically apply opposing cues to reinforce target behaviors and dissuade others. Moreover, the automatic detection of USVs employed here could allow for training of specific vocalization characteristics. By systematically changing USV detection criteria over the course of behavioral shaping, animals could be trained to perform specific types of vocalizations. For example, during the behavioral shaping period detection criteria for duration of USVs could be steadily lengthened as performance increases, preferentially rewarding and reinforcing longer and longer vocalizations until a target duration is met. Similar shaping paradigms could potentially reinforce vocalizations of a specific frequency, intensity, or type (e.g., frequency modulated versus flat tone) by slowly narrowing USV detection criteria for frequency, amplitude, bandwidth, and sinuosity. Our preliminary results with this new semi-automated training yielded similar progressive increases in the number of overall USVs produced as previously found in manual training. However, it is important to note that while performance continued to increase over 8 weeks in the manual training, there was a dip in performance after 6 weeks with the semi-automated training. These differences need to be considered

when developing and conducting experimental designs and may indicate an inherent limitation of long-term automated training, or that a change in behavioral training paradigm around week 6 is needed to counteract this downturn in performance.

Through automation of USV detection and reward administration, staffing requirements, human error, and subject behavioral variability may be minimized while scalability and reproducibility is increased. This automation may also result in greater experimental flexibility, allowing USV training paradigms to become more customizable for a wider array of applications. Here, we outline a semi-automated USV behavioral training paradigm that improves upon manual training techniques by increasing the ease, speed, and quality of data collection.

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 the Institutional Animal Care and Use Committee at the New York University Grossman School of Medicine.

AUTHOR CONTRIBUTIONS

AJ, AS, and CL contributed to conception and design of the study. AS, CL, DR, ES, and RM conducted the training trials. All authors wrote the sections of the manuscript and contributed to the manuscript revision, read, and approved the submitted version.

FUNDING

Support for this work was provided by grants from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health, R21 DC018107 (AJ), T32 DC009401 (CL).

ACKNOWLEDGMENTS

We thank Adam Mar, Director of the Rodent Behavior Core at NYU Langone Health, for his advice during the early stages of protocol development, and the service and support technicians at Coulbourn Instruments for their help troubleshooting the equipment setup.

REFERENCES

Brudzynski, S. M. (2009). Communication of adult rats by ultrasonic vocalization: biological, sociobiological, and neuroscience approaches. *ILAR J.* 50, 43–50. doi: 10.1093/ilar.50.1.43

Caruso, A., Ricceri, L., and Scattoni, M. L. (2020). Ultrasonic vocalizations as a fundamental tool for early and adult behavioral phenotyping of Autism spectrum disorder rodent models. *Neurosci. Biobehav. Rev.* 116, 31–43. doi: 10.1016/j.neubiorev.2020.06.011

Ciucci, M. R., Ma, S. T., Kane, J. R., Ahrens, A. M., and Schallert, T. (2008). Limb use and complex ultrasonic vocalization in a rat model of Parkinson's disease: deficit-targeted training. *Park. Relat. Disord.* 14(Suppl. 2), S172–S175. doi: 10.1016/j.parkreldis.2008.04.027

Coffey, K. R., Marx, R. G., and Neumaier, J. F. (2019). DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 859–868. doi: 10.1038/s41386-018-0303-6

Johnson, A. M., Ciucci, M. R., and Connor, N. P. (2013). Vocal training mitigates age-related changes within the vocal mechanism in old rats. *J. Gerontol. A Biol. Sci. Med. Sci.* 68, 1458–1468. doi: 10.1093/gerona/glt044

Johnson, A. M., Doll, E. J., Grant, L. M., Ringel, L., Shier, J. N., and Ciucci, M. R. (2011). Targeted training of ultrasonic vocalizations in aged and Parkinsonian rats. *J. Vis. Exp.* 54:e2835. doi: 10.3791/2835

Johnson, A. M., Grant, L. M., Schallert, T., and Ciucci, M. R. (2015). Changes in rat 50-kHz ultrasonic vocalizations during dopamine denervation and aging: relevance to neurodegeneration. *Curr. Neuropharmacol.* 13, 211–219. doi: 10.2174/1570159x1302150525122416

Kelm-Nelson, C. A., Brauer, A. F., and Ciucci, M. R. (2016). Vocal training, levodopa, and environment effects on ultrasonic vocalizations in a rat neurotoxin model of Parkinson disease. *Behav. Brain Res.* 307, 54–64. doi: 10.1016/j.bbr.2016.03.006

Kelm-Nelson, C. A., Lenell, C., Johnson, A. M., and Ciucci, M. R. (2018). "Laryngeal activity for production of ultrasonic vocalizations in rats," in *Handbook of Ultrasonic Vocalization: a Window into the Emotional Brain*, ed. S. M. Brudzynski (San Diego, CA: Elsevier Academic Press Inc), 37–43.

Krasko, M. N., Hoffmeister, J. D., Schaen-Heacock, N. E., Welsch, J. M., Kelm-Nelson, C. A., and Ciucci, M. R. (2021). Rat models of vocal deficits in Parkinson's disease. *Brain Sci.* 11:925. doi: 10.3390/brainsci11070925

Leenaars, C. H. C., Pels, E. G. M., Joosten, R. N. J. M. A., and Ritskes-Hoitinga, M. (2019). Wistar rats do not show preference for either of two commonly used nutritionally sound food rewards in a T-maze. *J. Vet. Behav.* 31, 22–27. doi: 10.1016/j.jveb.2019.01.007

Lenell, C., and Johnson, A. M. (2021). The effects of the estrous cycle, menopause, and recording condition on female rat ultrasonic vocalizations. *Physiol. Behav.* 229:113248. doi: 10.1016/j.physbeh.2020.113248

Lenell, C., Newkirk, B., and Johnson, A. M. (2019). Laryngeal neuromuscular response to short- and long-term vocalization training in young male rats. *J. Speech Lang. Hear. Res.* 62, 247–256. doi: 10.1044/2018_JSLHR-S-18-0316

Shembel, A. C., Lenell, C., Chen, S., and Johnson, A. M. (2021). Effects of vocal training on thyroarytenoid muscle neuromuscular junctions and myofibers in young and older rats. *J. Gerontol. A Biol. Sci. Med. Sci.* 76, 244–252. doi: 10.1093/gerona/glaa173

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Behavioral Neuroscience

REVIEW
published: 28 March 2022
doi: 10.3389/fnbeh.2022.833168

From Mating to Milk Access: A Review of Reproductive Vocal Communication in Mice

Sara Capas-Peneda[1,2,3], Yolanda Saavedra Torres[1], Jan-Bas Prins[1,4] and I. Anna S. Olsson[2,3]*

[1] Biological Research Facility, Francis Crick Institute, London, United Kingdom, [2] i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [3] ICBAS – School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal, [4] Leiden University Medical Centre, Leiden, Netherlands

Vocalisations play a central role in rodent communication, especially in reproduction related behaviours. In adult mice (*Mus musculus*) the emission of ultrasonic vocalisations (USVs) has been observed in courtship and mating behaviour, especially by males. These have been found to have distinctive individual signatures that influence female choice of mating partner. The most recent findings show that vocal communication also has a role in parental cooperation, in that female mice communicate with male partners in ultrasonic frequencies to induce paternal behaviour. Infant vocalisations form the other important part of reproductive vocal communication. Although born deaf, neonatal mice are capable of producing vocalisations since birth. As an altricial species, successful mother-infant communication is essential for survival, and these vocalisations are important modulators of maternal behaviour. Three main types of infant vocalisations have been identified and characterised. Most research has addressed pure USVs, related to stressful situations (e.g., cold, isolation, handling, presence of unfamiliar males or predators), which usually elicit maternal search and retrieval. In addition, broad-band spectrum signals, emitted post-partum during cleaning of foetal membranes, inhibit biting and injury by adults and "wriggling calls," emitted during suckling, release maternal behaviour (such as licking). Several variables have been identified to modulate vocalisations in mice, including individual characteristics such as strain/genotype, age, sex, and experimental factors such as pharmacological compounds and social context. In recent years, there has been a big increase in the knowledge about the characteristics of vocal communication in rodents due to recent technological advances as well as a growing interest from the neuroscience community. Vocalisation analysis has become an essential tool for phenotyping and evaluating emotional states. In this review, we will (i) provide a comprehensive summary of the current knowledge on mouse reproductive vocal communication and (ii) discuss the most recent findings in order to provide a broad overview on this topic.

Keywords: mouse, vocal communication, ultrasonic vocalisation (USV), courtship, neonatal vocalisation

OPEN ACCESS

Edited by:
Susanna Pietropaolo, UMR 5287 Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (INCIA), France

Reviewed by:
Valerie Lemaire, Université de Bordeaux, France
Luigi Bellocchio, INSERM U1215 Neurocentre Magendie, France

****Correspondence:***
Sara Capas-Peneda
sara.capas@crick.ac.uk

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *10 December 2021*
Accepted: *07 March 2022*
Published: *28 March 2022*

Citation:
Capas-Peneda S, Saavedra Torres Y, Prins J-B and Olsson IAS (2022) From Mating to Milk Access: A Review of Reproductive Vocal Communication in Mice. Front. Behav. Neurosci. 16:833168. doi: 10.3389/fnbeh.2022.833168

INTRODUCTION

Mice live in a complex social environment. Communication between conspecifics occurs by different types of cues that provide distinct information. Recent evidence suggest that body language, such as body posture and facial expressions play a part in social communication (Ebbesen and Froemke, 2021). But, the most well-studied forms of communication are still the odour cues, mainly pheromones, and vocal communication that play a role in informing on the location of animals, presence of food or threat, sexual attraction, courtship and dam-pup interactions (Bind et al., 2013; Portfors and Perkel, 2014).

Vocal communication has been gathering much interest since the first description of ultrasonic vocalisations (USVs) in neonatal mice by Zippelius and Schleidt (1956). USVs (in the range of 30–90 kHz)[1] are now known to serve important functions in sexual behaviour and pup and dam communication in mice.

Furthermore, although significantly less studied, mice are also able to produce vocalisations in the audible (to the human ear) range that serve different functions which are generally related to negative affective states, such as in response to a predatory attack (Blanchard et al., 1998), human handling (Whitney, 1969), or by females in response to male sexual behaviour when in a non-receptive state (Sugimoto et al., 2011; Neunuebel et al., 2015).

Communication in the ultrasonic range is thought to pose an evolutionary advantage due to the potential predatory evasion since the most common predators of house mice have hearing ranges below the frequencies of USVs (Musolf and Penn, 2012). It seems that USVs appeared first in neonatal mice, and it is theorised that they were later exploited by males. This is due to the fact that neonatal USVs reduce female aggression and most matings occur in the post-partum oestrous, when the female is especially receptive to neonatal USVs (Whitney et al., 1973).

Ultrasonic vocalisations can be grouped according to internal frequency changes, duration and sonographic shape in ten different categories, as proposed by Scattoni et al. (2008) (see **Table 1**).

Murine vocalisations are produced by airflow through the larynx. The mouse larynx is a tube-shaped musculocartilaginous organ through which the air passes from the pharynx to the trachea and, along with other functions, is the organ of phonation. As in other mammals, it is composed by three unpaired cartilages (epiglottic, thyroid, and cricoid cartilages) and two arythenoid cartilages. The vocal folds are located in the vestibule of the larynx, where there is also a large laryngeal recess. It is when air passes during expiration generating vibration of the vocal folds that sound is emitted (Navarro et al., 2017a).

Different mechanisms are involved when USV versus audible sounds are produced (Roberts, 1975). Audible sounds are produced by the vibration of vocal cords in the larynx, in a similar process to other mammals. In contrast, ultrasonic vocalisations are produced by a whistle like mechanism that is currently not completely understood. Currently, two main theories have been proposed, the planar impinging model by Mahrt et al. (2016) that suggests USVs are produced by an intralaryngeal air jet created by a glottal constriction impinging on the interior surface of the thyroid cartilage, and the edge-tone model proposed by Riede et al. (2017) that suggest USVs are produced by an interaction between the glottal exit and the edge of an intralaryngeal ventral pouch. Although they differ in exact mechanisms, both these theories indicate that glottal constriction and laryngeal muscle contraction are necessary to produce USVs.

Since USVs play an important role in mouse vocal communication, the hearing range of mice needs to be compatible with the frequencies at which these animals vocalise. Hearing range is dependent on the physics of the structures involved in the middle and inner ear and, thus, on the anatomy of the ear. The soundwaves are received by the auricle, in the external ear, and are carried in the external acoustic meatus toward the tympanic membrane that marks the beginning of the middle ear. The vibration is transmitted through the three main auditory ossicles, malleus, incus, and stapes, where the sound waves are amplified. Ultimately, the sound waves will reach the cochlea where they will be converted into mechanical waves *via* the displacement of the perilymph and endolymph. The last step of this process is triggered by the hair cells in the cochlea that translates this into a nerve impulse, ultimately reaching the central nervous system. In the mouse, the high flexibility of the ossicular joints in the middle ear is responsible for the ability to hear higher frequencies, such as the ones that are characteristic of ultrasonic vocalisations (Navarro et al., 2017b).

TABLE 1 | Different types of ultrasonic vocalisations (Scattoni et al., 2008).

Denomination	Characteristics
Complex calls	One syllable containing two or more directional changes in pitch, each ≥6.25 kHz
Harmonics	One main call, resembling the complex calls, but with additional calls of different frequencies surrounding the main call
Two-syllable calls	Consist of two components: a main call (flat or downward) with an additional punctuated component toward the end
Upward-modulated calls	Exhibit a continuous increase in pitch that was ≥12.5 kHz, with a terminal dominant frequency at least 6.25 kHz more than the pitch at the beginning of the vocalisation.
Downward-modulated calls	Exhibit a continuous decrease in pitch that was ≥12.5 kHz, with a terminal dominant frequency at least 6.25 kHz less than the pitch at the beginning of the vocalisation
Flat calls	Display a constant beginning and the ending of the pitch frequency remained constant -≤3 kHz of each other.
Chevron calls	Resemble an "inverted-U," which was identified by a continuous increase in pitch ≥ 12.5 kHz followed by a decrease that was ≤ 6.25 kHz
Short calls	punctuated and shorter than 5 ms
Composite calls	Formed by two harmonically independent components, emitted simultaneous
Frequency steps	Instantaneous frequency changes appearing as a vertically discontinuous "step" on a spectrogram, but with no interruption in time.

[1]We will adhere to the established terminology and refer to vocalisations beyond the human hearing range as ultrasonic; however this anthropocentric terminology is questionable when referring to vocalisations in species of non-human animals with a different hearing range.

Vocal communication plays a crucial role in parental care and especially maternal behaviour in laboratory and wild mice. Mice are altricial species and pups are born hairless, with very limited motor capacity and without thermoregulatory abilities, which makes them completely dependent on maternal care for survival (Weber and Olsson, 2008). USVs are an important component of pup and dam communication, playing a role in maternal bonding, along with odour and tactile stimuli (Nagasawa et al., 2012). Although pups are born deaf and only develop hearing abilities by day 4 or 5 (Weber and Olsson, 2008), they are able to emit vocalisations in a wide range of frequencies that females are able to recognise and respond to in a variety of different maternal behaviours such as, nest building and retrieval (Ehret and Haack, 1981). These include ultrasonic vocalisations, usually associated with isolation (between 30 and 90 kHz), broadband spectrum signals inhibit biting and injury by adults (4–40 kHz) (Haack et al., 1983) and low-frequency calls with a major energy below 10 kHz and a frequency range rarely exceeding 20 kHz. The low-frequency calls are often called wriggling calls and release maternal behaviour, such as licking of the pups (Ehret and Bernecker, 1986).

Females are able to discriminate between sounds produced from pups and sounds emitted by other sources, and show a preference for pup isolation calls when given the choice between these and male USVs or artificial tone bursts with the same frequency properties (Hammerschmidt et al., 2009). It is also known that infant USVs promote cerebral cortex plasticity in mothers, by promoting the activation of certain neuron populations in the auditory cortex in response to ultrasonic and low frequency pup calls (Tasaka et al., 2020).

In this review, we will focus on vocal communication in breeding contexts, covering sexual, parental behaviours and neonatal vocalisations in wild-type and wild derived mice.

MATERIALS AND METHODS

All the literature reviewed and included in this review is organised in the table provided as **Supplementary Material**. Information was extracted and organised into four sections—courtship, copulation, neonatal vocalisations, parental cooperation—and each section is organised into six categories:

- Strain, age, number of animals used
- Detection method
- Testing condition
- Variables measured
- Major findings

The literature included in this review was obtained from PubMed, Web of Science, and Google Scholar using the keywords "mouse," "vocalisation," "vocalization," "ultrasonic vocalisation," "ultrasonic vocalization," "courtship," "neonatal," and "infant." The data were collected in a period between April 2020 and November 2021.

Only original research in wild-type or wild-derived mice was selected with no restriction on the date of publication. Exceptions were made for original research using genetically modified lines or artificially induced models that were studied to elucidate crucial mechanisms of vocal communication in mice. Only research published in English or with at least an abstract in English was included.

COURTSHIP

Historically, vocalisations during heterosexual encounters of mice have been attributed to the male (Sales, 1972; Whitney et al., 1973; Warburton et al., 1989; Barthelemy et al., 2004). Several reasons led to this conclusions, including limitations related to the recording equipment that was used to detect vocalisations or to the way some experiments were conducted which prevented the emission of vocalisations by the female either by devocalisation, by anaesthesia (Whitney et al., 1973) or by replacing the female by olfactory stimulation (Whitney et al., 1974). Nevertheless, Sales (1972) had already noted that ultrasound emission was not exclusive of male mice and that "audible" cries were also detected coming from the female. In more recent research it has been established that females ultrasonically interact with males during courtship displays (Neunuebel et al., 2015) and also produce broadband sounds (Finton et al., 2017).

Characteristics of Male Courtship Vocalisations

Male mice emit ultrasonic vocalisations with frequencies ranging over 30–110 kHz that meet the criteria of song in that they are composed of different syllable types organised in a non-random temporal sequence (Holy and Guo, 2005). Furthermore, they have an individual signature that can distinguish between individual laboratory mice (Holy and Guo, 2005; Ronald et al., 2020; Melotti et al., 2021) and house mice (Marconi et al., 2020). Male USVs are triggered by the presence of a female (Sales, 1972) and several other related odour stimuli such as female-soiled cage shavings (Whitney et al., 1974), female urine (Nyby et al., 1977), female saliva (Byatt and Nyby, 1986), and female vaginal fluids (Nyby et al., 1977). Nevertheless, the specific chemo signal responsible for this effect has not yet been identified and freezing urine has a deleterious effect on its ability to elicit vocalisations, leading to a decrease in the number of USVs emitted when compared to fresh urine (Hoffmann et al., 2009).

When presented with a novel female, males produced short upward and one-jump syllables soon after the introduction of the female and the number of long syllables with frequency jumps increased approximately 1 min after the introduction of females (Matsumoto and Okanoya, 2016).

Effects of Age, Social Experience, Relatedness, Genetics

Increasing age was found to negatively affect the number of courtship USVs, with 30 weeks old male mice emitting fewer USVs when compared with younger males (Kanno and Kikusui, 2018). However, when taking into account social experience, the number of USVs is directly related to the past sociosexual experience of the male, and the effect of age can be mitigated by the latter (Kanno and Kikusui, 2018).

The genetic relatedness of the mating partner also seems to modulate male USVs, for instance, when presented with an unrelated female partner, male mice emit more, longer and a higher number of complex USVs when compared to the same interaction with a related female (Nicolakis et al., 2020).

F_1 hybrids of certain strains have also been demonstrated to have higher calling rate than parent strains, for instance, both F_1 progeny of C57BL/10Bg.D1-Y and DBA/1Bg have higher calling rates than the parental strains but this is not verified in all strains (Maggio and Whitney, 1986). It has been suggested that there is a directional dominance mode of inheritance of high rate of calling (Hahn et al., 1987).

Cross-fostering experiments with BALB/c and C57BL/6 suggest that adult courtship calls are innate, the calls from fostered males kept the characteristics of the parental strain (Kikusui et al., 2011).

Effects of Female Oestrous State and Hormonal State

Although males will mount females independently of their oestrous state, they seem to modulate syllable parameters such as dominant frequency, duration and bandwidth according to it. The lowest dominant frequency and highest duration and bandwidth were detected when males were exposed to females in proestrus and the opposite with dioestrus females; intermediate parameters were detected with oestrus females (Hanson and Hurley, 2012). Androgens also influence male USVs: in castrated males, the latency to produce USVs is higher and this is surpassed when males are supplemented with testosterone (Dizinno and Whitney, 1977).

Effects of Social Context

Social status seems to affect the amount of USVs produced by male mice when interacting with a female: dominant males emit more 70-kHz vocalisations when compared to subordinates, in a courtship context (Nyby et al., 1976). Furthermore, an "audience effect" was described by Seagraves et al. (2016), in that male mice modified their call rate, acoustic structure and syllable complexity in response to the presence of male body odour and an anaesthetised male audience. Interestingly, this effect was only observed in the simultaneous presence of male odour and an anaesthetised male and not when the experimental subject was exposed to only male odour or only male USVs.

Characteristics of Female Vocalisations During Courtship Displays

Females produce two types of vocalisations during interaction with males, USVs, and audible squeaks or broadband vocalisations (BBVs) (Lupanova and Egorova, 2015; Neunuebel et al., 2015; Ronald et al., 2020). They appear to have two distinctive functions: USVs are mainly produced during pursuit by males and when in close proximity with males (Neunuebel et al., 2015) and BBVs are often produced in accompanying behaviours such as kicking or lunging at males, often signalling rejection (Sugimoto et al., 2011). Ronald et al. (2020) investigated this further by exposing male mice to female urine with USVs or BBVs playback and found that male mice adjust their vocal courtship according to the type of vocalisation emitted by females: male USVs production was highest when exposed to female urine and USVs and lowest when they were exposed only to female urine, BBVs or the combination of BBVs and female urine, with no differences between all these experimental conditions. BBVs were used in this case because they are associated with rejection from the female. This also shows that olfactory stimuli provide context for female vocalisations.

Female Preferences and Reproductive Success

When exposed to male ultrasonic vocalisations, females seem to be attracted to them independently of their oestrous status as shown by Hammerschmidt et al. (2009). Furthermore, females prefer to interact with vocally intact males or surgically devocalised males with a playback of artificial 70 kHz tones when compared to devocalised males (Pomerantz et al., 1983). When given the choice between a compartment with a playback of male USVs against a silent compartment, females spent more time in the first. Such a preference was not found when the USVs were obtained from mouse pups (isolation calls) or artificial tone bursts of constant frequency between 70 and 80 kHz (Hammerschmidt et al., 2009).

Females also seem to have a preference for complex versus simple male vocalisations, as shown by Chabout et al. (2015) in a Y-maze choice test paradigm, this indicates that the female response to male courtship USVs is multidimensional, depending on the characteristics of male USVs.

Furthermore, female mice prefer male USV playback from unfamiliar non-kin compared to that from familiar siblings which indicates that females are able to discriminate between familiar siblings and unfamiliar non-siblings by their USVs. This occurs in both laboratory mice (Asaba et al., 2014) and wild-derived mice (Musolf et al., 2010). One interesting point to note is that this preference is not genetically determined. This was shown in cross-fostering experiments, where C57BL/6 females raised by BALB/c foster mice showed a preference for male USVs produced by C57BL/6 males. In contrast, C57BL/6 females that were raised by C57BL/6 parents showed a preference for male BALB/c USVs. Furthermore, song preference disappeared when females were raised in fatherless conditions. This demonstrates that the environment during development plays an important role in the subsequent reproductive life of females (Asaba et al., 2014).

Less is known about the relationship between male USVs and subsequent reproductive success, nevertheless, Nicolakis et al. (2020) described a negative correlation between the mean number and length of vocalisations with the latency of the pairs first litter.

COPULATION

In seminal work, Sales (1972) first described the occurrence of ultrasonic calls during mating behaviour in laboratory mice in three different strains, specifically during investigative behaviour of the female by the male (sniffing) and during mounting and

intromission. This work detected 40 kHz vocalisations during mating (Sales, 1972); however subsequent investigations focused on 70 kHz vocalisations. It was not until over two decades later that White et al. (1998) first documented the occurrence of both 40 and 70 kHz vocalisations during copulation. These two different vocalisations occur at different times, 70 kHz are emitted throughout the precopulatory and copulatory periods and are interspersed with 40 kHz calls that are seen most often in the later phases of a copulatory sequence (Sales, 1972; White et al., 1998). In addition, spectrographic analysis of USVs emitted during mounting revealed that males produced longer and more complex syllables with harmonics during mounting behaviour (Matsumoto and Okanoya, 2016).

NEONATAL VOCALISATIONS

Neonatal mice are capable of producing vocalisations in a wide range of frequencies. Three different call types have been described in infant mice: distress calls in ultrasonic range, broadband or "pain calls" and low-frequency or "wriggling calls" (Zippelius and Schleidt, 1956; Ehret and Bernecker, 1986). Each of these are involved in mother-offspring communication and recognition and play different roles in eliciting maternal behaviour (Ehret and Haack, 1981; Ehret and Bernecker, 1986).

Due to the altricial nature of mouse pups (Weber and Olsson, 2008), mother-infant communication plays a crucial role in survival, and neonatal vocalisations are central in infant-mother communication, as vocalisation is virtually the only mean to convey information from the pup to the dam.

Ultrasonic Calls of Infant Mice

Isolation ultrasonic calls in neonatal mice were first described as "Pfeifen des Verlassenseins" ("whistles of abandonment") by Zippelius and Schleidt (1956). USVs emitted by neonatal mice are whistle-like sounds characterised by frequencies ranging between 30 and 90 kHz, duration of 10–200 ms, and sound pressures of 60–100 dB (Branchi et al., 2001).

Development of Ultrasonic Vocalisations in Infant Mice

The temporal organisation is closely related to landmark developmental stages of neonates (Elwood and Keeling, 1982). The variations in calling rate mirror the stages of development of homoeothermy in mice which can be divided into three different phases: from 2 to 5 days of age, there is near poikilothermy, this is followed by a second phase, from 8 to 15 days of age, with an increasing ability to produce body heat and a final stage from 18 days until 30 days of age where homeothermy is achieved (Nagy, 1993). Call rate increases during the first 6–7 days of age, reaches a peak at 8 days and then decreases gradually until the end of the second week after birth (Hahn et al., 1998; Thornton et al., 2005). Other developmental milestones have been associated with changes in calling rate, Noirot (1966) described that calling rate increased until D4 which coincided with the opening of the ears and decreased almost to zero on the day which the opening of eyes was recorded.

During this period, call length and frequency range characteristics also vary, declining linearly over time; on the other hand, other frequency measures (beginning and ending frequency, highest and lowest frequency of call) increased with age (Hahn et al., 1998). In addition, regarding the spectrographic characteristics of USVs, there is an increased proportion of harmonic calls over the initial 13 days (Grimsley et al., 2011).

Influence of Environmental Factors

The intensity and calling rate of ultrasonic vocalisations is inversely related to the environmental temperature at which the pups are exposed (Okon, 1970a; Sales and Skinner, 1979). Furthermore, in experimental conditions, the day that pups cease to call is related to the environmental temperature, for instance, at 33°C isolated pups ceased to call at 11 days of age and, at 2–3°C, calls ceased at 19 days of age.

Other factors that elicit ultrasounds in neonatal mice are tactile stimuli (handling, loss of balance, retrieval by the mother) (Okon, 1970b; Henessy et al., 1980; Hahn and Schanz, 2002), and odour cues (Marchlewska-Koj et al., 1999).

Cross-fostering experiments also reveal that call features are primarily dependent on the genotype of the pups and not on early environment, with the exception of call amplitude that seems to be dependent on the mother genotype (Wohr et al., 2008).

Social context also influences the characteristics of infant USVs. Odour seems to play an important role on how mouse neonates perceive the external environment in the early neonatal period, olfaction is one of the best developed senses in neonatal mice (Walz et al., 2006) and this becomes even more evident due to the fact that odour is the main cue for nipple grasping and, thus, successful suckling (Al Ain et al., 2013) which is crucial for neonatal survival.

Pups on post-natal day 8 cease to call when exposed to the odours of unfamiliar males (Branchi et al., 1998). But, an increase calling rate was reported on pups aged between 2–10 days old when exposed to odours from non-infanticidal males, defined as unfamiliar males that either ignored or showed parental toward the pup and a clear increase in call rate when the same pup was exposed to urine of non-infanticial male followed by urine of an infanticidal male (Santucci et al., 1994). Furthermore, infant mice seem to be able to distinguish genotype, Marchlewska-Koj et al. (1999) reported an increase in the calling rate of CBA pups when exposed to bedding from lactating females of a different genetic background (C57BL/6J), and this occurred even when CBA pups were fostered by C57BL/6J females. Furthermore, Kapusta and Szentgyorgyi (2004) observed an increase in calling rate when CBA pups were exposed to bedding from two unrelated strains (C57BL and DBA).

Environmental conditions have also been found to influence the spectrographic characteristics of ultrasonic vocalisations in neonatal mice, as described by Branchi et al. (1998). Eight days old CD-1 pups of both sexes exposed to five different conditions (odour from the nest, social isolation, low-temperature isolation, tactile stimuli, or odour from conspecific unfamiliar adult mice) emitted differently shaped USVs in each of the situations. More frequency steps were

observed in low frequency range calls by animals exposed to low temperature isolation or to male odour. Isolated pups produced more frequency steps calls in medium frequency range calls. Pups exposed to low temperature isolation emitted a higher number of modulated frequency signals in high frequency range calls.

Strain and Genetic Influences

Pups of different strains of mice present different characteristics of distress calls. For instance, C57BL produces less number of ultrasounds than BALB/c and SEC strains and the peak of ultrasound emission also differs between this strains with C57 reaching a peak sooner (D2–D4) than SEC strain (D4-8) with BALB/c reaching a peak in an intermediate period (D3–D7), all presenting a fundamental frequency of 60–70 kHz (Robinson and D'Udine, 1982). It is interesting to note that SEC strain carries the short ear mutation ($Bmp5^{se}$) which is related to the occurrence of short, slightly ruffled external ears due to defective cartilage framework and an abnormal skeleton with numerous local defects in homozygotes and this can potentially have an impact on the evolution of vocalisations in this strain. Bell et al. (1972) also reported rates of calling higher in the BALB/cJ strain compared to C57BL/6J and C3H/HeJ at 3 days of age but this pattern changes across time, with C3H/HeJ calling more frequently at 9 days of age when compared to the other two strains, frequency and call duration also differ among the three strains, all with a modal peak frequency situated between 70 and 80 kHz.

Furthermore, between post-natal days 2–12, C57BL/6J (B6), 129 × 1 and FVB/NJ produced a wide repertoire of calls, which included high numbers of frequency steps and complex USVs. In more detail, B6 pups emitted more downward, chevron and short calls and, less two-syllable calls than the C57BL/6J, 129 × 1 and FVB/NJ. When comparing the calls emitted by C57BL/6J, 129 × 1 and FVB/NJ with the ones emitted by outbred CD-1 pups at the same age—8 days old—the latter produced a higher percentage of frequency steps and complex calls but low numbers of flat, short and complex calls (Branchi et al., 1998; Scattoni et al., 2008).

Sex can also have an influence on call characteristics (call rate, duration, and frequency) of mouse pups, female pups emit fewer calls with a lower frequency and length when compared to males (Hahn et al., 1998).

Effects on Maternal Behaviour

Ultrasonic vocalisations of infant mice affect maternal behaviour in several ways. USVs emitted by mouse pups, stimulate maternal retrieval and nest building (Zippelius and Schleidt, 1956; Noirot, 1964, 1966; Smotherman et al., 1974). And, as Ehret and Haack (1984) suggests they seem to have two major functions, to release maternal behaviour, such as searching for a lost pup, and to serve as a cue for the pup's location.

Infant USVs seem to play a part in offspring recognition (Mogi et al., 2017) but, interestingly, its role and importance seem to differ between strains. In experiments performed in ICR mice, infant USVs seem to be enough stimulus for maternal recognition (Uematsu et al., 2007; Okabe et al., 2010) but C57BL/6 dams require simultaneous presentation with pup odour (Okabe et al., 2013).

Furthermore, ICR females are able to discriminate between own and alien pups through their USVs (Mogi et al., 2017).

Interestingly, maternal responsiveness in C57BL/6 mice is negatively correlated with pup USV calling rate (D'Amato et al., 2005).

Further studies were performed in genetically deaf mice to elucidate the importance of vocal communication in the mother-offspring interaction, D'Amato and Populin (1987) performed observations in deaf females rearing deaf pups and cross-foster experiments with normal hearing mothers and deaf pups and vice-versa. The results indicate that maternal behaviour was not affected by the deafness of the mother when taking care of her own pups. Females seemed to compensate deafness by increasing activity levels but neither nest building or pup retrieval were affected. In contrast, cross-fostering of deaf pups that emitted less calls with normal hearing mothers seemed to decrease maternal care, possibly due to the fact that the lesser number of calls was interpreted as no need for maternal care.

Broadband and Low Frequency Calls of Infant Mice

Two other types of vocalisations have been described in infant mice: broadband spectrum and low frequency calls. Broadband spectrum signals inhibit biting and injury by adults. These can be emitted post-partum, when the dam is cleaning away foetal membranes, and are characterised by a frequency between 4 and 40 kHz (Haack et al., 1983). Low-frequency wriggling calls are characterised by having a major energy below 10 kHz and a frequency range rarely exceeding 20 kHz. These stimulate maternal behaviour, such as licking of the pups, and can be emitted when pups press to reach their mother's teats, when the litter is alone inside the nest or when pups older than 5 days crawl over each other (Ehret and Bernecker, 1986).

PARENTAL COOPERATION

A new function for ultrasonic vocalisations in parental cooperation has been described. A 38 kHz USV emitted by female ICR mice was detected when dams were separated from pups, defined as pairmate-dependent retrieval as it seems to elicit retrieval of pups by the father-(Liu et al., 2013). This vocalisation was only detected in ICR mice and not in C57BL/6J or BALB/c (Liang et al., 2014), and seems to be associated with the activation of aromatase which synthetises oestrogen from androgen in several brain regions of male mice (Akther et al., 2015).

CONCLUSION AND FUTURE PERSPECTIVES

In recent years, research into mouse vocal communication has gained a lot of interest. Neonatal USVs are very well-characterised in the most common mouse strains and have been an invaluable tool in neuroscience research (Premoli et al., 2021), for instance,

in neurodevelopmental disorders such as autism (Scattoni et al., 2008, 2011; Wohr et al., 2011). But some aspects of vocal communication in breeding contexts are still poorly understood. Literature on neonatal low-frequency vocalisations is scarce and information about vocal communication in established social groups is lacking.

Most research in vocal communication has been performed in laboratory conditions due to technical constraints. Detection of sounds from these animals is usually done in highly controlled environments to avoid contamination by other sounds and requires highly specialised equipment. Social interactions are usually brief and there is a lack of data on vocal communication between established groups. So, although there is published data on laboratory and wild-derived mice, no studies were performed in naturalistic conditions. Nevertheless, efforts have been made to surpass this, and de Chaumont et al. (2021) have recently developed a sophisticated apparatus that allows the recording of video and sounds during long periods of time which can provide novel insights into vocal communication and its correlation with behaviour.

In laboratory animal facilities, breeding is mainly performed in trios where two females are kept with the male continuously. This allows females to nest communally and also permits the occurrence of post-partum mating. Communal nesting occurs in wild mice and poses several advantages such as sharing of rearing tasks and, even, communal nursing (Weber and Olsson, 2008). As such, it would be expected that females communicate vocally in these contexts and that this would have an impact on maternal care and pup survival. But the current literature only addresses vocal communication in same sex or heterosexual encounters (Warren et al., 2018, 2020; Sasaki et al., 2020; de Chaumont et al., 2021). It is know that females vocalise in same-sex encounters and produce more complex USVs in these situations (Matsumoto and Okanoya, 2018) but it is still unknown how this would be affected by the presence of a male.

Further studies are needed to understand the social dynamics between females in a breeding trio and the potential impact on pup survival. The development of home cage monitoring systems with the ability to detect vocalisations will be a useful tool to fill this gap of knowledge.

Laboratory mouse strains have different profiles with regards to their hearing range and susceptibility for hearing loss (Davis et al., 2001). The most prominent example is the well-characterised age-related hearing loss in C57BL/6J mice which is caused by a mutation in cadherin 23 (Noben-Trauth et al., 2003). Hearing capabilities of this strain start to decline as soon as 3 months old; higher frequencies are initially affected, and by 1 year of age, these mice present a severe-to-profound high frequency hearing loss (Ison et al., 2007). Furthermore, sex differences have been reported with females presenting a more severe degree of hearing loss when compared to males of the same age (Henry, 2002). On the other hand, CBA/CaJ mice maintain relatively stable hearing over age, making them a preferred strain to use in vocal communication studies (Erway et al., 1993). It is not clear how gradual hearing loss can impact mouse vocal communication, and especially, communication between dam and pup since most studies were performed in genetically modified strains which are deaf since birth (D'Amato et al., 2005) or using artificially induced deafness (Ehret and Bernecker, 1986). More research is needed to elucidate on the impact of hearing loss in vocal communication in a breeding context in laboratory strains.

The spectrographic analysis of USVs in mouse communication is also one of the aspects that has gained a lot of research interest in the last decade. Mice produce different shapes of USVs (see **Table 1**) that can potentially deliver different messages to the receiver. Spectographic analysis has been performed to gain insight into the social deficits of autism models, such as the BTBR strain (Scattoni et al., 2008, 2011). Several authors hypothesised about the meaning of these different types of USVs that were first described by Scattoni et al. (2008), Branchi et al. (1998) suggested that, since high frequency sounds travel for a short distance and low frequency sounds for a longer distance, the first could serve to be used in intraspecific communication (i.e., intralitter, nest area range) while the latter would be used in other types of intraspecific communication, such as out of the nest range. Grimsley et al. (2011) also suggested that the changing proportions of different syllable types could be used as cues for mothers to distinguish different aged pups. Matsumoto and Okanoya (2018) hypothesised that since females interacting in same-sex groups produce more complex calls and males tend to produce simple calls, these could be useful to maintaining group structure. Further research is needed to elucidate the context and behavioural meaning of different syllable types.

AUTHOR CONTRIBUTIONS

SC-P performed the literature research and wrote the first draft. SC-P and IO designed the manuscript. All authors discussed, provided input, contributed to the article, and approved the final manuscript.

FUNDING

This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2022.833168/full#supplementary-material

REFERENCES

Akther, S., Huang, Z., Liang, M., Zhong, J., Fakhrul, A. A., Yuhi, T., et al. (2015). Paternal Retrieval Behavior Regulated by Brain Estrogen Synthetase (Aromatase) in Mouse Sires that Engage in Communicative Interactions with Pairmates. *Front. Neurosci.* 9:450. doi: 10.3389/fnins.2015.00450

Al Ain, S., Belin, L., Schaal, B., and Patris, B. (2013). How does a newly born mouse get to the nipple? Odor substrates eliciting first nipple grasping and sucking responses. *Dev. Psychobiol.* 55, 888–901. doi: 10.1002/dev.21082

Asaba, A., Okabe, S., Nagasawa, M., Kato, M., Koshida, N., Osakada, T., et al. (2014). Developmental social environment imprints female preference for male song in mice. *PLoS One* 9:e87186. doi: 10.1371/journal.pone.0087186

Barthelemy, M., Gourbal, B. E., Gabrion, C., and Petit, G. (2004). Influence of the female sexual cycle on BALB/c mouse calling behaviour during mating. *Naturwissenschaften* 91, 135–138. doi: 10.1007/s00114-004-0501-4

Bell, R. W., Nitschke, W., and Zachman, T. A. (1972). Ultra-sounds in three inbred strains of young mice. *Behav. Biol.* 7, 805–814. doi: 10.1016/s0091-6773(72)80172-x

Bind, R. H., Minney, S. M., Rosenfeld, S., and Hallock, R. M. (2013). The role of pheromonal responses in rodent behavior: future directions for the development of laboratory protocols. *J. Am. Assoc. Lab. Anim. Sci.* 52, 124–129.

Blanchard, R. J., Hebert, M. A., Ferrari, P., Palanza, P., Figueira, R., Blanchard, D. C., et al. (1998). Defensive behaviors in wild and laboratory (Swiss) mice: the mouse defense test battery. *Physiol. Behav.* 65, 201–209. doi: 10.1016/s0031-9384(98)00012-2

Branchi, I., Santucci, D., and Alleva, E. (2001). Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behav. Brain Res.* 125, 49–56. doi: 10.1016/s0166-4328(01)00277-7

Branchi, I., Santucci, D., Vitale, A., and Alleva, E. (1998). Ultrasonic Vocalizations by Infant Laboratory Mice: a Preliminary Spectographic Characterization under Different Conditions. *Dev. Psychobiol.* 33, 249–256. doi: 10.1002/(sici)1098-2302(199811)33:3<249::aid-dev5>3.0.co;2-r

Byatt, S., and Nyby, J. (1986). Hormonal regulation of chemosignals of female mice that elicit ultrasonic vocalizations from males. *Horm. Behav.* 20, 60–72. doi: 10.1016/0018-506x(86)90029-2

Chabout, J., Sarkar, A., Dunson, D. B., and Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Front. Behav. Neurosci.* 9:76. doi: 10.3389/fnbeh.2015.00076

D'Amato, F., Scalera, E., Sarli, C., and Moles, A. (2005). Pups Call, Mothers Rush: does Maternal Responsiveness Affect the Amount of Ultrassonic Vocalizations in Mouse Pups? *Behav. Genet.* 35, 103–112. doi: 10.1007/s10519-004-0860-9

D'Amato, F. R., and Populin, R. (1987). Mother-offspring interaction and pup development in genetically deaf mice. *Behav. Genet.* 17, 465–475. doi: 10.1007/BF01073113

Davis, R. R., Newlander, J. K., Ling, X.-B., Cortopassi, G. A., Krieg, E. F., and Erway, L. C. (2001). Genetic basis for susceptibility to noise-induced hearing loss in mice. *Hear. Res.* 155, 82–90. doi: 10.1016/s0378-5955(01)00250-7

de Chaumont, F., Lemiere, N., Coqueran, S., Bourgeron, T., and Ey, E. (2021). LMT USV Toolbox, a Novel Methodological Approach to Place Mouse Ultrasonic Vocalizations in Their Behavioral Contexts-A Study in Female and Male C57BL/6J Mice and in Shank3 Mutant Females. *Front. Behav. Neurosci.* 15:735920. doi: 10.3389/fnbeh.2021.735920

Dizinno, G., and Whitney, G. (1977). Androgen influence on male mouse ultrasounds during courtship. *Horm. Behav.* 8, 188–192. doi: 10.1016/0018-506x(77)90035-6

Ebbesen, C. L., and Froemke, R. C. (2021). Body language signals for rodent social communication. *Curr. Opin. Neurobiol.* 68, 91–106. doi: 10.1016/j.conb.2021.01.008

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

Ehret, G., and Haack, B. (1981). Categorical perception of mouse pup ultrasound by lactating females. *Naturwissenschaften* 68, 208–209. doi: 10.1007/BF01047208

Ehret, G., and Haack, B. (1984). Motivation and Arousal Influence Sound-Induced Maternal Pup-retrieving Behavior in Lactating House Mice. *Z. Tierpsychol.* 65, 25–39. doi: 10.1111/j.1439-0310.1984.tb00370.x

Elwood, R. W., and Keeling, F. (1982). Temporal organization of ultrasonic vocalizations in infant mice. *Dev. Psychobiol.* 15, 221–227. doi: 10.1002/dev.420150306

Erway, L. C., Willott, J. F., Archer, J. R., and Harrison, D. E. (1993). Genetics of age-related hearing loss in mice: I. Inbred and F1 hybrid strains. *Hear. Res.* 65, 125–132. doi: 10.1016/0378-5955(93)90207-h

Finton, C. J., Keesom, S. M., Hood, K. E., and Hurley, L. M. (2017). What's in a squeak? Female vocal signals predict the sexual behaviour of male house mice during courtship. *Anim. Behav.* 126, 163–175. doi: 10.1016/j.anbehav.2017.01.021

Grimsley, J. M., Monaghan, J. J., and Wenstrup, J. J. (2011). Development of social vocalizations in mice. *PLoS One* 6:e17460. doi: 10.1371/journal.pone.0017460

Haack, B., Markl, H., and Ehret, G. (1983). "Sound Communication Between Parents and Offspring," in *The Auditory Psychobiology of the Mouse*, ed. J. F. Willott (Springfield, Illinois: Charles C. Thomas), 57–97. doi: 10.11647/obp.0097.06

Hahn, M. E., Hewitt, J. K., Adams, M., and Tully, T. (1987). Genetic influences on ultrasonic vocalizations in young mice. *Behav. Genet.* 17, 155–166. doi: 10.1007/BF01065994

Hahn, M. E., Karkowski, L., Weinreb, L., Henry, A., Schanz, N., and Hahn, E. M. (1998). Genetic and developmental influences on infant mouse ultrasonic calling. II. Developmental patterns in the calls of mice 2-12 days of age. *Behav. Genet.* 28, 315–325. doi: 10.1023/a:1021679615792

Hahn, M. E., and Schanz, N. (2002). The effects of cold, rotation, and genotype on the production of ultrasonic calls in infant mice. *Behav. Genet.* 32, 267–273. doi: 10.1023/a:1019728813891

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., and Fischer, J. (2009). Female mice respond to male ultrasonic 'songs' with approach behaviour. *Biol. Lett.* 5, 589–592. doi: 10.1098/rsbl.2009.0317

Hanson, J. L., and Hurley, L. M. (2012). Female presence and estrous state influence mouse ultrasonic courtship vocalizations. *PLoS One* 7:e40782. doi: 10.1371/journal.pone.0040782

Henessy, M. B., Li, J., Lowe, E. L., and Levine, S. (1980). Maternal Behavior, Pup Vocalizations, and Temperature Changes Following Handling in Mice of 2 Inbred Strains. *Dev. Psychobiol.* 13, 573–584. doi: 10.1002/dev.420130603

Henry, K. R. (2002). Sex- and age-related elevation of cochlear nerve envelope response (CNER) and auditory brainstem response (ABR) thresholds in C57BL/6 mice. *Hear. Res.* 170, 107–115. doi: 10.1016/s0378-5955(02)00391-x

Hoffmann, F., Musolf, K., and Penn, D. J. (2009). Freezing urine reduces its efficacy for eliciting ultrasonic vocalizations from male mice. *Physiol. Behav.* 96, 602–605. doi: 10.1016/j.physbeh.2008.12.014

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

Ison, J. R., Allen, P. D., and O'neill, W. E. (2007). Age-related hearing loss in C57BL/6J mice has both frequency-specific and non-frequency-specific components that produce a hyperacusis-like exaggeration of the acoustic startle reflex. *J. Assoc. Res. Otolaryngol.* 8, 539–550. doi: 10.1007/s10162-007-0098-3

Kanno, K., and Kikusui, T. (2018). Effect of Sociosexual Experience and Aging on Number of Courtship Ultrasonic Vocalizations in Male Mice. *Zool. Sci.* 35, 208–214. doi: 10.2108/zs170175

Kapusta, J., and Szentgyorgyi, H. (2004). Ultrasonic Response of CBA Pups to the Odour of Genetically Different Mice. *Behaviour* 141, 157–167. doi: 10.1163/156853904322890799

Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K., and Okanoya, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PLoS One* 6:e17721. doi: 10.1371/journal.pone.0017721

Liang, M., Zhong, J., Liu, H. X., Lopatina, O., Nakada, R., Yamauchi, A. M., et al. (2014). Pairmate-dependent pup retrieval as parental behavior in male mice. *Front. Neurosci.* 8:186. doi: 10.3389/fnins.2014.00186

Liu, H. X., Lopatina, O., Higashida, C., Fujimoto, H., Akther, S., Inzhutova, A., et al. (2013). Displays of paternal mouse pup retrieval following communicative interaction with maternal mates. *Nat. Commun.* 4:1346. doi: 10.1038/ncomms2336

Lupanova, A. S., and Egorova, M. A. (2015). Vocalization of sex partners in the house mouse (Mus Musculus). *J. Evol. Biochem. Physiol.* 51, 324–331. doi: 10.1134/s0022093015040080

Maggio, J. C., and Whitney, G. (1986). Heterosis of adult mouse (Mus musculus) ultrasonic vocalizing. *Behav. Genet.* 16, 493–506. doi: 10.1007/BF01074267

Mahrt, E., Agarwal, A., Perkel, D., Portfors, C., and Elemans, C. P. (2016). Mice produce ultrasonic vocalizations by intra-laryngeal planar impinging jets. *Curr. Biol.* 26, R880–R881. doi: 10.1016/j.cub.2016.08.032

Marchlewska-Koj, A., Kapusta, J., and Olejniczak, P. (1999). Ultrasonic Response of CBA Newborn Mice to Bedding Odour. *Behaviour* 136, 269–278. doi: 10.1163/156853999501324

Marconi, M. A., Nicolakis, D., Abbasi, R., Penn, D. J., and Zala, S. M. (2020). Ultrasonic courtship vocalizations of male house mice contain distinct individual signatures. *Anim. Behav.* 169, 169–197. doi: 10.1016/j.anbehav.2020.09.006

Matsumoto, Y. K., and Okanoya, K. (2016). Phase-Specific Vocalizations of Male Mice at the Initial Encounter during the Courtship Sequence. *PLoS One* 11:e0147102. doi: 10.1371/journal.pone.0147102

Matsumoto, Y. K., and Okanoya, K. (2018). Mice modulate ultrasonic calling bouts according to sociosexual context. *R. Soc. Open Sci.* 5:180378. doi: 10.1098/rsos.180378

Melotti, L., Siestrup, S., Peng, M., Vitali, V., Dowling, D., Von Kortzfleisch, V. T., et al. (2021). Individuality, as well as genetic background, affects syntactical features of courtship songs in male mice. *Anim. Behav.* 180, 179–196. doi: 10.1016/j.anbehav.2021.08.003

Mogi, K., Takakuda, A., Tsukamoto, C., Ooyama, R., Okabe, S., Koshida, N., et al. (2017). Mutual mother-infant recognition in mice: the role of pup ultrasonic vocalizations. *Behav. Brain Res.* 325, 138–146. doi: 10.1016/j.bbr.2016.08.044

Musolf, K., Hoffmann, F., and Penn, D. J. (2010). Ultrasonic courtship vocalizations in wild house mice, Mus musculus musculus. *Anim. Behav.* 79, 757–764. doi: 10.1016/j.anbehav.2009.12.034

Musolf, K., and Penn, D. J. (2012). "Ultrasonic vocalizations in house mice: a cryptic mode of acoustic communication," in *The Evolution of the House Mouse*, eds M. MacholáN, S. J. E. Baird, P. Munclinger, and J. PiáLek (Cambridge, UK: Cambridge University Press).

Nagasawa, M., Okabe, S., Mogi, K., and Kikusui, T. (2012). Oxytocin and mutual communication in mother-infant bonding. *Front. Hum. Neurosci.* 6:31. doi: 10.3389/fnhum.2012.00031

Nagy, Z. M. (1993). Development of homeothermy in infant C3H mice. *Bull. Psychon. Soc.* 31, 221–224. doi: 10.3758/bf03337329

Navarro, M., Ruberte, J., and Carretero, A. (2017a). "Respiratory apparatus," in *Morphological Mouse Phenotyping*, eds J. Ruberte, A. Carretero, and M. Navarro (Madrid: Academic Press), 147–178.

Navarro, M., Ruberte, J., and Carretero, A. (2017b). "Vestibulocochlear organ," in *Morphological Mouse Phenotyping*, eds J. Ruberte, A. Carretero, and M. Navarro (Cambridge, MA: Academic Press), 521–539.

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., and Egnor, S. E. (2015). Female mice ultrasonically interact with males during courtship displays. *Elife* 4:e06203. doi: 10.7554/eLife.06203

Nicolakis, D., Marconi, M. A., Zala, S. M., and Penn, D. J. (2020). Ultrasonic vocalizations in house mice depend upon genetic relatedness of mating partners and correlate with subsequent reproductive success. *Front. Zool.* 17:10. doi: 10.1186/s12983-020-00353-1

Noben-Trauth, K., Zheng, Q. Y., and Johnson, K. R. (2003). Association of cadherin 23 with polygenic inheritance and genetic modification of sensorineural hearing loss. *Nat. Genet.* 35, 21–23. doi: 10.1038/ng1226

Noirot, E. (1964). Changes in Responsiveness to Young in the Adult Mouse: The Effect of External Stimuli. *J. Comp. Physiol. Psychol.* 57, 97–99. doi: 10.1037/h0042864

Noirot, E. (1966). Ultra-sounds in young rodents. I. Changes with age in albino mice. *Anim. Behav.* 14, 459–462. doi: 10.1016/s0003-3472(66)80045-3

Nyby, J., Dizinno, G. A., and Whitney, G. (1976). Social status and ultrasonic vocalizations of male mice. *Behav. Biol.* 18, 285–289. doi: 10.1016/s0091-6773(76)92198-2

Nyby, J., Wysocki, C. J., Whitney, G., and Dizinno, G. (1977). Pheromonal regulation of male mouse ultrasonic courtship (Mus musculus). *Anim. Behav.* 25, 333–341. doi: 10.1016/0003-3472(77)90009-4

Okabe, S., Nagasawa, M., Kihara, T., Kato, M., Harada, T., Koshida, N., et al. (2010). The effects of social experience and gonadal hormones on retrieving behavior of mice and their responses to pup ultrasonic vocalizations. *Zool. Sci.* 27, 790–795. doi: 10.2108/zsj.27.790

Okabe, S., Nagasawa, M., Kihara, T., Kato, M., Harada, T., Koshida, N., et al. (2013). Pup odor and ultrasonic vocalizations synergistically stimulate maternal attention in mice. *Behav. Neurosci.* 127, 432–438. doi: 10.1037/a0032395

Okon, E. E. (1970a). The effect of envrionmental temperature on the production of ultrasounds by isolated non-handled albino mouse pups. *J. Zool.* 162, 71–83. doi: 10.1111/j.1469-7998.1970.tb01258.x

Okon, E. E. (1970b). The ultrasonic responses of albino mouse pups to tactile stimuli. *J. Zool. Lond.* 162, 485–492. doi: 10.1111/j.1469-7998.1970.tb01283.x

Pomerantz, S. M., Nunez, A. A., and Bean, N. J. (1983). Female Behavior is Affected by Male Ultrasonic Vocalizations in House Mice. *Physiol. Behav.* 31, 91–96. doi: 10.1016/0031-9384(83)90101-4

Portfors, C. V., and Perkel, D. J. (2014). The role of ultrasonic vocalizations in mouse communication. *Curr. Opin. Neurobiol.* 28, 115–120. doi: 10.1016/j.conb.2014.07.002

Premoli, M., Memo, M., and Bonini, S. A. (2021). Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen. Res.* 16, 1158–1167. doi: 10.4103/1673-5374.300340

Riede, T., Borgard, H. L., and Pasch, B. (2017). Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. *R. Soc. Open Sci.* 4:170976. doi: 10.1098/rsos.170976

Roberts, L. H. (1975). The rodent ultrasound production mechanism. *Ultrasonics* 13, 83–88. doi: 10.1016/0041-624x(75)90052-9

Robinson, D. J., and D'Udine, B. (1982). Ultrasonic calls produced by three laboratory strains of *Mus musculus*. *J. Zool.* 197, 383–389. doi: 10.1111/jzo.1982.197.3.383

Ronald, K. L., Zhang, X., Morrison, M. V., Miller, R., and Hurley, L. M. (2020). Male mice adjust courtship behavior in response to female multimodal signals. *PLoS One* 15:e0229302. doi: 10.1371/journal.pone.0229302

Sales, G. D. (1972). Ultrasound and mating behaviour in rodents with some observations on other behavioural situations. *J. Zool.* 168, 149–164. doi: 10.1111/j.1469-7998.1972.tb01345.x

Sales, G. D., and Skinner, N. C. (1979). The effect of ambient temperature on body temperature and on ultrasonic behaviour in litters of albino laboratory mice deprived of their mothers. *J. Zool. Lond.* 187, 265–281. doi: 10.1111/j.1469-7998.1979.tb03948.x

Santucci, D., Masterson, D., and Elwood, R. W. (1994). Effects of age, sex, and odour from conspecific adult males on ultrasonic vocalizations of infant CS1 mice. *Behav. Process.* 32, 285–296. doi: 10.1016/0376-6357(94)90048-5

Sasaki, E., Tomita, Y., and Kanno, K. (2020). Sex differences in vocalizations to familiar or unfamiliar females in mice. *R. Soc. Open Sci.* 7:201529. doi: 10.1098/rsos.201529

Scattoni, M. L., Gandhy, S. U., Ricceri, L., and Crawley, J. N. (2008). Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS One* 3:e3067. doi: 10.1371/journal.pone.0003067

Scattoni, M. L., Ricceri, L., and Crawley, J. N. (2011). Unusual repertoire of vocalizations in adult BTBR T+tf/J mice during three types of social encounters. *Genes Brain Behav.* 10, 44–56. doi: 10.1111/j.1601-183X.2010.00623.x

Seagraves, K. M., Arthur, B. J., and Egnor, S. E. (2016). Evidence for an audience effect in mice: male social partners alter the male vocal response to female cues. *J. Exp. Biol.* 219, 1437–1448. doi: 10.1242/jeb.129361

Smotherman, W. P., Bell, R. W., Starzec, J., Elias, J., and Zachman, T. A. (1974). Maternal responses to infant vocalizations and olfactory cues in rats and mice. *Behav. Biol.* 12, 55–66. doi: 10.1016/s0091-6773(74)91026-8

Sugimoto, H., Okabe, S., Kato, M., Koshida, N., Shiroishi, T., Mogi, K., et al. (2011). A role for strain differences in waveforms of ultrasonic vocalizations during male-female interaction. *PLoS One* 6:e22093. doi: 10.1371/journal.pone.0022093

Tasaka, G. I., Feigin, L., Maor, I., Groysman, M., Denardo, L. A., Schiavo, J. K., et al. (2020). The Temporal Association Cortex Plays a Key Role in Auditory-Driven Maternal Plasticity. *Neuron* 107:e567. doi: 10.1016/j.neuron.2020.05.004

Thornton, L. M., Hahn, M. E., and Schanz, N. (2005). Genetic and Developmental Influences on Infant Mouse Ultrasonic Calling. III. Patterns of Inheritance in the Calls of Mice 3-9 Days of Age. *Behav. Genet.* 35, 73–83. doi: 10.1007/s10519-004-0857-4

Uematsu, A., Kikusui, T., Kihara, T., Harada, T., Kato, M., Nakano, K., et al. (2007). Maternal approaches to pup ultrasonic vocalizations produced by a nanocrystalline silicon thermo-acoustic emitter. *Brain Res.* 1163, 91–99. doi: 10.1016/j.brainres.2007.05.056

Walz, A., Omura, M., and Mombaerts, P. (2006). Development and topography of the lateral olfactory tract in the mouse: imaging by genetically encoded

and injected fluorescent markers. *J. Neurobiol.* 66, 835–846. doi: 10.1002/neu.20266

Warburton, V. L., Sales, G. D., and Milligan, S. R. (1989). The emission and elicitation of mouse ultrasonic vocalizations: the effects of age, sex and gonadal status. *Physiol. Behav.* 45, 41–47. doi: 10.1016/0031-9384(89)90164-9

Warren, M. R., Clein, R. S., Spurrier, M. S., Roth, E. D., and Neunuebel, J. P. (2020). Ultrashort-range, high-frequency communication by female mice shapes social interactions. *Sci. Rep.* 10:2637. doi: 10.1038/s41598-020-59418-0

Warren, M. R., Spurrier, M. S., Roth, E. D., and Neunuebel, J. P. (2018). Sex differences in vocal communication of freely interacting adult mice depend upon behavioral context. *PLoS One* 13:e0204527. doi: 10.1371/journal.pone.0204527

Weber, E. M., and Olsson, I. A. S. (2008). Maternal behaviour in Mus musculus sp.: an ethological review. *Appl. Anim. Behav. Sci.* 114, 1–22. doi: 10.1016/j.applanim.2008.06.006

White, N. R., Prasad, M., Barfield, R. J., and Nyby, J. G. (1998). 40- and 70-kHz Vocalizations of Mice (Mus musculus) during Copulation. *Physiol. Behav.* 63, 467–473. doi: 10.1016/s0031-9384(97)00484-8

Whitney, G., Alpern, M., Dizinno, G., and Horowitz, G. (1974). Female odors evoke ultrasounds from male mice. *Anim. Learn. Behav.* 2, 13–18. doi: 10.3758/bf03199109

Whitney, G., Coble, J. R., Stockton, M. D., and Tilson, E. F. (1973). Ultrasonic emissions: do they facilitate courtship of mice. *J. Comp. Physiol. Psychol.* 84, 445–452. doi: 10.1037/h0034899

Whitney, G. D. (1969). Vocalization of mice: a single genetic unit effect. *J. Hered.* 60, 337–340. doi: 10.1093/oxfordjournals.jhered.a108009

Wohr, M., Dahlhoff, M., Wolf, E., Holsboer, F., Schwarting, R. K., and Wotjak, C. T. (2008). Effects of genetic background, gender, and early environmental factors on isolation-induced ultrasonic calling in mouse pups: an embryo-transfer study. *Behav. Genet.* 38, 579–595. doi: 10.1007/s10519-008-9221-4

Wohr, M., Roullet, F. I., and Crawley, J. N. (2011). Reduced scent marking and ultrasonic vocalizations in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav.* 10, 35–43. doi: 10.1111/j.1601-183X.2010.00582.x

Zippelius, H.-M., and Schleidt, W. M. (1956). Ultraschall-Laute bei jungen Mäusen. *Die Naturwissenschaften* 43, 502–502. doi: 10.1007/bf00632534

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fnbeh.2022.884688

Effects of Congenital Blindness on Ultrasonic Vocalizations and Social Behaviors in the ZRDBA Mouse

*Nouhaila Bouguiyoud[1,2], Elena Morales-Grahl[3], Gilles Bronchti[1], Johannes Frasnelli[1,2], Florence I. Roullet[4] and Syrina Al Aïn[1,2]**

[1] Department of Anatomy, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, [2] Cognition, Neurosciences, Affect et Comportement (CogNAC) Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, [3] Carleton College, Northfield, MN, United States, [4] Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada

Mice produce ultrasonic vocalizations (USVs) at different ages and social contexts, including maternal-pup separation, social play in juveniles, social interactions, and mating in adults. The USVs' recording can be used as an index of sensory detection, internal state, and social motivation. While sensory deprivation may alter USVs' emission and some social behaviors in deaf and anosmic rodents, little is known about the effects of visual deprivation in rodents. This longitudinal study aimed to assess acoustic communication and social behaviors using a mouse model of congenital blindness. Anophthalmic and sighted mice were assayed to a series of behavioral tests at three different ages, namely, the maternal isolation-induced pup USV test and the home odor discrimination and preference test on postnatal day (PND) 7, the juvenile social test on PND 30–35, and the female urine-induced USVs and scent-marking behavior at 2–3 months. Our results evidenced that (1) at PND 7, USVs' total number between both groups was similar, all mice vocalized less during the second isolation period than the first period, and both phenotypes showed similar discrimination and preference, favoring exploration of the home bedding odor; (2) at PND 30–35, anophthalmic mice engaged less in social behaviors in the juvenile play test than sighted ones, but the number of total USVs produced is not affected; and (3) at adulthood, when exposed to a female urine spot, anophthalmic male mice displayed faster responses in terms of USVs' emission and sniffing behavior, associated with a longer time spent exploring the female urinary odor. Interestingly, acoustic behavior in the pups and adults was correlated in sighted mice only. Together, our study reveals that congenital visual deprivation had no effect on the number of USVs emitted in the pups and juveniles, but affected the USVs' emission in the adult male and impacted the social behavior in juvenile and adult mice.

Keywords: congenital blindness, acoustic communication, social behavior, mice, development

OPEN ACCESS

Edited by:
Sara Anna Bonini, University of Brescia, Italy

Reviewed by:
Maya Opendak, Johns Hopkins University, United States
Anne-Marie Mouly, INSERM U1028 Centre de Recherche en Neurosciences de Lyon, France

***Correspondence:**
Syrina Al Aïn
syrina.alain@uqtr.ca

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *26 February 2022*
Accepted: *29 March 2022*
Published: *03 May 2022*

Citation:
Bouguiyoud N, Morales-Grahl E, Bronchti G, Frasnelli J, Roullet FI and Al Aïn S (2022) Effects of Congenital Blindness on Ultrasonic Vocalizations and Social Behaviors in the ZRDBA Mouse. Front. Behav. Neurosci. 16:884688. doi: 10.3389/fnbeh.2022.884688

INTRODUCTION

In the animal kingdom, most of the species emit vocalizations in response to various social stimuli. House mice are known to produce mainly ultrasonic vocalizations (USVs), characterized by fundamental frequency spanning a range of 35–110 kHz (Sales, 1972; Branchi et al., 1998; Holy and Guo, 2005). A vast number of studies have documented that the USVs' features may be

modulated as a function of social contexts and the developmental stage of the mouse emitter (Nyby, 1983; Maggio and Whitney, 1985; Ehret, 2005; Holy and Guo, 2005; Wang et al., 2008; Williams et al., 2008; Grimsley et al., 2011; Hanson and Hurley, 2012; Ey et al., 2013; Sangiamo et al., 2020). Specifically, vocalization behavior was first studied in rodent pups during the isolation of the pups from their mothers and littermates, resulting in USVs' calls between birth and postnatal day (PND) 14 with a peak emission at the age of 7–9 days (Sales, 1972; Branchi et al., 1998; Ehret, 2005; Fischer and Hammerschmidt, 2011). USVs' emission has been reported in juvenile mice during social interactions/play with an age-/sex-matched congener (Panksepp et al., 2007), just like in both male and female adult mice during dyadic encounters, courtship, and mating (Pomerantz et al., 1983; Maggio and Whitney, 1985; Holy and Guo, 2005; Hammerschmidt et al., 2009; Scattoni et al., 2009; Grimsley et al., 2011; Roullet et al., 2011; Wöhr and Schwarting, 2013; von Merten et al., 2014). The acoustic behavior is also regulated by internal state, such as the strength of arousal and emotion (Brudzynski, 2013; Gaub et al., 2016; Grimsley et al., 2016; Boulanger-Bertolus et al., 2017; Demir et al., 2020), and by external factors/conditions, such as the presence of a predator/attractive congener (Sales, 1972; Mun et al., 2015). Furthermore, social behavior and communication are tightly linked as demonstrated by other studies focusing on acoustic communication and aberrant social interactions in models of neurodevelopmental disorders, such as mouse models of autism spectrum disorder or fragile X syndrome (Jamain et al., 2008; Scattoni et al., 2008; Bozdagi et al., 2010; Wöhr et al., 2011; Schmeisser et al., 2012; Ey et al., 2013; Wöhr, 2014; Belagodu et al., 2016). Taken together, quantifying USVs' emission in diverse social contexts helps in assessing the dynamics underlying socioaffective communication in rodent models, and is thus relevant to better interpret alterations in vocal communication and sociability seen in rodent models of disorders (Wöhr and Scattoni, 2013).

Importantly, USVs' emission may rely on the integrity of sensory systems as sensory—mainly auditory and olfactory—disruption leads to substantial changes in USVs' features. Indeed, early deafened mice emitted intact USVs' rates in pup isolation and male's courtship contexts (Hammerschmidt et al., 2012; Mahrt et al., 2013), whereas late deafened male mice resulted in increased female urine-induced social vocalizations (Arriaga et al., 2012). Moreover, disruption of the vomeronasal system led to considerable reduction of USVs' levels emitted by males in response to female stimulus, although $ZnSO_4$-induced anosmia did not alter the USV numbers (Bean, 1982). Notwithstanding, the importance of the visual inputs on acoustic communication and related social behaviors has received little attention so far. Interestingly, Langford et al. (2006) highlighted that, in mice, visual cues are required to trigger empathic responses toward a congener in pain as an opaque screen abolished empathic responses, whereas deaf and anosmia did not. Nevertheless, congenitally blind women displayed increased vocalizations toward their newborn, accompanied by heightened duration of contact/proximity and breastfeeding compared with sighted dyads (Thoueille et al., 2006; Chiesa et al., 2015; Ganea et al., 2018). To the best of our knowledge, the link between visual inputs, acoustic communication, and related social behaviors has not been examined in rodent models of visual deprivation.

This study aimed to investigate USVs and social behaviors across development in a congenitally blind mouse model, called ZRDBA strain. Both anophthalmic and sighted phenotypes were assayed to four behavioral tests, namely, (1) the maternal isolation-induced pup USV test consists in recording USVs numbers emitted by PND 7 pups for twice 5-min periods of maternal isolation; (2) the home odor discrimination and preference test consists in measuring USVs' levels associated to time spent exploring the home and clean bedding odors in PND 7 pups; (3) the juvenile social test consists in quantifying USVs' calls, social and nonsocial behaviors of an experimental mouse exposed to a non-familiar congener (sex and age matched); and (4) the female urine-induced USVs and scent-marking behavior test consists in exposing adult males to female urine and recording their USVs emission and sniffing/markings behaviors.

The ZRDBA strain has been obtained by crossbreeding between the anophthalmic ZRDCT and the sighted DBA-6 strains (Touj et al., 2019). The anophthalmic ZRDCT mice have orbits but neither eyes, nor optic tracts and afferents retina-hypothalamus due to a mutation on chromosome 18 of the Rx/Rax gene (Chase and Chase, 1941). The crossbreeding results in a litter with an equal number of anophthalmic Rx/Rax homozygous and sighted Rx/Rax heterozygous pups. Interestingly, a deformation-based morphometry study conducted on ZRDBA adult mice highlighted structural alterations of the ventromedial hypothalamus, the preoptic area, and the bed nucleus of stria terminalis (Touj et al., 2020, 2021); these cerebral regions being implicated in the mediation of the social communication and social behaviors, such as aggression, mating, parental care, and defense (Lebow and Chen, 2016). In the light of these neuroimaging data, we hypothesized that anophthalmic mice might show deficits in social behavior associated with lower numbers of USVs compared with sighted mice (once eyes opening).

METHODS

Animals

In this study, we used 26 anophthalmic mice (10 females and 16 males) and 20 sighted mice (four females and 16 males) in total. All mice were bred and housed in mixed cages of 3–6 blind and sighted individuals in the animal facility. Standard and constant housing conditions were applied including a 14/10 h light/dark cycle, a free and *ad libitum* access for water/food, and a controlled room temperature set at 21°C and 40–60% humidity. Pups were tattooed on the paw for identification at PND 1. In the ZRDBA strain, breeding a blind mouse with a sighted mouse generates litters comprising half the pups born blind (homozygous, results in the absence of eyes and optic nerves), and the other half born sighted (heterozygous). All experimental procedures were approved by the animal care committee of Université du Québec à Trois-Rivières in accordance with the Canadian Council on Animal Care guidelines.

Behavioral Procedures

Behavioral experiments carried out during the light phase between 9:00 a.m. and 5:00 p.m. All mice were assayed to the following behavioral tests conducted at three developmental stages (see **Figures 1**, **2**), namely, (1) at PND 7, the maternal isolation-induced pup USV test, followed by the home odor discrimination and preference test; (2) at PND 30–35, the juvenile social test; and (3) at 2–3 months, the female urine-induced USVs and scent-marking behavior test. At PND 7, all the pups from all litters were assayed to the maternal isolation-induced pup USVs test and the home odor discrimination and preference test ($n = 26$ blind and 20 sighted mice). At PND 35, all of the mice from all litters were tested as experimental mice in the juvenile social test, except for mice used as social stimuli ($n = 15$ blind and 14 sighted). All males were assessed in female urine-induced USVs and scent-marking behavior ($n = 16$ blind and 16 sighted mice).

All behavioral recordings and codings were performed using Ethovision XT software (Noldus, VA, USA). To record USVs' rates, an ultrasonic microphone was suspended above or attached to the wall of the experimental cage according to the behavioral tests. USVs' recordings were processed and analyzed through the UltraVox XT system (Noldus Information Technology, The Netherlands). The system can record the full spectrum of sound and has a maximum frequency of 160 kHz.

Maternal Isolation-Induced Pup USV Test

The maternal isolation-induced pup USV test was adapted from Branchi et al. (2006). A 30-min habituation period was performed to acclimate the dams and their littermates to the experimental room. Dams were then isolated from their littermates and placed in an individual cage (28 × 18 × 12 cm). The testing chamber was a sound-attenuation chamber (24 × 24 × 27.5 cm) with a temperature-controlled pad set at 22°C. The testing consisted in, first, introducing a PND 7 pup within the testing chamber and recording the rate of USVs' calls produced for 5 min using an ultrasonic microphone suspended at 5 cm above the center of the testing chamber (isolation 1). Secondly, the pup was placed into the cage containing his dam for 3 min (maternal-pup reunion). Finally, the pup was reintroduced into the testing chamber, and USVs' rates were again recorded for 5 min (isolation 2).

Home Odor Discrimination and Preference Test

The home odor discrimination and preference test, adapted from Roullet et al. (2010) and Meyer and Alberts (2016), was conducted 5 min after the previous test. Each PND 7 pup was individually placed to the center of the testing cage (35 × 20 × 20 cm) with mesh flooring under which one side contained a cup with clean bedding and the other side contained a cup with home bedding (3 g each). The time spent exploring each cup and USVs' rates were measured for 2 min. Testing cage was cleaned with 50% ethanol, and bedding was changed between each pup testing. At the end of the test, pups were weighed, and their body temperature was noticed.

Juvenile Social Test

Juvenile social behavior and USVs' assessments, based on Panksepp et al. (2007) and Kabir et al. (2014), consisted in studying dyadic encounters between non-familiar sex and age-matched mice (PND 30–35). USVs in the social play test are considered as a relevant index of social motivation (Panksepp et al., 2007). The experimental mouse and social stimulus mouse were habituated to the experimental room for 1 h before the testing, during which each mouse was placed into an individual cage (28 × 18 × 12 cm), with clean bedding and without access to food and water, to improve social interest. The stimulus mice were all sighted and were previously habituated to encounter unfamiliar mice to attenuate potential stress effects. The experimental mouse was placed first into the testing cage (44.5 × 24 × 19.5 cm) for a 10–min habituation period. Once the social stimulus was introduced in the cage, social, non-social behaviors and USVs exhibited by the pairs of experimental and stimulus juvenile mice (blind-sighted and sighted-sighted) were recorded for 10 min.

The scored behaviors were adapted from Willey et al. (2009) and Cox and Rissman (2011). Social behaviors refer to (1) social sitting (sitting when being in close contact with the mouse stimulus), (2) social sniffing (investigating the mouse stimulus by sniffing the nose or the anogenital zone), and (3) following (walking behind the mouse stimulus). Non-social behaviors refer to (1) self-grooming, (2) exploring (investigating the cage alone), and (3) sitting alone. Mice were eventually weighed and put back in their home cage.

Female Urine-Induced USVs and Scent-Marking Behavior

This protocol, based on Roullet et al. (2011) and Binder et al. (2020), was conducted on adult male mice at 2–3 months. USVs are considered as an indicator of social motivation and sexual arousal in this social context (Wöhr and Scattoni, 2013). First, 1 week before the experimental testing, the male mouse was previously exposed to an unfamiliar female mouse, which is a crucial step to elicit subsequent male USVs' emission. Practically, both male and female mice were introduced for 5 min in a clean cage separated by a mesh to prevent copulation. Second, 24 h before testing, we placed soiled bedding of the male into the home cage of the female to induce the estrus. On the day of testing, the vaginal area was checked, and, if it was red, inflamed, and open, the female mouse was considered to be in estrus phase. A urine sample was thus performed by holding the female and gently stroking her abdomen toward the caudal direction. Fresh urine was collected in an Eppendorf tube and used for testing within 15 min.

After a 30-min habituation period in the experimental room, the male was then acclimated to the experimental cage (44.5 × 24 × 19.5 cm) for 30 min with a Strathmore paper (Strathmore Sketch Paper Pad, microperforated, 300 series, WI, USA) lining the floor and a small amount of its own bedding placed in a corner. Mice were then placed back into their own cage, while the experimental cage was cleaned (removal of the bedding) and urine markings were detected using an ultraviolet lamp and were circled with a marker. Later, 100 μl of urine from the

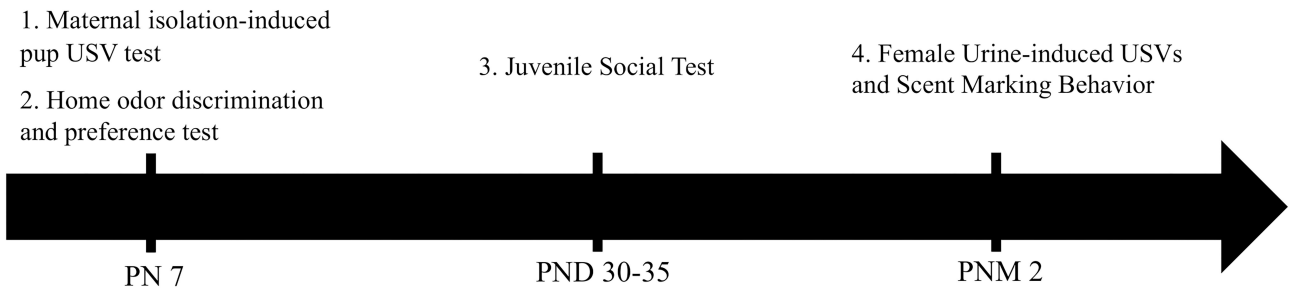


FIGURE 1 | Timeline of the experimental design.

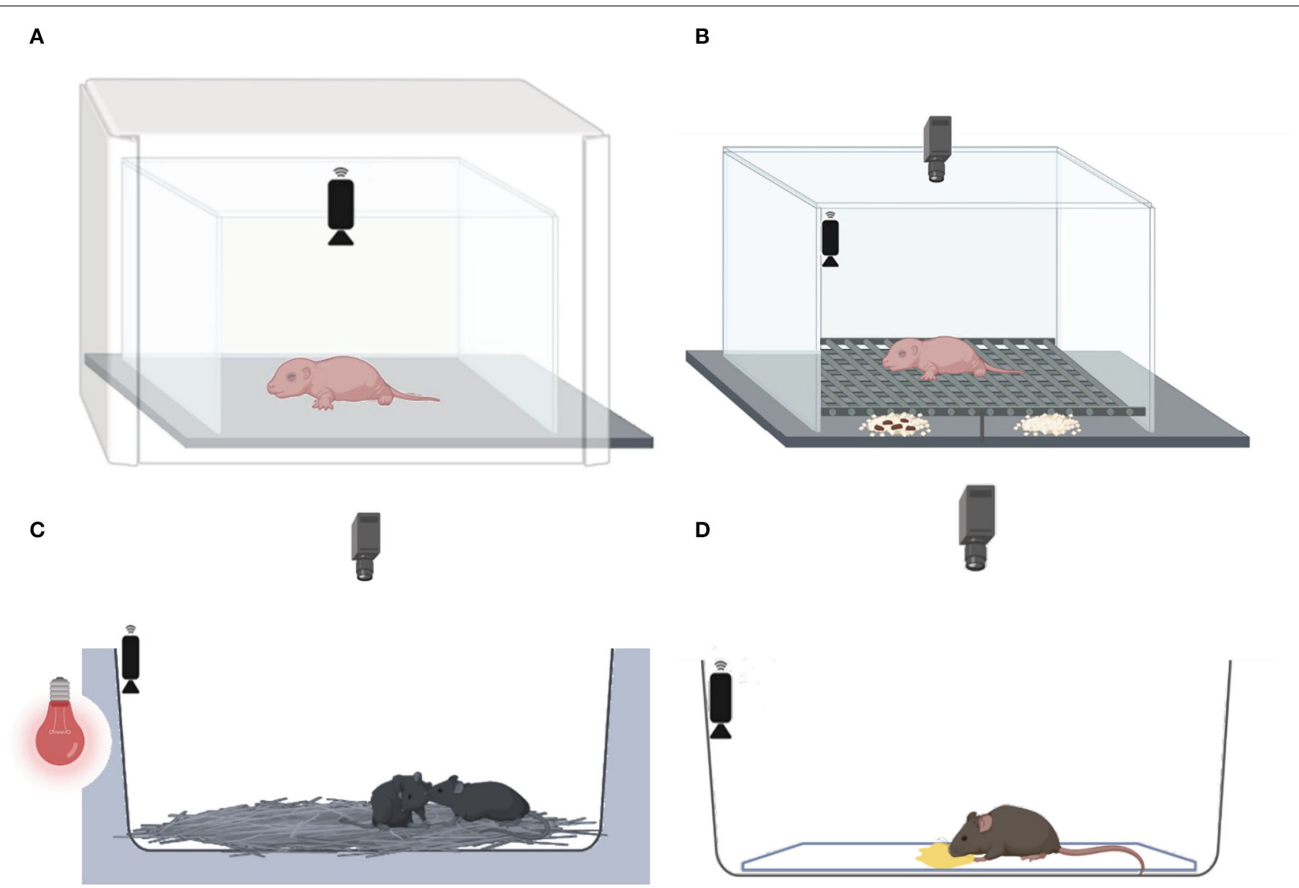


FIGURE 2 | Behavioral tests conducted at three developmental stages. **(A)** At PND 7, the maternal isolation-induced pup USV test, followed by **(B)** the home odor discrimination and preference test, **(C)** at PND 30–35, the juvenile social test, and **(D)** at 2–3 months, the female urine-induced USVs and scent-marking behavior test.

familiar female were placed on the Strathmore paper and the test started once the male mouse was placed in the experimental cage (Lehmann et al., 2013). The total numbers of USVs, total numbers of scent markings, time spent exploring the female urine, and total distance traveled were videotaped and analyzed for 5 min. Once the test ended, the Strathmore paper was removed, Ninhydrin was sprayed (DavTech Company, Canada), and scent markings were visible after 24 h. To quantify the scent marking, a transparent grid with 1 cm^2 squares was put on the Strathmore paper, and each square soiled by a urine marking was counted as one scent-mark unit. The total numbers of scent marks in the whole cage and within 10 cm^2 around the female urine spot were counted (Roullet et al., 2011).

Statistical Analysis

We used SPSS 22.0 (IBM, Armonk, NY, USA) for statistical analysis. We verified normal data distribution using the Shapiro-Wilk/Kolmogorov-Smirnov test. For data not normally distributed, we used the Mann-Whitney test to compare the experimental groups. All data are shown as mean ± SE.

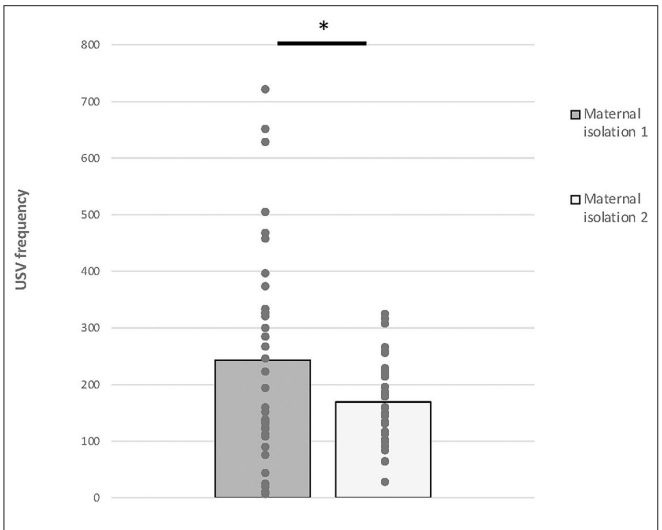


FIGURE 3 | Maternal isolation-induced pup USV test. Total numbers of USVs emitted by PND 7 mice (data from anophthalmic and sighted mice were compiled) during the 5-min isolation 1 and the 5-min isolation 2 (mean ± SE; $^*p < 0.05$).

Regarding the maternal isolation-induced pup USV test, we computed a repeated-measures ANOVA with time (two levels—isolation 1 and isolation 2) as within-subject factor and group (two levels—anophthalmic mice and sighted mice) as between-subject factor. Regarding the home odor discrimination and preference, we computed odor preference of each mouse as a difference index: time spent above home bedding odor—clean bedding odor (Meyer and Alberts, 2016). Consequently, we compared odor preference score and USV rates between the groups with Mann-Whitney *U*-tests for independent samples. Regarding the juvenile social and the scent-marking behavior tests, we compared both groups using a one-way MANOVA (independent groups—two visual groups; dependent variables—measures within a test). We examined correlations between variables for each group using Spearman's correlation. For all analyses, we set the significance level at $p < 0.05$, with appropriate corrections for multiple comparisons (Bonferroni's correction).

RESULTS

Statistical analysis revealed no significant effect of sex or of the interaction sex * vision on each dependent variable ($p > 0.05$); therefore, data from both sexes were combined.

Maternal Isolation-Induced Pup USV Test

Repeated-measures ANOVA analysis revealed a significant effect of *time* [$F_{(1,36)} = 6.885$, $p = 0.013$], but we did not find any effect of *visual status* [$F_{(1,36)} = 0.854$, $p = 0.362$] or the interaction *time* * *visual status* [$F_{(1,36)} = 2.973$; $p = 0.094$] (**Figure 3**).

Home Odor Discrimination and Preference Test

Given that data for odor preference index and USV rates did not pass the Shapiro-Wilk normality test ($W = 0.719$, $p < 0.001$ and $W = 0.852$, $p = 0.015$, **Figure 4**), we performed Mann-Whitney tests. There was no significant effect of odor preference index and USVs' rates between both groups ($U = 155$, p = 0.888; $U = 197.5$, $p = 0.236$). In addition, no correlation was found between the number of USVs and the odor preference index in both phenotypes [blind: $r(16) = -0.147$, sighted: -0.383, $p > 0.05$].

Juvenile Social Test

One-way MANOVA test revealed a significant effect of *visual status* on behavioral variables measured [$F_{(9,19)} = 2.925$, $p = 0.023$, **Figure 5**]. Specifically, blind mice exhibited less time spent in sniffing [$F_{(1,27)} = 15.635$, $p < 0.001$] and in following the congener [$F_{(1,27)} = 6.499$, $p = 0.017$] compared to sighted mice. Moreover, blind mice spent more time exploring the cage alone [$F_{(1,27)} = 3.977$, $p = 0.046$]. Overall, Student's *t*-tests confirmed that blind mice engaged less social behaviors [$t_{(1,27)} = 4.772$, $p < 0.001$] and consequently more frequent non-social behaviors [$t_{(1,27)} = -5.419$, $p < 0.001$]. Nevertheless, the total number of USVs emitted was similar between both groups [$F_{(1,27)} = 0.202$, $p = 0.657$].

In addition, we did not find any correlation between the total number of USVs and any social or non-social behaviors in both phenotypes [blind: r(16) = −0.335 to 0.253, sighted: −0.473 to 0.399, $p > 0.05$]. In blind mice, time spent exploring alone the environment was positively correlated with distance traveled [r(16) = 0.526, $p = 0.044$] and negatively associated with time spent sniffing the congener [r(16) = −0.601, $p = 0.018$], following the congener [r(16) = −0.532, $p = 0.041$], and sitting alone [r(16) = −0.561, $p = 0.030$]. In sighted mice, time spent following the congener was positively correlated with time spent sniffing this latter [r(16) = 0.572, $p = 0.033$] and negatively correlated with time spent sitting close to it [r(16) = −0.554, $p = 0.040$]. Besides, time spent exploring the cage in sighted mice was negatively associated with time spent self-grooming [r(16) = −0.688, $p = 0.007$].

Female Urine-Induced USVs and Scent-Marking Behavior

The one-way MANOVA revealed a significant effect of *visual status* on behavioral variables measured [$F_{(6,25)} = 2.560$, $p = 0.040$, **Figure 6**]. Specifically, blind mice were faster to sniff the urine spot and spent more time sniffing it [$F_{(1,30)} = 8.934$, $p = 0.006$, $F_{(1,30)} = 4.580$, $p = 0.040$, respectively]. Additionally, blind mice emitted USV faster than their sighted counterparts [$F_{(1,30)} = 4.679$, $p = 0.039$]. No difference was demonstrated in both groups in terms of total numbers of USVs, distance traveled, and total surface marked [$F_{(1,30)} = 0.715$; $F_{(1,30)} = 0.008$; $F_{(1,30)} = 0.404$, $p > 0.05$, respectively].

In addition, no correlation was found between the total number of USVs and any behavioral variables in blind mice [r(16) = −0.462 to 0.370, $p > 0.05$], whereas a positive correlation was found between the total number of USVs and (1) the total numbers of scent marks and (2) the latency before emitting USVs' calls in sighted mice [r(16) = 0.593, $p = 0.015$; r(16) = 0.535, $p = 0.033$, respectively]. Finally, the more sighted mice

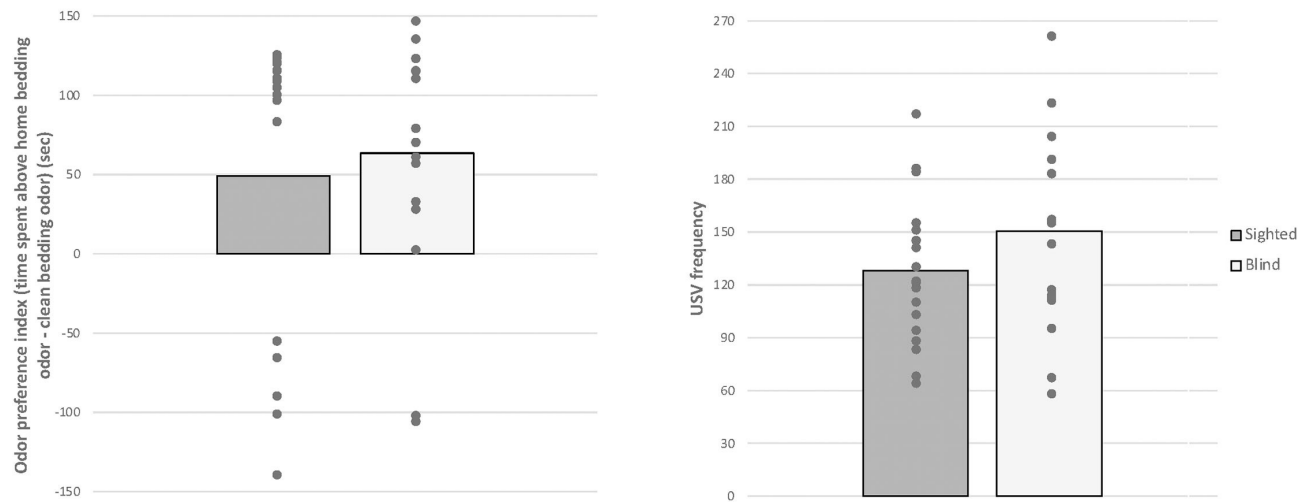


FIGURE 4 | Home odor discrimination and preference test. Anophthalmic and sighted mice assayed to the home odor discrimination and preference test: **(A)** Time spent (s) exploring the home bedding odor minus the clean bedding odor and **(B)** total numbers of USVs produced by PND 7 pups during the 2-min test (mean ± SE).

sniff the urine spot, the more they mark it [$r(16) = 0.620$, $p = 0.010$].

Correlations Between USVs' Rates

In blind mice, there was no correlation between the total number of USVs calls produced in the pup isolation, odor preference, social juvenile, and scent-marking tests. In sighted mice, a positive correlation was found between the total number of USVs emitted by PND 7 pups during the first 5-min isolation and during the scent-marking test in male adults [$r(16) = 0.599$, $p = 0.014$, **Figure 7**].

DISCUSSION

This study assessed acoustic communication and related behaviors in a mouse model of congenital blindness using a longitudinal approach. At PND 7, both sighted and blind mice exhibited similar social behavior and USVs' calls. In contrast, at PND 35, both groups differed significantly on several social behaviors, but no difference in USVs. Specifically, juvenile blind mice exhibited spent more time exploring the cage, and spent less time following and sniffing a congener than sighted ones. In male adults in response to a female urine stimuli, blind mice displayed shorter latency to vocalize and longer time spent sniffing the female urine spot, suggesting enhanced odor acuity and/or localization induced by early blindness. Finally, correlation analyses showed that in sighted mice the number of USVs emitted during the first maternal isolation at PND 7 was positively correlated with the number of USVs emitted during scent-marking test and a positive correlation was found, in the scent-marking test, between the number of USVs produced and the total number of scent markings. Conversely, in blind mice, no significant correlations were observed.

Early Developmental Stage (PND 7): USV Emission and Social Odors Discrimination and Preference

Isolation-induced USV has been widely employed as a marker for distress/anxiety in pups' rodents. USVs emitted by pup mice and rats are a spontaneous response to isolation from their mother (Barnes et al., 2017). Our results revealed no differences in the number of USVs emitted during the first and/or the second maternal isolation between sighted and blind pups. At PND 7, visual impairment had no incidence on isolation induced-vocalization behavior. Both groups having their eyes closed at this early developmental stage [the eyes opening in sighted mice occurs around PND 11–13 (*personal observations*)], we can assume that the visual status of both groups is similar. This indicates that the possible neuroanatomical differences between sighted and anophthalmic pups, such as a lack of optic nerve (Touj et al., 2019), have no impact on USVs' production behavior at PND 7. In this study, we also observed no difference between anophthalmic and sighted pups' calls emission during a second maternal isolation phase compared to the first phase, both groups emitting less calls in the second period. Hence, we did not observe the "maternal potentiation" phenomenon, namely, a higher acoustic response induced by a second isolation phase, which has been described previously in several studies in rats (Hofer et al., 1994; Shair, 2007) and in Swiss-Webster mice at PND 7–8 (Scattoni et al., 2008, 2009). Our results, however, are consistent with other studies, which showed no or decreasing number of USVs produced during the reisolation period in C57BL/6J mice (Barnes et al., 2017) and guinea pigs (Hennessy et al., 2006), respectively.

Besides USVs' emission, we also observed that both anophthalmic and sighted 7-day-old pups had similar olfactory

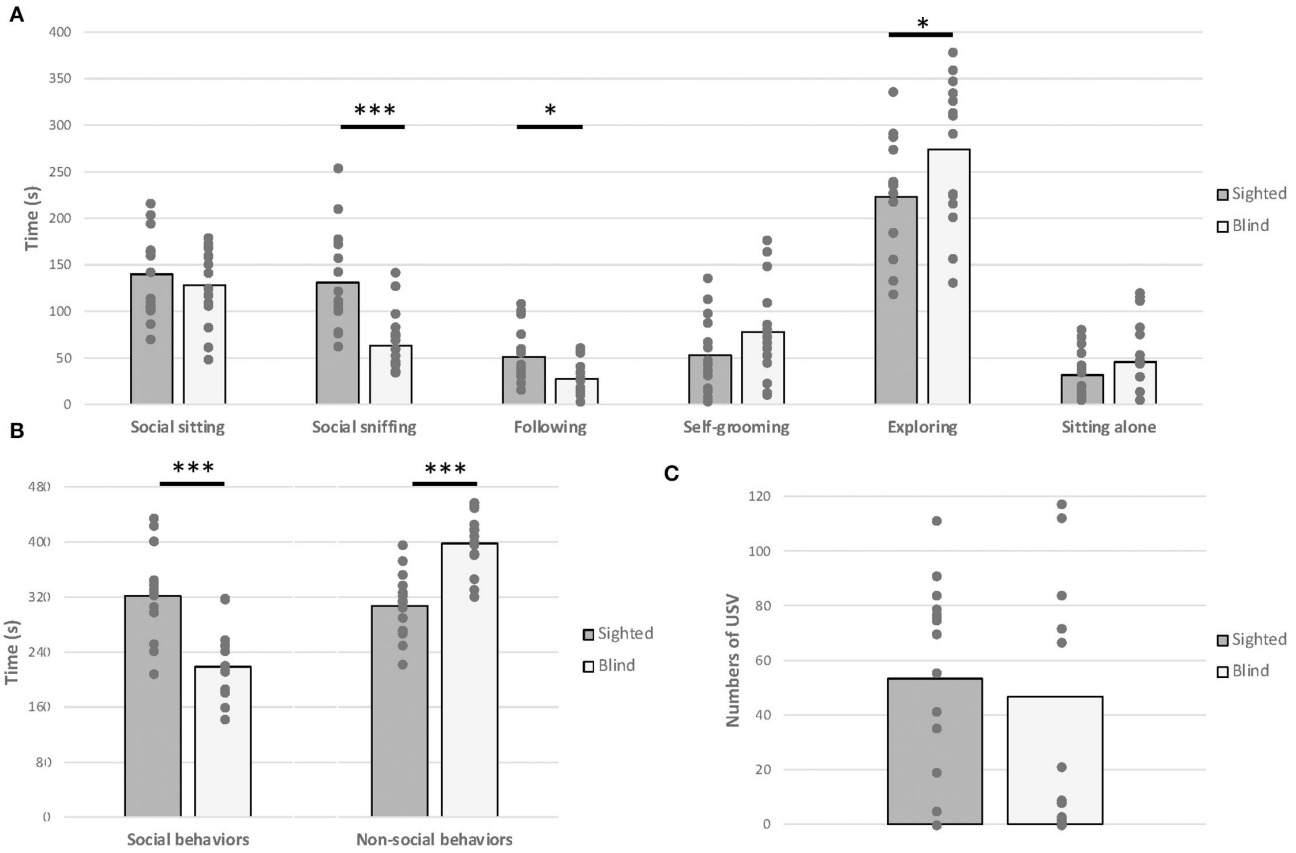


FIGURE 5 | Juvenile social test. Anophthalmic and sighted mice assayed to a 10-min juvenile social test at PND 30–35: **(A)** Mean (± SE) of the time spent (s) displaying social behaviors, such as social sitting, social sniffing, following the congener, and non-social behaviors, such as exploring, sitting alone, and self-grooming. **(B)** Mean (± SE) of the time spent (s) displaying social behaviors and non-social behaviors. **(C)** Mean (± SE) of total number of USV calls produced (*$p < 0.05$; ***$p < 0.001$).

behaviors. Indeed, both groups were able to discriminate and preferred exploring their home bedding's compared to the clean bedding's odors. These results are consistent with several studies showing that 9- to 13-day-old mice express their preference for their own nest to a clean one by searching and reaching the home nest/bedding area and spending more time on it (Scattoni et al., 2008; Lo Iacono et al., 2021; Premoli et al., 2021). It has also been suggested that, prior to the opening of the eyes pups' exploratory behavior is exclusively driven by olfactory cues (Freeman and Rosenblatt, 1978). Our findings indicate that both genotypes displayed the same odor performance in terms of odor discrimination and preference.

Taken together, our congenital blindness mouse model did not present alteration of neither the number of isolation-induced USVs emitted nor social olfactory behaviors at PND 7.

Juvenile Developmental Stage: PN30-35

During social play, the pairs of blind-sighted and sighted-sighted juvenile mice displayed a similar number of USVs. Interestingly, social behavior was altered in the blind mice, with a heightened time exploring the cage to the detriment of sniffing and following the congener. No differences between sighted and anophthalmic juvenile mice were observed otherwise in self-grooming, social sitting, and sitting-alone behaviors. Consistent with previous studies (Klein and Brown, 1969; Dyer and Weldon, 1975; Iura and Udo, 2014), we recently reported that congenitally blind mice exhibited an enhanced motivational level to explore a new environment, such as seen in the open-field, Elevated Plus Maze (EPM), and Forced Swim tests, compared to sighted controls (Bouguiyoud et al., 2022). The hyperactivity observed in blind mice when exploring a novel environment may reflect a compensatory mechanism by which the lack of visual information is counterbalanced by gathering and memorizing information from the physical world before it becomes familiar (Iura and Udo, 2014). Accordingly, blind animals may require a longer habituation period to explore the new environment. This could explain their increased general exploratory behavior despite the presence of a social stimulus mouse.

Furthermore, this apparent preference for exploring the environment over engaging in social interactions with a congener resembles the social deficits described in mouse models of autism (Moy et al., 2004; McFarlane et al., 2008; Silverman et al., 2010). Animal models of autism showed social behavior deficits such as less time sniffing, grooming, and following an unfamiliar

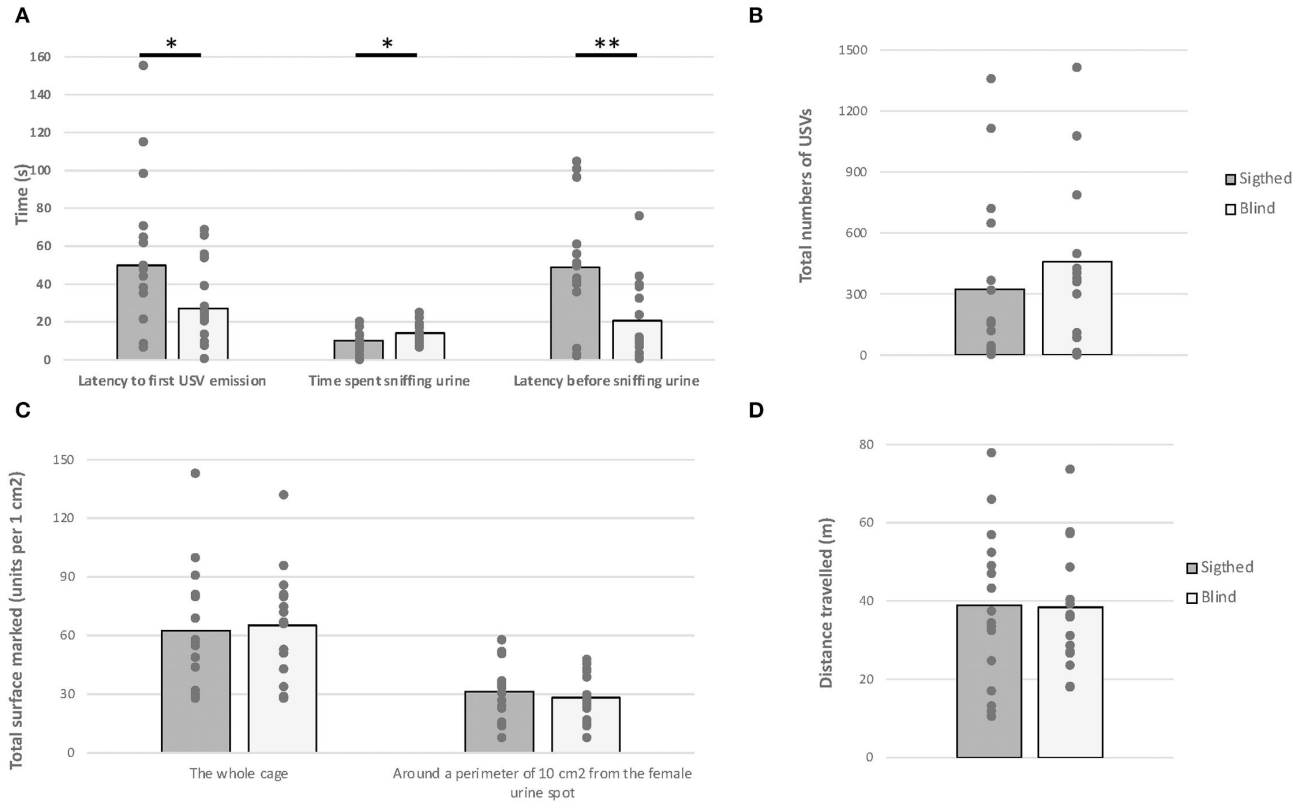


FIGURE 6 | Female urine-induced USVs and scent-marking behavior. Anophthalmic and sighted male mice were exposed to a 100–ml female urine spot for 5 min: **(A)** The latency before emitting the first USV call, the time spent sniffing the female urine, the latency before sniffing the female urine; **(B)** the total numbers of USVs emitted; **(C)** the total number of mark units; and **(D)** the total distance traveled (*$p < 0.05$; **$p < 0.01$).

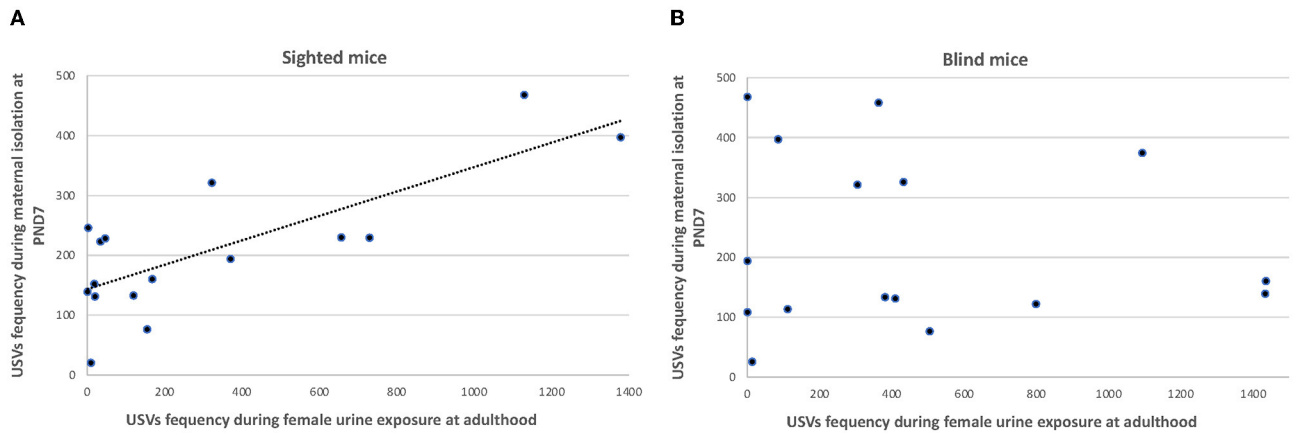


FIGURE 7 | Correlation graph of the total number of USVs emitted by PND 7 pups during the first 5-min isolation and during the scent-marking test in male adults **(A)** [sighted mice: $r(16) = 0.599$, $p = 0.014$], and **(B)** blind mice.

congener (Brodkin, 2007; McFarlane et al., 2008) and increased time allocated to self-grooming and inactivity (McFarlane et al., 2008).

Interestingly, in humans, a link between congenital blindness and autism was proposed based on the dramatic increased prevalence autism in individuals suffering visual impairments (Jure et al., 1991; Hobson et al., 1999; Hobson, 2011; Suhumaran et al., 2020). Moreover, children with profound visual impairment show delays in the development of joint attention behaviors, such as sharing or talking about their interests to others (Tadić et al., 2010; Dale et al., 2014). However, social behavioral differences in visually impaired individuals may be

caused by specific sensory limitations, as suggested by Chokron et al. (2020), and the link between early visual deprivation, social communication, and social behavior throughout development is still unclear in humans and non-human animal models.

Taken together, the similar amount of calls in the pairs of juveniles with a lower amount of social interaction in the blind-sighted mice indicates that either (1) blind-sighted pairs of mice produced higher amounts of USVs in a shorter total duration of social interactions than the sighted-sighted ones or (2) the calls were emitted during social and non-social exploratory behaviors. It should be noted that blind animals developed ultrasonic echolocation abilities to help them explore efficiently their surroundings (Griffin, 1944; Dufour et al., 2005; Schenkman and Nilsson, 2010; Thaler et al., 2011; Kupers and Ptito, 2014). Future studies should track each animal and record their acoustic behavior emitted specifically during social behaviors, such as allogrooming, sniffing, following, or specific non-social behaviors, such as exploration.

Adulthood: Female Urine-Induced USVs and Scent-Marking Behavior

Adult male mice exposed to females' urinary marks show a fast approach and prolonged sniffing behavior, emit ultrasonic vocalizations, a phenomenon known as the "ultrasonic courtship" vocalizations (Nyby, 1983; Holy and Guo, 2005; Arakawa et al., 2009; Wöhr and Schwarting, 2013), and deposit a large number of scent marks (Hurst, 1989; Arakawa et al., 2007), especially around the urinary source (Roullet et al., 2011). In this study, congenitally blind mice exhibited shorter latency to first sniff and higher duration sniffing the female urine spot, together with a shorter latency to emit the first female urine-induced USV call. The total number of calls, however, was similar between blind and sighted mice. Previous studies using the buried food test showed that blind rodents had a shorter latency to detect environmental odorants such as an appetent olfactory source than sighted congeners in congenitally blind ZRDBA mice (Touj et al., 2021) in visually deprived C57BL6 mice and rats (Zhou et al., 2017). To the best of our knowledge, our results highlight for the first time the importance of social olfactory cues (female urine) detection and their impact on acoustic behavior in anophthalmic male mice. In addition, histological and structural imaging studies reported larger olfactory bulbs and hypertrophy of brain areas involved in the olfactory processing (e.g., the anterior olfactory nucleus, the olfactory tubercle, the piriform cortex) in anophthalmic ZRDBA mice, supporting their heightened olfactory performance (Touj et al., 2020, 2021).

Accordingly, a longer time spent exploring the urinary source observed in blind mice may indicate a high attentional/motivational process toward odor stimuli. Similarly, increased attentional processes were found in our congenitally blinded mice in a two-odor choice test in response to positive odors (i.e., peanut butter and vanilla scent) compared to sighted mice (Touj et al., 2020). Thereby, visually impaired humans and non-human animals pay more attention to non-visual cues, processing them more efficiently compared to sighted individuals, which might enhance non-visual abilities (Kujala et al., 1997; Liotti et al., 1998; Hugdahl et al., 2004; Collignon et al., 2006; Collignon and De Volder, 2009; Ferdenzi et al., 2010; Beaulieu-Lefebvre et al., 2011; Pigeon and Marin-Lamellet, 2015; Topalidis et al., 2020) .

Regarding the scent-marking behavior, the total surface marked was not different between sighted and blind adult mice in the whole cage and at 10 cm^2 around the female urine drop. Dominant male mice tend to mark more than the subordinate ones (Desjardins et al., 1973; Rich and Hurst, 1998), so this result suggests that congenital blindness does not affect social ranking and reproductive-like behavior (both phenotypes are housed within the same cage). Further studies should examine in detail male social hierarchy in the context of sensory deficits such as visual deprivation. Finally, we reported a similar locomotor activity between the two groups in this test, which was also observed in Touj et al. (2020) when mice are assayed to an olfactory test.

Interestingly, in sighted, but not in the blind, adult male mice, there was a significant correlation between the number of female urine-induced calls and the total number of scent marks produced. This discrepancy may suggest an altered communication behavior in blind male mice in response to female urine, with a weaker coherence between the communication modalities (call emission and scent marking). This result is reminiscent of previous studies in a mouse model for autism, where the correlation between scent marking and USV emission appeared incoherent in the autism-like male group compared to the controls (Wöhr et al., 2011).

In conclusion, blind and sighted adult male mice exhibited differences in social communication and behavior in response to female urine stimuli.

Longitudinal Correlation of Acoustic Communication With Other Behaviors

The longitudinal analysis of the acoustic and related social behaviors across the three developmental stages examined in this study, namely, early, juvenile, and adult stages, show that in sighted mice, but not in the blind, the number of USVs emitted during the first maternal isolation is correlated with the total number of scent marks in the adult. The calls emitted by the mice at PND 7 during the isolation might indicate a high level of arousal, attention, and motivation in the pups, which may then be reflected in and predict the scent-marking behavior in the adult.

It should be noted that both behaviors are crucial for the social life in the mouse, and this correlation may indicate a common biological foundation and neurological pathway (Demir et al., 2020). In contrast, social play behavior in the juvenile has been described as a unique category of behavior and does not predict future adult social, sexual, or aggressive behavior (Vanderschuren et al., 1997).

CONCLUSION AND PERSPECTIVES

This study investigated acoustic communication and associated behaviors in a mouse model of congenital blindness using a longitudinal approach. Early blind mice compared to sighted

counterparts showed (1) no deficit in USVs' emission and social odor discrimination and preference at PND 7; (2) no difference in social play-induced USVs' emission but deficits in social behaviors (following and sniffing a congener) together with an increased exploratory behavior of the cage, without affecting the number of USV calls at PND 35; and (3) faster and longer exploration of female urine stimuli, faster emission to the first call, no differences in the number of calls and scent marks deposited in male adults. These findings indicate that congenital visual deprivation results in altered acoustic communication in the adult and modified social behaviors in juvenile and adult mice. Future studies should explore the qualitative analysis of the USVs emitted and examine the behavioral impact of playback calls, particularly within the mother-pup dyad, to understand with further details the foundation and development of acoustic communication of our mouse model of congenital blindness.

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 Animal Care Committee of the Université du Québec à Trois-Rivières, in accordance with the guidelines of the Canadian Council on Animal Care.

AUTHOR CONTRIBUTIONS

NB, EM-G, FIR, and SA: study conception and design. NB and SA: data collection. NB, JF, FIR, and SA: analysis and interpretation of results. NB and EM-G: draft manuscript preparation. All authors revised and approved the final version of this manuscript.

FUNDING

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) (2017-06942) and the MITACS Globalink Research Program.

REFERENCES

Arakawa, H., Arakawa, K., Blanchard, D. C., and Blanchard, R. J. (2007). Scent marking behavior in male C57BL/6J mice: sexual and developmental determination. *Behav. Brain Res.* 182, 73–79. doi: 10.1016/j.bbr.2007.05.007

Arakawa, H., Arakawa, K., Blanchard, D. C., and Blanchard, R. J. (2009). Social features of scent-donor mice modulate scent marking of C57BL/6J recipient males. *Behav. Brain Res.* 205, 138–145. doi: 10.1016/j.bbr.2009.07.007

Arriaga, G., Zhou, E. P., and Jarvis, E. D. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLoS ONE* 7, e46610. doi: 10.1371/journal.pone.0046610

Barnes, T. D., Rieger, M. A., Dougherty, J. D., and Holy, T. E. (2017). Group and individual variability in mouse pup isolation calls recorded on the same day show stability. *Front. Behav. Neurosci.* 11, 243. doi: 10.3389/fnbeh.2017.00243

Bean, N. J. (1982). Olfactory and vomeronasal mediation of ultrasonic vocalizations in male mice. *Physiol. Behav.* 28, 31–37. doi: 10.1016/0031-9384(82)90097-X

Beaulieu-Lefebvre, M., Schneider, F. C., Kupers, R., and Ptito, M. (2011). Odor perception and odor awareness in congenital blindness. *Brain Res. Bull.* 84, 206–209. doi: 10.1016/j.brainresbull.2010.12.014

Belagodu, A. P., Johnson, A. M., and Galvez, R. (2016). Characterization of ultrasonic vocalizations of Fragile X mice. *Behav. Brain Res.* 310, 76–83. doi: 10.1016/j.bbr.2016.04.016

Binder, M. S., Jones, D. G., Hodges, S. L., and Lugo, J. N. (2020). NS-Pten adult knockout mice display both quantitative and qualitative changes in urine-induced ultrasonic vocalizations. *Behav. Brain Res.* 378, 112189. doi: 10.1016/j.bbr.2019.112189

Bouguiyoud, N., Roullet, F., Bronchti, G., Frasnelli, J., and Al Aïn, S. (2022). Anxiety and depression assessments in a mouse model of congenital blindness. *Front. Neurosci.* 15, 807434. doi: 10.3389/fnins.2021.807434

Boulanger-Bertolus, J., Rincón-Cortés, M., Sullivan, R. M., and Mouly, A. M. (2017). Understanding pup affective state through ethologically significant ultrasonic vocalization frequency. *Sci. Rep.* 7, 13483. doi: 10.1038/s41598-017-13518-6

Bozdagi, O., Sakurai, T., Papapetrou, D., Wang, X., Dickstein, D. L., Takahashi, N., et al. (2010). Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism* 1, 15. doi: 10.1186/2040-2392-1-15

Branchi, I., Santucci, D., and Alleva, E. (2006). Analysis of ultrasonic vocalizations emitted by infant rodents. *Current Protoc. Toxicol.* doi: 10.1002/0471140856.tx1312s30

Branchi, I., Santucci, D., Vitale, A., and Alleva, E. (1998). Ultrasonic vocalizations by infant laboratory mice: a preliminary spectrographic characterization under different conditions. *Dev. Psychobiol.* 33, 249–256. doi: 10.1002/(SICI)1098-2302(199811)33:3<249::AID-DEV5>3.0.CO;2-R

Brodkin, E. S. (2007). BALB/c mice: low sociability and other phenotypes that may be relevant to autism. *Behav. Brain Res.* 176, 53–65. doi: 10.1016/j.bbr.2006.06.025

Brudzynski, S. M. (2013). Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. *Curr. Opin. Neurobiol.* 23, 310–317. doi: 10.1016/j.conb.2013.01.014

Chase, H. B., and Chase, E. B. (1941). Studies on an anophtalmic strain of mice. I, Embryology of the eye region. *J. Morph.* 68, 279–301. doi: 10.1002/jmor.1050680205

Chiesa, S., Galati, D., and Schmidt, S. (2015). Communicative interactions between visually impaired mothers and their sighted children: analysis of gaze, facial expressions, voice and physical contacts. *Child Care Health Dev.* 41, 1040–1046. doi: 10.1111/cch.12274

Chokron, S., Kovarski, K., Zalla, T., and Dutton, G. N. (2020). The inter-relationships between cerebral visual impairment, autism and intellectual disability. *Neurosci. Biobehav. Rev.* 114, 201–210. doi: 10.1016/j.neubiorev.2020.04.008

Collignon, O., and De Volder, A. G. (2009). Further evidence that congenitally blind participants react faster to auditory and tactile spatial targets. *Can. J. Exp. Psychol.* 63, 287–293. doi: 10.1037/a0015415

Collignon, O., Renier, L., Bruyer, R., Tranduy, D., and Veraart, C. (2006). Improved selective and divided spatial attention in early blind subjects. *Brain Res.* 1075, 175–82. doi: 10.1016/j.brainres.2005.12.079

Cox, K. H., and Rissman, E. F. (2011). Sex differences in juvenile mouse social behavior are influenced by sex chromosomes and social context. *Genes Brain Behav* 10, 465–472. doi: 10.1111/j.1601-183X.2011.00688.x

Dale, N. J., Tadi,ć V., and Sonksen, P. (2014). Social communicative variation in 1-3-year-olds with severe visual impairment. *Child Care Health Dev.* 40, 158–64. doi: 10.1111/cch.12065

Demir, E., Li, K., Bobrowski-Khoury, N., Sanders, J. I., Beynon, R. J., Hurst, J. L., et al. (2020). The pheromone darcin drives a circuit for innate and reinforced behaviours. *Nature* 578, 137–141. doi: 10.1038/s41586-020-1967-8

Desjardins, C., Maruniak, J. A., and Bronson, F. H. (1973). Social rank in house mice: differentiation revealed by ultraviolet visualization of urinary marking patterns. *Science* 182, 939–941. doi: 10.1126/science.182.4115.939

Dufour, A., Després, O., and Candas, V. (2005). Enhanced sensitivity to echo cues in blind subjects. *Exp. Brain Res.* 165, 515–519. doi: 10.1007/s00221-005-2329-3

Dyer, R. S., and Weldon, D. A. (1975). Blindness-induced hyperactivity in several strains of mice. *Physiol. Behav.* 15, 439–441. doi: 10.1016/0031-9384(75)90211-5

Ehret, G. (2005). Infant rodent ultrasounds – a gate to the understanding of sound communication. *Behav. Genetics* 35, 19–29. doi: 10.1007/s10519-004-0853-8

Ey, E., Torquet, N., le Sourd, A. M., Leblond, C. S., Boeckers, T. M., Faure, P., et al. (2013). The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. *Behav. Brain Res.* 256, 677–689. doi: 10.1016/j.bbr.2013.08.031

Ferdenzi, C., Schaal, B., and Roberts, S. C. (2010). Family scents: developmental changes in the perception of kin body odor? *J. Chem. Ecol.* 36, 847–854. doi: 10.1007/s10886-010-9827-x

Fischer, J., and Hammerschmidt, K. (2011). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes Brain Behav.* 10, 17–27. doi: 10.1111/j.1601-183X.2010.00610.x

Freeman, N. C., and Rosenblatt, J. S. (1978). The interrelationship between thermal and olfactory stimulation in the development of home orientation in newborn kittens. *Dev. Psychobiol.* 11, 437–457. doi: 10.1002/dev.420110508

Ganea, N., Hudry, K., Vernetti, A., Tucker, L., Charman, T., Johnson, M. H., et al. (2018). Supplemental material for development of adaptive communication skills in infants of blind parents. *Dev. Psychol.* 54, 2265–2273. doi: 10.1037/dev0000564

Gaub, S., Fisher, S. E., and Ehret, G. (2016). Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. *Genes Brain Behav.* 15, 243–259. doi: 10.1111/gbb.12274

Griffin, D. R. (1944). Echolocation by blind men, bats and radar. *Science (New York, N.Y.)* 100, 589–590. doi: 10.1126/science.100.2609.589

Grimsley, J. M. S., Monaghan, J. J. M., and Wenstrup, J. J. (2011). Development of social vocalizations in mice. *PLoS ONE* 6, e17460. doi: 10.1371/journal.pone.0017460

Grimsley, J. M. S., Sheth, S., Vallabh, N., Grimsley, C. A., Bhattal, J., Latsko, M., et al. (2016). Contextual modulation of vocal behavior in mouse: newly identified 12 kHz "Mid-frequency" vocalization emitted during restraint. *Front. Behav. Neurosci.* 10, 38. doi: 10.3389/fnbeh.2016.00038

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., and Fischer, J. (2009). Female mice respond to male ultrasonic "songs" with approach behaviour. *Biol. Letters* 5, 589–592. doi: 10.1098/rsbl.2009.0317

Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenzke, N., and Fischer, J. (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neurosci.* 13, 40. doi: 10.1186/1471-2202-13-40

Hanson, J. L., and Hurley, L. M. (2012). Female presence and estrous state influence mouse ultrasonic courtship vocalizations. *PLoS ONE* 7, e40782. doi: 10.1371/journal.pone.0040782

Hennessy, M. B., Miller, E. E., and Shair, H. N. (2006). Brief exposure to the biological mother "potentiates" the isolation behavior of precocial Guinea pig pups. *Dev. Psychobiol.* 48, 653–659. doi: 10.1002/dev.20180

Hobson, R. P. (2011). Congenital blindness and autism. *J. Autism Dev. Disord.* 41, 1593–1594. doi: 10.1007/s10803-011-1187-z

Hobson, R. P., Lee, A., and Brown, R. (1999). Autism and congenital blindness. *J. Autism Dev. Disord.* 29, 45–56. doi: 10.1023/A:1025918616111

Hofer, M. A., Brunelli, S. A., and Shair, H. N. (1994). Potentiation of isolation-induced vocalization by brief exposure of rat pups to maternal cues. *Dev. Psychobiol.* 27, 503–517. doi: 10.1002/dev.420270804

Holy, T. E., and Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biol.* 3, 1–10. doi: 10.1371/journal.pbio.0030386

Hugdahl, K., Ek, M., Takio, F., Rintee, T., Tuomainen, J., Haarala, C., et al. (2004). Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. *Brain Res. Cogn. Brain Res.* 19, 28–32. doi: 10.1016/j.cogbrainres.2003.10.015

Hurst, J. L. (1989). The complex network of olfactory communication in populations of wild house mice, Mus domesticus rutty: urine marking and investigation within family groups. *Anim. Behav.* 37, 705–725. doi: 10.1016/0003-3472(89)90057-2

Iura, Y., and Udo, H. (2014). Behavioral analyses of visually impaired Crx knockout mice revealed sensory compensation in exploratory activities on elevated platforms. *Behav. Brain Res.* 258, 1–7. doi: 10.1016/j.bbr.2013.10.020

Jamain, S., Radyushkin, K., Hammerschmidt, K., Granon, S., Boretius, S., Varoqueaux, F., et al. (2008). Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *PNAS* 105, 1710–1715. doi: 10.1073/pnas.0711555105

Jure, R., Rapin, I., and Tuchman, R. F. (1991). Hearing-impaired autistic children. *Dev. Med. Child Neurol.* 33, 1062–1072. doi: 10.1111/j.1469-8749.1991.tb14828.x

Kabir, Z. D., Kennedy, B., Katzman, A., Lahvis, G. P., and Kosofsky, B. E. (2014). Effects of prenatal cocaine exposure on social development in mice. *Dev. Neurosci.* 36, 338–346. doi: 10.1159/000360524

Klein, D., and Brown, T. S. (1969). Exploratory behavior and spontaneous alternation in blind and anosmic rats. *J. Comp. Physiol. Psychol.* 68, 107–110. doi: 10.1037/h0027657

Kujala, T., Lehtokoski, A., Alho, K., Kekoni, J., and Näätänen, R. (1997). Faster reaction times in the blind than sighted during bimodal divided attention. *Acta Psych.* 96, 75–82. doi: 10.1016/S0001-6918(97)00007-3

Kupers, R., and Ptito, M. (2014). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neurosci. Biobehav. Rev.* 41, 36–52. doi: 10.1016/j.neubiorev.2013.08.001

Langford, D. J., Crager, S. E., Shehzad, Z., Smith, S. B., Sotocinal, S. G., Levenstadt, J. S., et al. (2006). Social modulation of pain as evidence for empathy in mice. *Science* 312, 1967–1970. doi: 10.1126/science.1128322

Lebow, M. A., and Chen, A. (2016). Overshadowed by the amygdala: the bed nucleus of the stria terminalis emerges as key to psychiatric disorders. *Mol. Psych.* 21, 450–463. doi: 10.1038/mp.2016.1

Lehmann, M. L., Geddes, C. E., Lee, J. L., and Herkenham, M. (2013). Urine Scent Marking (USM): a novel test for depressive-like behavior and a predictor of stress resiliency in mice. *PLoS ONE* 8, e69822. doi: 10.1371/journal.pone.0069822

Liotti, M., Ryder, K., and Woldorff, M. G. (1998). Auditory attention in the congenitally blind: where, when and what gets reorganized? *Neuroreport* 9, 1007–1012. doi: 10.1097/00001756-199804200-00010

Lo Iacono, L., Mancini, C., Babicola, L., Pietrosanto, M., Di Segni, M., D'Addario, S. L., et al. (2021). Early life adversity affecting the attachment bond alters ventral tegmental area transcriptomic patterning and behavior almost exclusively in female mice. *Neurobiol. Stress* 15, 100406. doi: 10.1016/j.ynstr.2021.100406

Maggio, J. C., and Whitney, G. (1985). Ultrasonic vocalizing by adult female mice (*Mus musculus*). *J. Comp. Psych.* 99, 420–436. doi: 10.1037/0735-7036.99.4.420

Mahrt, E. J., Perkel, D. J., Tong, L., Rubel, E. W., and Portfors, C. V. (2013). Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J. Neurosci.* 33, 5573–5583. doi: 10.1523/JNEUROSCI.5054-12.2013

McFarlane, H. G., Kusek, G. K., Yang, M., Phoenix, J. L., Bolivar, V. J., and Crawley, J. N. (2008). Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav.* 7, 152–163. doi: 10.1111/j.1601-183X.2007.00330.x

Meyer, P. M., and Alberts, J. R. (2016). Non-nutritive, thermotactile cues induce odor preference in infant mice (*Mus musculus*). *J. Comp. Psychol.* 130, 369–379. doi: 10.1037/com0000044

Moy, S. S., Nadler, J. J., Perez, A., Barbaro, R. P., Johns, J. M., Magnuson, T. R., et al. (2004). Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav.* 3, 287–302. doi: 10.1111/j.1601-1848.2004.00076.x

Mun, H. S., Lipina, T., and v., Roder, J. C. (2015). Ultrasonic vocalizations in mice during exploratory behavior are context-dependent. *Front. Behav. Neurosci.* 9, 316 doi: 10.3389/fnbeh.2015.00316

Nyby, J. (1983). Ultrasonic vocalizations during sex behavior of male house mice (*Mus musculus*): a description. *Behav. Neural. Biol.* 39, 128–134. doi: 10.1016/S0163-1047(83)90722-7

Panksepp, J. B., Jochman, K. A., Kim, J. U., Koy, J. K., Wilson, E. D., Chen, Q., et al. (2007). Affiliative behavior, ultrasonic communication and social reward

are influenced by genetic variation in adolescent mice. *PLoS ONE* 2, e351 doi: 10.1371/journal.pone.0000351

Pigeon, C., and Marin-Lamellet, C. (2015). Evaluation of the attentional capacities and working memory of early and late blind persons. *Acta Psychol. (Amst)* 155, 1–7. doi: 10.1016/j.actpsy.2014.11.010

Pomerantz, S. M., Nunez, A. A., and Bean, N. J. (1983). Female behavior is affected by male ultrasonic vocalizations in house mice. *Physiol Behav*. 31, 91–6. doi: 10.1016/0031-9384(83)90101-4

Premoli, M., Memo, M., and Bonini, S. (2021). Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen. Res*. 16, 1158–1167. doi: 10.4103/1673-5374.300340

Rich, T. J., and Hurst, J. L. (1998). Scent marks as reliable signals of the competitive ability of mates. *Anim. Behav*. 56, 727–735. doi: 10.1006/anbe.1998.0803

Roullet, F. I., Wöhr, M., and Crawley, J. N. (2011). Female urine-induced male mice ultrasonic vocalizations, but not scent-marking, is modulated by social experience. *Behav. Brain Res*. 216, 19–28. doi: 10.1016/j.bbr.2010.06.004

Roullet, F. I., Wollaston, L., deCatanzaro, D., and Foster, J. A. (2010). Behavioral and molecular changes in the mouse in response to prenatal exposure to the anti-epileptic drug valproic acid. *Neuroscience* 170, 514–522. doi: 10.1016/j.neuroscience.2010.06.069

Sales, G. D. (1972). Ultrasound and aggressive behaviour in rats and other small mammals. *Anim. Behav*. 20, 88–100. doi: 10.1016/S0003-3472(72)80177-5

Sangiamo, D. T., Warren, M. R., and Neunuebel, J. P. (2020). Ultrasonic signals associated with different types of social behavior of mice. *Nat. Neurosci*. 23, 411–422. doi: 10.1038/s41593-020-0584-z

Scattoni, M. L., Crawley, J., and Ricceri, L. (2009). Ultrasonic vocalizations: a tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neurosci. Biobehav. Rev*. 33, 508–515. doi: 10.1016/j.neubiorev.2008.08.003

Scattoni, M. L., Gandhy, S. U., Ricceri, L., and Crawley, J. N. (2008). Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS ONE* 3, e3067. doi: 10.1371/journal.pone.0003067

Schenkman, B. N., and Nilsson, M. E. (2010). Human echolocation: blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object. *Perception* 39, 483–501. doi: 10.1068/p6473

Schmeisser, M.J., Ey, E., Wegener, S., Bockmann, J., Stempel, A.V., Kuebler, A., et al. (2012). Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature* 486, 256–260. doi: 10.1038/nature11015

Shair, H. N. (2007). Acquisition and expression of a socially mediated separation response. *Behav. Brain Res*. 182, 180–192. doi: 10.1016/j.bbr.2007.02.016

Silverman, J. L., Yang, M., Lord, C., and Crawley, J. N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nat. Rev. Neurosci*. 11, 490–502. doi: 10.1038/nrn2851

Suhumaran, S., Yeleswarapu, S. P., Daniel, L. M., and Wong, C. M. (2020). Congenital blindness and autism spectrum disorder (ASD): diagnostic challenges and intervention options. *BMJ Case Reports* 13, e232981. doi: 10.1136/bcr-2019-232981

Tadić, V., Pring, L., and Dale, N. (2010). Are language and social communication intact in children with congenital visual impairment at school age? *J. Child Psychol. Psychiatry* 51, 696–705. doi: 10.1111/j.1469-7610.2009.02200.x

Thaler, L., Arnott, S. R., and Goodale, M. A. (2011). Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE* 6, e20162. doi: 10.1371/journal.pone.0020162

Thoueille, É., Candilis-Huisman, D., Soul,é, M., and Vermillard, M. (2006). La maternité des femmes aveugles. Du désir d'enfant au bébé réel. *La psychiatrie de l'enfant*. 49, 285–348. doi: 10.3917/psye.492.0285

Topalidis, P., Zinchenko, A., Gädeke, J. C., and Föcker, J. (2020). The role of spatial selective attention in the processing of affective prosodies in congenitally blind adults: an ERP study. *Brain Res*. 1739, 146819. doi: 10.1016/j.brainres.2020.146819

Touj, S., Cloutier, S., Jemâa, A., Piché, M., Bronchti, G., and Al Aïn, S. (2020). Better Olfactory performance and larger olfactory bulbs in a mouse model of congenital blindness. *Chem. Senses* 45, 523–531. doi: 10.1093/chemse/bjaa052

Touj, S., Gallino, D., Chakravarty, M. M., Bronchti, G., and Piché, M. (2021). Structural brain plasticity induced by early blindness. *Eur. J. Neurosci*. 53, 778–795. doi: 10.1111/ejn.15028

Touj, S., Tokunaga, R., Al Aïn, S., Bronchti, G., and Piché, M. (2019). Pain hypersensitivity is associated with increased amygdala volume and c-fos immunoreactivity in anophthalmic mice. *Neuroscience* 418, 37–49. doi: 10.1016/j.neuroscience.2019.08.035

Vanderschuren, L. J., Niesink, R. J., and Van Ree, J. M. (1997). The neurobiology of social play behavior in rats. *Neurosci. Biobehav. Rev*. 21, 309–326. doi: 10.1016/S0149-7634(96)00020-6

von Merten, S., Hoier, S., Pfeifle, C., and Tautz, D. (2014). A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (*Mus musculus domesticus*). *PLoS ONE* 9, e97244. doi: 10.1371/journal.pone.0097244

Wang, H., Liang, S., Burgdorf, J., Wess, J., and Yeomans, J. (2008). Ultrasonic vocalizations induced by sex and amphetamine in M2, M4, M5 muscarinic and D2 dopamine receptor knockout mice. *PLoS ONE* 3, e1893. doi: 10.1371/journal.pone.0001893

Willey, A. R., Varlinskaya, E. I., and Spear, L. P. (2009). Social interactions and 50 kHz ultrasonic vocalizations in adolescent and adult rats. *Behav. Brain Res*. 202, 122–129. doi: 10.1016/j.bbr.2009.03.025

Williams, W. O., Riskin, D. K., and Mott, K. M. (2008). Ultrasonic sound as an indicator of acute pain in laboratory mice. *JAALAS* 47, 8–10.

Wöhr, M. (2014). Ultrasonic vocalizations in Shank mouse models for autism spectrum disorders: detailed spectrographic analyses and developmental profiles. *Neurosci. Biobehav. Rev*. 43, 199–212. doi: 10.1016/j.neubiorev.2014.03.021

Wöhr, M., Roullet, F. I., and Crawley, J. N. (2011). Reduced scent marking and ultrasonic vocalizations in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav*. 10, 35–43. doi: 10.1111/j.1601-183X.2010.00582.x

Wöhr, M., and Scattoni, M. L. (2013). Behavioral methods used in rodent models of autism spectrum disorders: current standards and new developments. *Behav. Brain Res*. 251, 5–17. doi: 10.1016/j.bbr.2013.05.047

Wöhr, M., and Schwarting, R. K. W. (2013). Affective communication in rodents: ultrasonic vocalizations as a tool for research on emotion and motivation. *Cell Tissue Res*. 354, 81–97. doi: 10.1007/s00441-013-1607-9

Zhou, Y., Fang, F. H., Pan, P., Liu, Z. R., and Ji, Y. H. (2017). Visual deprivation induce cross-modal enhancement of olfactory perception. *Biochem. Biophys. Res. Commun*. 486, 833–838. doi: 10.1016/j.bbrc.2017.03.140

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Behavioral Neuroscience

ORIGINAL RESEARCH
published: 14 July 2022
doi: 10.3389/fnbeh.2022.883353

Ultrasonic Vocalizations in Adult C57BL/6J Mice: The Role of Sex Differences and Repeated Testing

Marika Premoli[1], *Valeria Petroni*[2], *Ronald Bulthuis*[3], *Sara Anna Bonini*[1] *and Susanna Pietropaolo*[2]*

[1] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy, [2] Univ. Bordeaux, CNRS, INCIA, UMR 5287, Bordeaux, France, [3] Metris B.V., Hoofddorp, Netherlands

Ultrasonic vocalizations (USVs) are a major tool for assessing social communication in laboratory mice during their entire lifespan. At adulthood, male mice preferentially emit USVs toward a female conspecific, while females mostly produce ultrasonic calls when facing an adult intruder of the same sex. Recent studies have developed several sophisticated tools to analyze adult mouse USVs, especially in males, because of the increasing relevance of adult communication for behavioral phenotyping of mouse models of autism spectrum disorder (ASD). Little attention has been instead devoted to adult female USVs and impact of sex differences on the quantitative and qualitative characteristics of mouse USVs. Most of the studies have also focused on a single testing session, often without concomitant assessment of other social behaviors (e.g., sniffing), so little is still known about the link between USVs and other aspects of social interaction and their stability/variations across multiple encounters. Here, we evaluated the USVs emitted by adult male and female mice during 3 repeated encounters with an unfamiliar female, with equal or different pre-testing isolation periods between sexes. We demonstrated clear sex differences in several USVs' characteristics and other social behaviors, and these were mostly stable across the encounters and independent of pre-testing isolation. The estrous cycle of the tested females exerted quantitative effects on their vocal and non-vocal behaviors, although it did not affect the qualitative composition of ultrasonic calls. Our findings obtained in B6 mice, i.e., the strain most widely used for engineering of transgenic mouse lines, contribute to provide new guidelines for assessing ultrasonic communication in male and female adult mice.

Keywords: ultrasonic communication, social behaviors, sex, repeated testing, isolation

OPEN ACCESS

Edited by:
Valerie J. Bolivar,
Wadsworth Center, United States

Reviewed by:
Erich David Jarvis,
Duke University, United States
Claudia Pitzer,
University of Heidelberg, Germany

***Correspondence:**
Susanna Pietropaolo
susanna.pietropaolo@u-bordeaux.fr

Specialty section:
This article was submitted to Individual and Social Behaviors, a section of the journal Frontiers in Behavioral Neuroscience

Received: *25 February 2022*
Accepted: *20 June 2022*
Published: *14 July 2022*

Citation:
Premoli M, Petroni V, Bulthuis R, Bonini SA and Pietropaolo S (2022) Ultrasonic Vocalizations in Adult C57BL/6J Mice: The Role of Sex Differences and Repeated Testing. Front. Behav. Neurosci. 16:883353. doi: 10.3389/fnbeh.2022.883353

INTRODUCTION

Mice emit ultrasonic vocalizations (USVs) to communicate with each other in a social context during defined phases of their life: in newborn (i.e., during the first 15 post-natal days) to summon the mother, in the juvenile phase (i.e., between 3 and 7 weeks of age) during playing in same-sex dyads and at adulthood most commonly during male-female or female-female interactions (Lahvis et al., 2011; Arriaga and Jarvis, 2013; Caruso et al., 2020). USVs have a frequency range between 30 and 110 kHz and are of innate nature, since it has been demonstrated that mice are not vocal learners (Kikusui et al., 2011; Mahrt et al., 2013). At all ages, different types of vocalizations exist

with specific spectrographic characteristics that several researchers have struggled to classify using different technologies and approaches (Holy and Guo, 2005; Gaub et al., 2016; Grimsley et al., 2016; Premoli et al., 2021a). However, the understanding of the precise significance of different types of calls is still unknown to date, and additional data are needed to better unravel this issue, allowing an essential step forward in the field of mouse behavioral neuroscience.

Studies on USVs have become a widely used behavioral assay to monitor the emotional state of mice (Simola and Granon, 2019) and their sociability. The research interest in USVs is indeed justified from both ethological and preclinical points of view: a growing number of studies have applied USVs as a valuable tool to study pathologies characterized by deficits in communication and sociability in mouse models (such as autism spectrum disorder, ASD) and to investigate the effects of therapeutic approaches, since USVs can be modulated by diverse pharmacological treatments (Premoli et al., 2021b). A large part of studies on adult ultrasonic communication have focused on USVs emitted by male mice, since these are more commonly employed than those by females for behavioral phenotyping of animal models of ASD and avoid the well-known impact of the estrous cycle on females' USVs (Moles et al., 2007). USVs produced by adult mice during male-female dyadic interactions are also the most extensively characterized, and they allow for easier identification of the emitting animal, since it has been demonstrated that in this context USVs are mostly produced by male mice to attract female mice (Sugimoto et al., 2011; Hammerschmidt et al., 2012; Egnor and Seagraves, 2016).

Nonetheless, USVs can also be emitted by adult females: recent studies have described that adult receptive female mice produced USVs either in the presence of male urines enriched with pheromones or in groups of four mixed-sex individuals (Neunuebel et al., 2015; Demir et al., 2020). However, in these studies, the female mice were assessed under experimental conditions maximizing the expression of their sexual interest, including being tested in a receptive estrous state, during the dark phase, and for long sessions (more than 30 min). Even under these conditions, the largest proportion of USVs registered during group interactions was produced by males (Neunuebel et al., 2015). Instead, female mice show their most prominent vocalizing abilities during adult female-female dyadic interactions, and in these cases USVs are used to establish affiliative relationships (Moles and D'Amato, 2000; Moles et al., 2007; Zala et al., 2017). In particular, Moles and D'Amato (2000), Moles et al. (2007) studied USVs of adult female mice in a resident-intruder setting, i.e., the resident being isolated in the testing cage 3 days before assessing the USVs with a female intruder. They demonstrated that in these experimental settings most of the USVs are uttered by the resident and suggested that the calls can facilitate proximity with the intruder and reduce its potential aggressiveness. Female USVs in this context can be also used as an index of sociability and social memory, since (i) a strong positive correlation was found between the number of calls and the time spent by the resident female mouse sniffing the intruder, and (ii) a marked decline was observed in the number of USVs emitted by a resident female mouse when exposed multiple consecutive times to the same female intruder. USVs are mostly emitted during close contacts and approach behaviors in female-female interactions and male-female encounters (Ferhat et al., 2015, 2016). Nonetheless, the precise link between multiple USV characteristics and other social behaviors is still not fully understood in adult female mice, as in male mice.

Several studies have tried to analyze differences in USVs between male and female mice (Hammerschmidt et al., 2012; von Merten et al., 2014; Zala et al., 2017; Matsumoto and Okanoya, 2018; de Chaumont et al., 2021), also with novel technical approaches such as deep learning networks (Ivanenko et al., 2020), yielding to the emergence of a variety of either quantitative or qualitative differences (or both) without, so far, a univocal pattern. Divergences among these studies mainly arise from differences in testing procedures, e.g., sex of the stimulus animal and pre-testing isolation conditions. The sex of the "receiver" is known to critically modulate several characteristics of USVs of the "emitter" in mice of both sexes (Zala et al., 2017); for instance, quantitative sex differences in USVs were described between female-female and male-female interactions (von Merten et al., 2014), while both quantitative and qualitative differences were observed in same-sex interactions (de Chaumont et al., 2021). Also, most of the studies on female USVs have applied relatively long periods (more than 24 h) of pre-testing isolation in order to induce a resident status in the subject and assure the identification of the emitter (Moles et al., 2007; Hammerschmidt et al., 2012). In contrast, pre-testing isolation is not commonly applied in USV studies on adult male mice, as this manipulation is not necessary to induce their USV production. In general, isolation is known to alter USV emission in adult mice (Lefebvre et al., 2020; Zhao et al., 2021) and to modulate the correlation between USVs and other social behaviors (Chabout et al., 2012). Finally, most of the studies on sex differences in USVs have employed single testing sessions or multiple repeated sessions but with the same social stimulus in order to assess habituation (Moles et al., 2007). Hence, it is not clear whether sex differences in ultrasonic communication may be dependent on testing experience or are a stable trait across multiple social encounters with an unfamiliar stimulus.

Here, we performed an extensive quantitative and qualitative characterization of sex differences in USVs emitted by adult C57BL/6J (B6) mice. To this end, we compared the USVs uttered either by an adult male or female toward the same type of stimulus, i.e., an adult CD1 female. The female mice were isolated for 3 days before testing in order to acquire the status of resident, i.e., becoming the major emitter of USVs during interaction with a female intruder. Their USVs were compared with those of males that were either isolated for the same duration before testing (study 1) or only habituated to the testing cage for 10 min before tests (study 2). These two studies allow us to evaluate sex differences respectively (1) in the same resident-intruder settings, thus controlling for isolation effects and (2) using the most common (and practically more suitable) experimental settings for USV assessment used in previous studies with ASD mouse models, i.e., in the resident-intruder context for females and without pre-testing long isolation in males (Pietropaolo et al., 2011, 2014; Hebert et al., 2014; Oddi et al., 2015; Gaudissard

et al., 2017; Gauducheau et al., 2017; Fyke et al., 2021). In both studies, three subsequent social tests with a novel intruder were performed with an interval of 7–10 days in order to evaluate the potential stability of sex differences and their dependency on previous testing experience without confounding effects of social memory. For each encounter, social affiliative behaviors were also assessed in order to evaluate their potential link with USV changes. Estrous cycle phases were assessed for experimental female subjects and female stimuli before each social encounter to control for potential hormonal modulation of social interaction and communication (Moles et al., 2007; Hanson and Hurley, 2012; Egnor and Seagraves, 2016; Kim et al., 2016).

We chose B6 mice as the experimental subjects of our study because of the well-known relevance of this strain for behavioral neuroscience due to its large use for engineering genetically modified mouse lines. Stimulus females for all social tests were instead chosen from the CD1 strain because of its common use in social studies (Moles and D'Amato, 2000; Moles et al., 2007), especially those using genetic mouse models of ASD (Hebert et al., 2014; Pietropaolo et al., 2014; Oddi et al., 2015; Gaudissard et al., 2017; Gauducheau et al., 2017; Lemaire-Mayo et al., 2017; Fyke et al., 2021). This strain is preferentially employed in social interaction tests, since it is characterized by high levels of sociability and it facilitates behavioral analysis during social encounters with B6 animals because of its albino phenotype.

MATERIALS AND METHODS

Animals and Housing Conditions

Forty adult male and female C57BL/6J mice (10–12 weeks old, $n = 10$ per sex in each experiment), used as experimental subjects, and forty adult CD1 females, used as social stimuli, were purchased from Janvier (Le Genest St Isle, France). Upon arrival at our animal facility at Bordeaux University they were all housed in same-sex and same–strain groups of 5 individuals in standard polycarbonate cages (37 × 21 × 15 cm in size; Tecniplast, Limonest, France) and provided with sawdust bedding (SAFE, Augy, France) enriched with cotton nestlets. Food chow (SAFE, Augy, France) and water were provided *ad libitum*. The animals were maintained in a temperature- (22°C) and humidity- (55%) controlled vivarium under a 12:12 h light–dark cycle (lights on at 7 a.m.). The mice were left undisturbed for 2 weeks upon their arrival before the behavioral tests began.

As illustrated in **Supplementary Figure 1**, two separate groups of mice were used for the two experiments of the study, each consisting of 20 B6 experimental subjects (10 male mice and 10 female mice) and 20 CD1 stimulus females. The CD1 female mice were all naïve to social experience with B6 mice at the time of the first testing session; each female stimulus was employed for a total of 3 times for each experiment but only once for each testing session. B6 mice of either sex encountered a novel female at each session. Separate batches of CD1 female mice were employed for male-female and female-female interactions in each experiment, so that a CD1 stimulus encountered either B6 female or male mice during the 3 sessions. In experiment 1, both male and female B6 subjects were single-caged for the same time (72 h) in the test cage before each testing session to assess social behaviors and USVs: this allowed for us to assess sex differences in the same resident-intruder settings and pre-testing social isolation conditions. In experiment 2, the male mice were subjected to 10 min of isolation in the test cage, and their social behavior was compared with that of female mice exposed to 72-h pretesting isolation: this comparison served to evaluate sex differences under conditions that are commonly employed to assess male and female USVs in ASD mouse studies (Hebert et al., 2014; Pietropaolo et al., 2014; Oddi et al., 2015; Gaudissard et al., 2017; Gauducheau et al., 2017; Lemaire-Mayo et al., 2017; Fyke et al., 2021) and that are also more suitable for male behavioral assessment. USVs can in fact be also evaluated in male-female interactions without inducing a resident state in the male [e.g., Hebert et al., 2014; Oddi et al., 2015; Gaudissard et al., 2017; Lemaire-Mayo et al., 2017; Fyke et al., 2021], thus avoiding applying a social isolation period of at least 72 h that could interfere with several other behaviors. In contrast, female mouse USVs are most commonly assessed in a resident-intruder setting, at least in female-female dyadic interactions. In both experiments, after each testing session, the experimental mice were re-housed in groups with the same cagemates. The CD1 stimulus mice were kept under grouped conditions during the entire duration of the study.

Behavioral Procedures

Behavioral testing was carried out during the light phase of the cycle. All the experimental procedures were performed in accordance with the European Communities Council Directive 2010/63/EEC and approved by the Local Ethical Committee ("Comité d'Ethique pour l'experimentation animale de Bordeaux", CE 50) and the French Ministry ("Ministere de l'enseignement superieur de la recherché et de l'innovation").

Social behavior and ultrasonic communication were assessed in a 33 × 15 × 14 cm plastic cage with 3 cm of sawdust and a metal flat cover during 3 testing sessions of 3 min each and with an interval of 7–10 days. In experiments 1 and 2, the female B6 subjects were isolated in the testing cage for 72 h prior to testing in order to induce the status of resident in the adult female mice and therefore promote the emission of ultrasonic vocalizations (USVs) toward an adult female intruder (Moles et al., 2007). The male B6 subjects were isolated either for 72 h (experiment 1) or for 10 min (experiment 2) in the testing cage before each social encounter. In all the experiments, an unfamiliar adult female CD1 stimulus was then introduced into the testing cage of either male or female subjects and left there for 3 min. Previous studies alternately anesthetizing each pair member have shown that in these experimental settings adult stimulus females do not emit ultrasonic vocalizations (USVs) that are instead mostly uttered by the experimental male (Whitney et al., 1973; Maggio and Whitney, 1985) or the resident female (Maggio and Whitney, 1985; D'Amato and Moles, 2001). The lack of concomitant emission of USVs by the two interacting animals was indeed confirmed here by additional inspection of spectrograms, excluding the presence of "double calls", i.e., overlapping in their timing, but with different, non-harmonic characteristics (e.g., different peak and mean frequency, modulation). After each testing session of both

experiments, the experimental and stimulus mice were returned to their home cages and kept with their original cagemates until the subsequent testing session.

The testing sessions were recorded with a camera placed on the side of the cage, and videos were analyzed with Observer XT (Noldus, The Netherlands). One observer who was unaware of the sex and experimental assignment of the animals scored the behavior of the test B6 mice only, quantifying the time spent performing the following behaviors (Pietropaolo et al., 2011, 2014; Oddi et al., 2015; Gaudissard et al., 2017; Gauducheau et al., 2017):

- affiliative behaviors: nose/anogenital/body sniffing (sniffing the head and snout of the partner/its anogenital region/any other part of the body), contact with the partner (traversing the partner's body by crawling over/under from one side to the other or allogrooming)
- nonsocial activities: rearing (standing on the hind limbs and sometimes with the forelimbs against the walls of the cage), exploring the cage (locomotion and wall rearing), digging, grid-climbing, self-grooming (the animal licks and mouths its own fur).

An ultrasonic microphone, UltraSoundGate Condenser Microphone CM 16 (Avisoft Bioacoustics, Berlin, Germany), was mounted 2 cm above the cover of the testing cage; it was connected *via* an UltraSoundGate 116 USB audio device (Avisoft Bioacoustics) to a personal computer, with which acoustic data were recorded with a sampling rate of 250 kHz in 16-bit format with Avisoft Recorder (version 2.97; Avisoft Bioacoustics). The recordings were then transferred to Avisoft SASLab Pro (Version 5.20; Avisoft, Berlin, Germany) and Fast Fourier transformation was applied (512 FFT length, 100% frame, Hamming window, and 75% time window overlap).

Spectrograms were generated with Avisoft at a frequency resolution of 488 Hz and a time resolution of 0.512 ms. Signals below 30 kHz were cut to reduce background noise to 0 dB (Premoli et al., 2019). For USV detection, an interactive function with section labels was used. This tool permits to define manually USV borders by inserting section labels, and it is useful when automatic threshold-based USV separation may not work satisfactorily because of ambient noise or because of poorly structured vocalizations (manual guide of Avisoft Bioacoustics). Number, mean duration, peak frequency, and peak amplitude were calculated for each vocalization together with the calling time of the mice based on previous studies (Wohr et al., 2011). Call subtypes were also determined for a more detailed qualitative analysis. For this purpose, USVs were automatically classified with the Sonotrack Call Classification (version 1.4.7, Metris B.V., The Netherlands) software, using the categories described in detail in **Figure 1**, based on previous literature on mouse USVs (Caruso et al., 2020). To deal with background noise and artifacts in the ultrasound recordings, the Sonotrack Call Classification software applies various filters to remove unwanted signals such as white noise and artificial sound sources. In addition, a logic filter is used that further processes the recorded signal by removing sounds that are too short or appear at many frequencies at the same time. The logic filter also reduces echo that is found in the recording and merges spectral elements that are interrupted by a very short time and a frequency gap.

On each testing day, the vaginal estrous phases of both testing and stimulus female mice were assessed from the analysis of their vaginal smears (Caligioni, 2009). In both experiments, all stimulus females used for male-female interactions were in non-receptive diestrous phase in order to minimize mounting attempts and other sexual behaviors that could confound the evaluation of sex differences in USVs and social behaviors. The stimulus females for female-female interactions were either in diestrous (non-receptive) or estrous (receptive) phase, and their assignment was counterbalanced depending on the estrous phase of the experimental subjects (i.e., approximately half of the intruders in the estrous phase encountered a resident in estrous and the other half was assigned to a resident in diestrous; the same design was applied to the intruders in diestrous). The estrous phases of the female residents and intruders for each testing session of female-female interactions are illustrated in **Table 1**.

Statistical Analyses

Normality of data distribution was confirmed by Shapiro-Wilks test for each sex and testing session and for each variable of interest. Behavioral data from each experiment were separately analyzed by ANOVA with sex as the between-subject factor and testing session as the within-subject factor. Furthermore, behavioral data from the female mice in each testing session were subjected to an additional ANOVA with the estrous phase (estrous or diestrous) of the experimental B6 female and the estrous phase (e.g., **Supplementary Figures 2, 3**) of the stimulus CD1 female as the between subject-factors. The analysis of the data from male mice did not include the estrous phase of the stimulus, since all CD1 females selected for testing males were in diestrous phase. A further ANOVA with experiment as the additional between-subject factor was performed on the data from female mice only in order to quantify the replicability of the female phenotype (under the same testing conditions) across the two experiments.

Post-hoc comparisons (Fisher's LSD test) and separate ANOVAs were performed when appropriate. All the analyses were conducted using the software Statview and SPSS, and α was set at 0.05. The data were inspected for exclusion of outliers (by Grubbs' ESD test adapted for small sample size). Outlier values were excluded only from a specific dataset (e.g., body sniffing time on session 1), except for the analysis of repeated measures, when values for all the 3 sessions had to be excluded for the affected variable. The results are expressed as mean $\pm$ SEM throughout the text. Individual data of social behaviors and USV parameters are also provided for all the animals in **Supplementary Figures 4, 5**; in addition, the individual composition of call types is illustrated in **Supplementary Figures 6, 7** for half (i.e., 5 over 10) of individuals for each sex emitting the higher rates of USVs in each experiment.

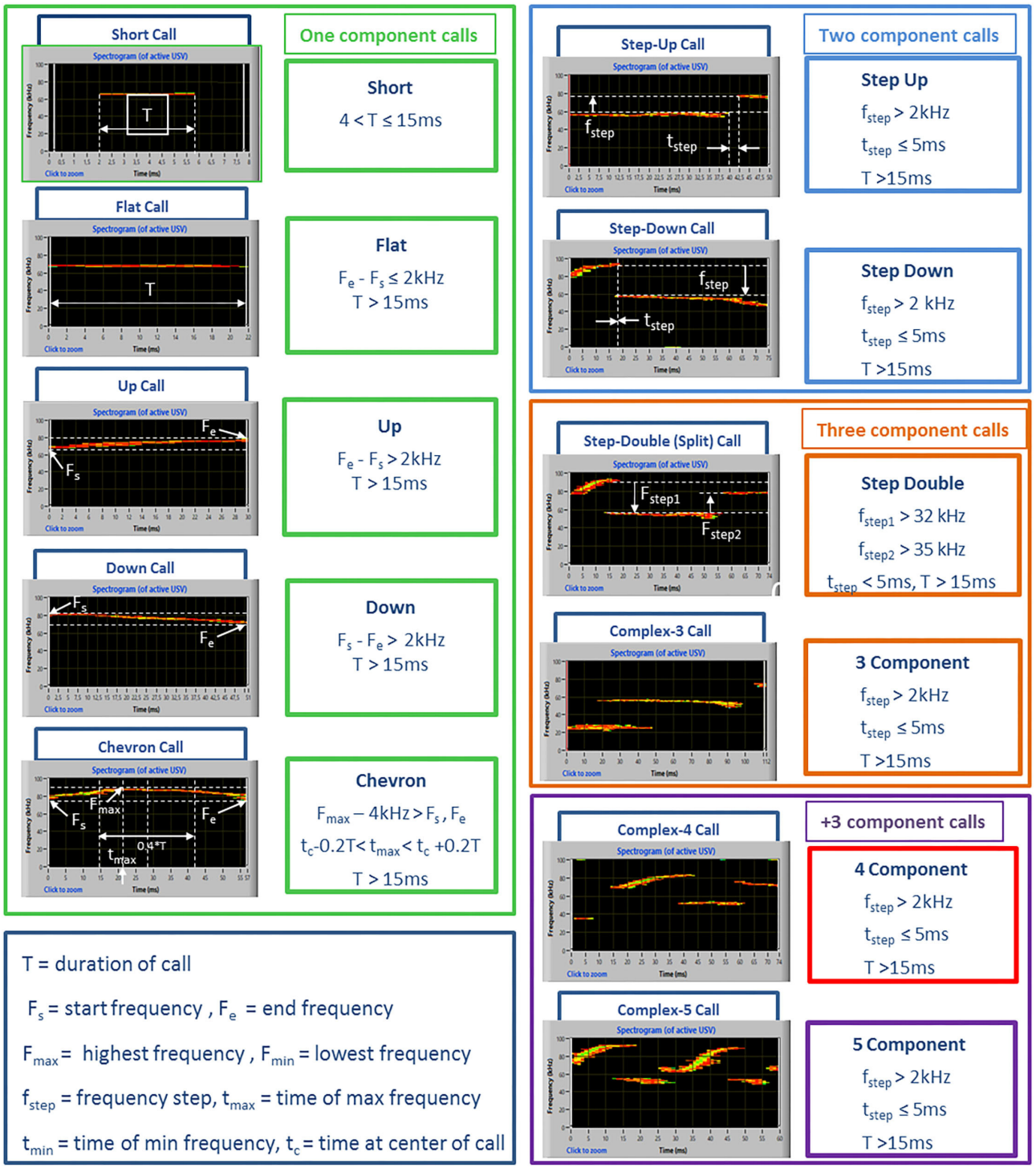


FIGURE 1 | Examples of ultrasonic call types used to classify USVs in the study. The call types were automatically classified using the software Sonotrack and based on the parameters described above. Definitions of the call types were mutually exclusive. Overlap of components was removed when more than 70 % to prevent wrong call durations. Short gaps between components in both frequency (≤6 kHz) and time (≤5 ms) were interpolated (gaps can be caused by changes in microphone sensitivity or direction of vocalization). Complex "3 component" and "+3 component" calls were summed up into a "complex tot" category.

TABLE 1 | Estrous phase of the resident B6 female mice and the CD1 stimulus female intruders for each testing session.

Experiment	Estrous phase	Session 1		Session 2		Session 3	
		Estrous	Diestrous	Estrous	Diestrous	Estrous	Diestrous
1	Resident	5*	5	4*	6	4*	6
1	Intruder	5	5*	4*	6	5*	5
2	Resident	4	6	7	3	7	3
2	Intruder	5	5	6	4	5	5

*While the female CD1 intruders used for assessing males' USVs and social behaviors in male-female interactions were all in diestrous phases, the intruders used for female-female interactions were either in diestrous or estrous phase. Their assignment on each testing day was counterbalanced according to the estrous phase of the B6 experimental female mice (i.e., approximately half of the intruders in estrous phase encountered a resident in estrous and the other half was assigned to a resident in diestrous; the same design was applied to the intruders in diestrous). The female CD1 stimuli at the time of the first testing session were naïve to encounters with mice of different strains (i.e., they have had pre-testing social interactions only with their same-sex and same–strain cagemates); they were used for a total of 3 times on the 3 subsequent testing sessions, but they were tested only once for each session. Separate batches of stimulus females were used for female-female interactions and male-female tests, and for each experiment (for a total of 40 CD1 female and 10 B6 male + 10 B6 female mice for each experiment). *One female B6 was excluded from the analysis of social behaviors because she was a statistical outlier on the time spent in affiliative behaviors (based on Grubbs' ESD test).*

RESULTS

Experiment 1: Same Pre-Testing Isolation Time in Both Sexes

Sex Differences: Social Interaction and USVs

Social behaviors were overall more markedly expressed in the female mice than in the male mice and tended to decrease with testing sessions; furthermore, sex differences depended on specific type of considered social behavior (**Figures 2A–E**). The time spent performing nose sniffing was overall similar in mice of both sexes and tended to decrease with repeated testing sessions [sex effect and its interaction, n.s., session effect: $F_{(2,34)} = 10.18$, $p < 0.001$, **Figure 2A**]. The female mice displayed more body sniffing than the male mice [sex effect: $F_{(1,17)} = 22.31$, $p < 0.001$, **Figure 2B**], and this effect was mostly observed during the first 2 sessions, since on the 3rd encounter body sniffing decreased in the female mice but increased in the male mice [interaction sex × session: $F_{(2,34)} = 18.59$, $p < 0.0001$, **Figure 2B**; *post-hoc*: $p < 0.05$]. The time spent performing anogenital sniffing did not differ overall between sexes, but it decreased with testing sessions in the male mice only [overall interaction sex × session: $F_{(2,34)} = 2.78$, $p < 0.07$, session effect on the male mice: $F_{(2,18)} = 14.68$, $p < 0.01$; on the female mice: ns, **Figure 2C**]. The female mice showed a tendency to display more affiliative behaviors than the male mice and explored significantly less the testing cage than males [sex effect, respectively: $F_{(1,17)} = 3.22$, 6.43, $p = 0.09$ and $p < 0.05$; **Figures 2D,E**]. In mice of both sexes, the levels of affiliative behaviors tended to decrease with testing sessions while those of cage exploration increased [session effect, respectively: $F_{(2,34)} = 25.01$ and 10.26, $p < 0.001$; **Figures 2D,E**].

Several characteristics of USVs differed between the two sexes (**Figures 2F–J**). Although the number of USVs emitted was not significantly different (**Figure 2F**), the total calling time and the mean duration were higher in the female mice [sex effect, respectively: $F_{(1,18)} = 4.85$ and 4.44, $p < 0.05$; **Figures 2G,H**], while the peak frequency tended to be lower than in the male mice [sex effect: $F_{(1,18)} = 3.62$, $p = 0.07$; **Figure 2I**]. All the USV parameters did not significantly change across the testing sessions, with the exception of peak amplitude that increased from the first to the second session in mice of both sexes [session effect: $F_{(2,36)} = 4.27$, $p < 0.05$; **Figure 2J**]. Although the interaction sex × session did not reach statistical significance, it was evident that the number of USVs decreased across the testing sessions in the male mice only [separate ANOVAs on the male mice: $F_{(2,18)} = 9.59$, $p < 0.01$; in the female mice:, ns; **Figure 2F**].

The types of ultrasonic calls, as classified based on most common spectrographic categories, differed between sexes, and this pattern of results seemed more evident on the first 2 testing sessions (**Figures 3**, **4**). The female mice tended to emit less "short" calls on the first testing session than the male mice [interaction sex × session: $F_{(2,36)} = 2.73$, $p = 0.08$, sex effect on session 1: $F_{(1,18)} = 3.17$, $p = 0.09$; **Figure 3A**]. Especially during the second session, the female mice also produced overall less "up" calls [sex effect: $F_{(1,18)} = 7.27$, $p < 0.05$; interaction sex × session: $F_{(2,36)} = 3.05$, $p = 0.06$, sex effect on session 2: $F_{(1,18)} = 16.01$, $p < 0.001$; **Figure 3C**] and more "down" calls [sex effect: $F_{(1,18)} = 4.91$, $p < 0.05$; interaction sex × session: $F_{(2,36)} = 3.2$, $p = 0.05$, sex effect on session 2: $F_{(1,18)} = 10.93$, $p < 0.01$; **Figure 3D**]. The female mice also emitted more "step double and more "complex" calls with 3 or more components [sex effect, respectively: $F_{(1,18)} = 11.44$ and 7.3, $p < 0.05$; **Figures 3H,I**].

The Effects of Estrous Phase: Social Interaction and USVs in Female Mice

While the male B6 mice were all tested with a female CD1 stimulus in diestrous phase, the female B6 mice were tested on each session with a female CD1 either in estrous or diestrous phase, with a balanced assignment across sessions (refer also to **Table 1**). The estrous phase of the stimulus females did not affect the social behaviors in any of the testing sessions, neither any of the USVs' characteristics (stimulus' estrous phase effect for each testing session: all ns).

The estrous phase of the experimental B6 female mice affected their social behaviors on the first 2 testing sessions (**Figures 5A–E**): the female mice in estrous phase displayed less anogenital sniffing on session 1 and more on session 2 than those in diestrous [estrous phase effect on sessions 1 and 2, respectively: $F_{(1,7)} = 55.3$ and 5.96, $p < 0.001$ and <0.05; **Figure 5C**], and the

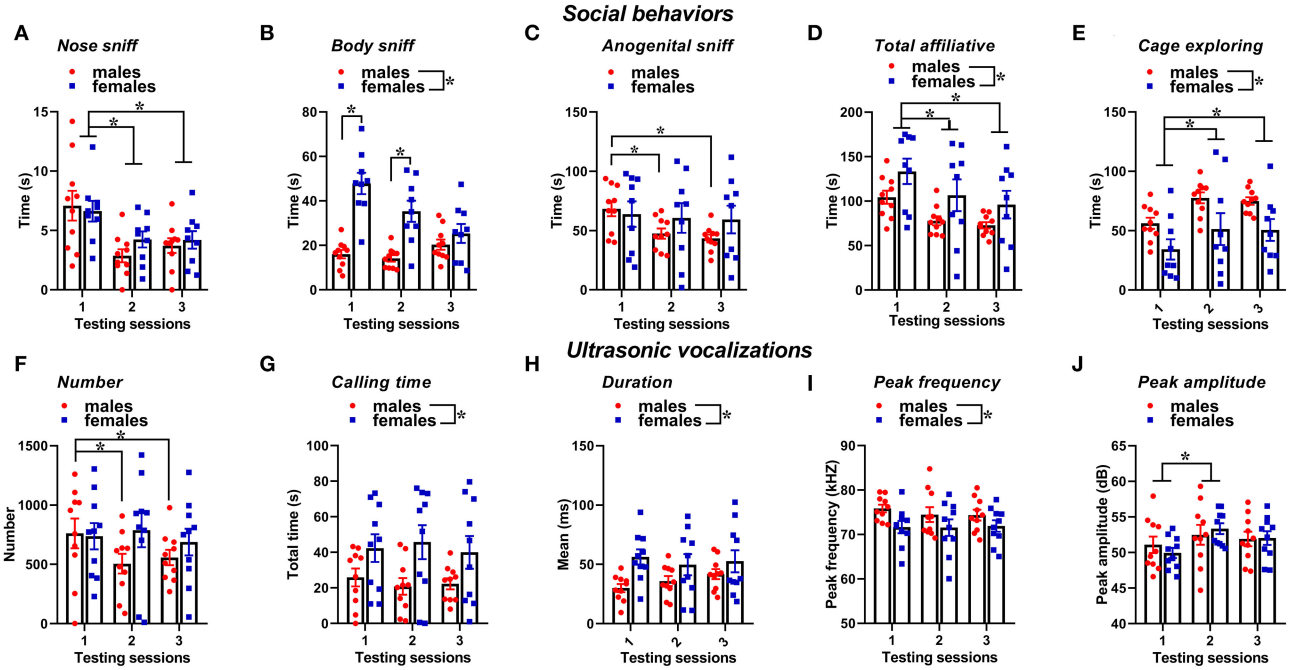


FIGURE 2 | Sex differences in social behaviors and USVs in experiment 1 (same pre-testing isolation). All behaviors were scored in the male and female experimental B6 subjects **(A–E)** toward a female CD1 stimulus during 3 testing sessions of 3 min each (inter-session interval: 7–10 days). Time spent performing social behaviors and cage exploration was scored from video files by an observer who was unaware of the sex of the subjects and their experimental assignment. Affiliative behaviors refer to all sniffing + crawl under/over + allogrooming (refer also to definitions in the text). Cage exploration refers to locomotion in the testing cage and wall rearing. USVs were quantified by spectrographic analysis using the software Avisoft SASLab Pro **(F–J)**. They were emitted either by male or female experimental B6 subjects toward a female CD1 stimulus during the 3 testing sessions of 3 min each. Data are mean ± SEM. $N = 10$ before exclusion of statistical outliers by Grubb's test for small samples.*$p < 0.05$. * Refers to a nonsignificant tendency ($0.05 < p \leq 0.09$). Sex differences are reported as * in each graph legend when a significant main effect of sex was detected in the absence of any interaction with testing session.

same pattern was observed for total affiliative behaviors [estrous phase effect on sessions 1 and 2, respectively: $F_{(1,7)} = 24.97$ and 6.39, $p < 0.01$ and <0.05; **Figure 5D**]. In parallel, cage exploration was more evident in the estrous than in the diestrous female mice in session 1 [estrous phase effect on session 1: $F_{(1,7)} = 7.89$, $p < 0.05$; **Figure 5E**] and tended to decrease afterward.

The estrous phase of the experimental B6 female mice also affected certain characteristics of the USVs they emitted, especially in the first testing session (**Figures 5F–J**). The female mice in estrous phase emitted less USVs in session 1 than those in diestrous phase [estrous phase effect on session 1: $F_{(1,8)} = 16.74$, $p < 0.05$; **Figure 5F**]. The female mice in estrous phase also spent less time calling in session 1 compared to those in diestrous phase, while an increase was observed in session 2 [estrous phase effect on s1 and s2, respectively: $F_{(1,8)} = 12.50$ and 5.99 $p < 0.05$; **Figure 5G**]. The peak frequency of the USVs emitted by the female mice in estrous phase was also higher, although this effect was again detectable only in session 1 [estrous phase effect on session 1: $F_{(1,8)} = 5.35$, $p < 0.05$; **Figure 5I**]. No difference was found in the distribution of call types between estrous and diestrous experimental female subjects (**Supplementary Figure 2**).

Experiment 2: Different Pre-Testing Isolation Time in Male and Female Mice

Sex Differences: Social Interaction and USVs

Similar to experiment 1, also, when the male mice were not isolated before testing, social behaviors were overall more expressed by the female mice than the male one and tended to decrease with testing sessions; furthermore, sex differences depended on specific type of behavior (**Figures 6A–E**). The female mice displayed more body sniffing than the male mice, especially in the first testing session [interaction sex × session: $F_{(2,34)} = 5.34$, $p < 0.01$; sex effect on session 1: $F_{(1,17)} = 10.36$, $p < 0.01$; **Figure 6B**]. The female mice were also overall engaged in more anogenital sniffing than the male mice [sex effect: $F_{(1,17)} = 16.72$, $p < 0.01$, **Figure 6C**], an effect that was stable across the sessions. The female mice display more affiliative behaviors and explored significantly less the testing cage than the male mice [sex effect, respectively: $F_{(1,17)} = 9.61$ and $F_{(1,18)} = 20.51$, $p < 0.01$; **Figures 6D,E**]. In mice of both sexes, the levels of affiliative behaviors tended to decrease with testing sessions, while those of cage exploration increased [session effect, respectively: $F_{(2,34)} = 6.55$ and $F_{(2,36)} = 10.98$, $p < 0.01$; **Figures 6D,E**].

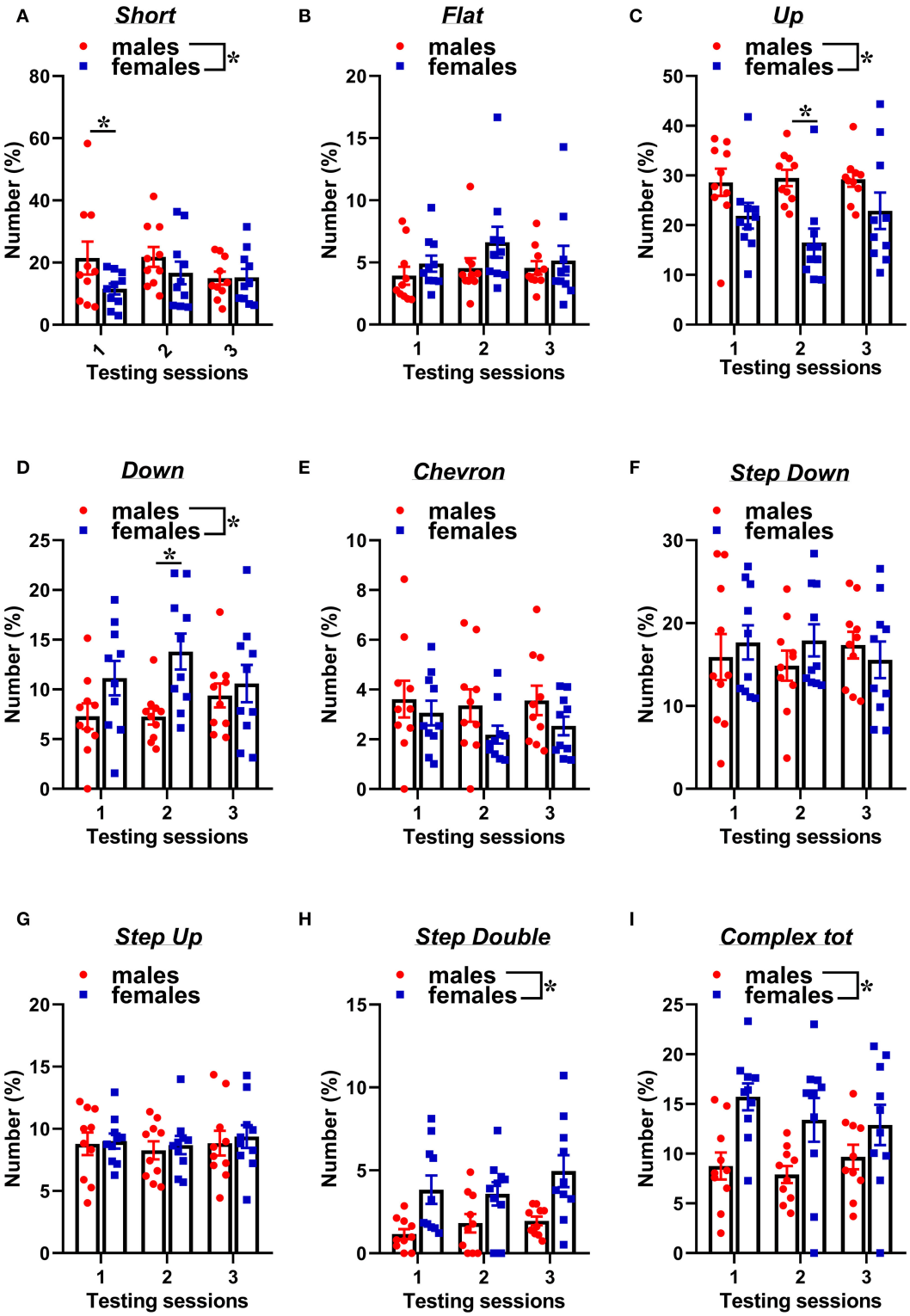


FIGURE 3 | Sex differences in ultrasonic call types in experiment 1 (same pre-testing isolation). **(A–I)** The different call types were automatically classified as detailed in **Figure 1**. Complex tot = complex 3 + complex 4+ complex 5. Data are expressed as percentages over the total number of USVs for each sex and session. Data are mean ± SEM. $N = 10$ for each sex. *$p < 0.05$. * Refers to a nonsignificant tendency ($0.05 < p \leq 0.09$). Sex differences are reported as * in each graph legend when a significant main effect of sex was detected in the absence of any interaction with testing session.

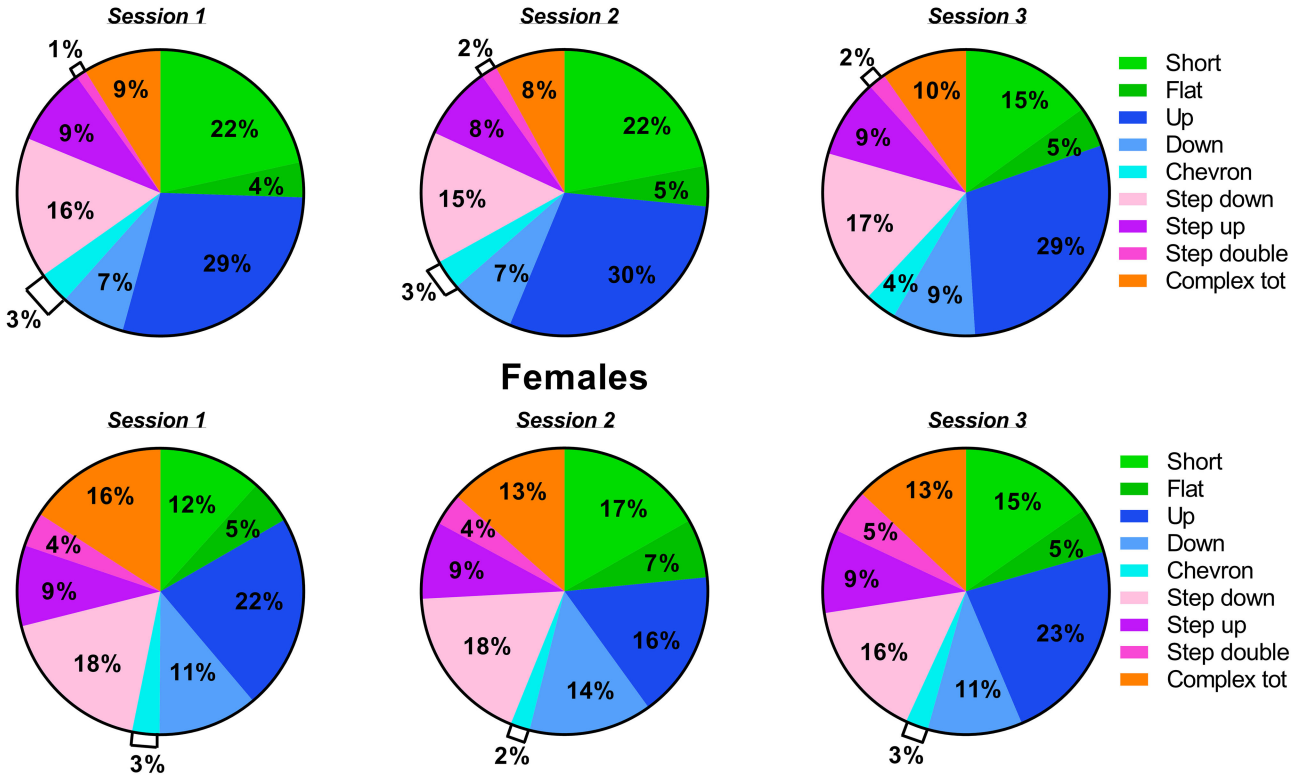


FIGURE 4 | Pie charts depicting sex differences in ultrasonic call types in experiment 1 (same pre-testing isolation). Distribution of call categories in each sex and testing session. Data are expressed as percentages over the total number of USVs for each sex and session. Data are mean ± SEM. $N = 10$ for each sex. Complex tot = complex 3 + complex 4 + complex 5 (refer also to **Figure 1**).

Several characteristics of USVs differed between the two sexes (**Figures 6F–J**), in a highly similar manner to what was observed in experiment 1. Although the number of USVs emitted was not significantly different (**Figure 6F**), the total calling time and the mean duration were higher in the female mice [sex effect, respectively: $F_{(1,16)} = 6.6$ and 5.32, $p < 0.05$; **Figures 6G,H**]. All the USV parameters did not significantly change across the testing sessions (session effect and its interaction with sex: all ns).

The classification of the call types revealed several sex differences that were mostly independent of the testing sessions (**Figures 7**, **8**). The female mice tended to emit less "short" calls than the male mice [sex effect: $F_{(1,18)} = 3.53$, $p = 0.08$; **Figure 7A**]. The female mice also produced more "down" calls [sex effect: $F_{(1,18)} = 9.15$, $p < 0.01$; **Figure 7D**], more "step double" [sex effect: $F_{(1,18)} = 4.12$, $p = 0.06$ **Figure 7H**], and more complex calls [sex effect: $F_{(1,18)} = 7.3$, $p < 0.05$; **Figure 7I**].

The Effects of Estrous Phase: Social Interaction and USVs in Female Mice

As for experiment 1, the female B6 mice were tested in each session with a female CD1 either in estrous or diestrous phase, with a balanced assignment across sessions and between resident's estrous phases (**Table 1**). As observed in experiment 1, the estrous phase of the female stimulus (intruder) did not affect the social behaviors in any of the testing sessions or any of the USV characteristics (stimulus' estrous phase effect for each testing session: all ns).

In contrast to what was observed in experiment 1, the estrous phase of the female resident did not modulate any of its social behaviors in any testing session (effects of estrous phase on all sessions: ns; **Figures 9A–E**). Furthermore, no significant effect of the resident's estrous phase was found on any of the USV characteristics (effects of estrous phase on all sessions: ns; **Figures 9F–J**) or on their composition based on call types (**Supplementary Figure 3**).

Comparison Between Experiments 1 and 2 in Female Mice: Social Interaction and USVs

The female mice in both experiments were tested under the same experimental conditions, i.e., following a 72-h isolation period before social encounters. Since the female subjects belonged to

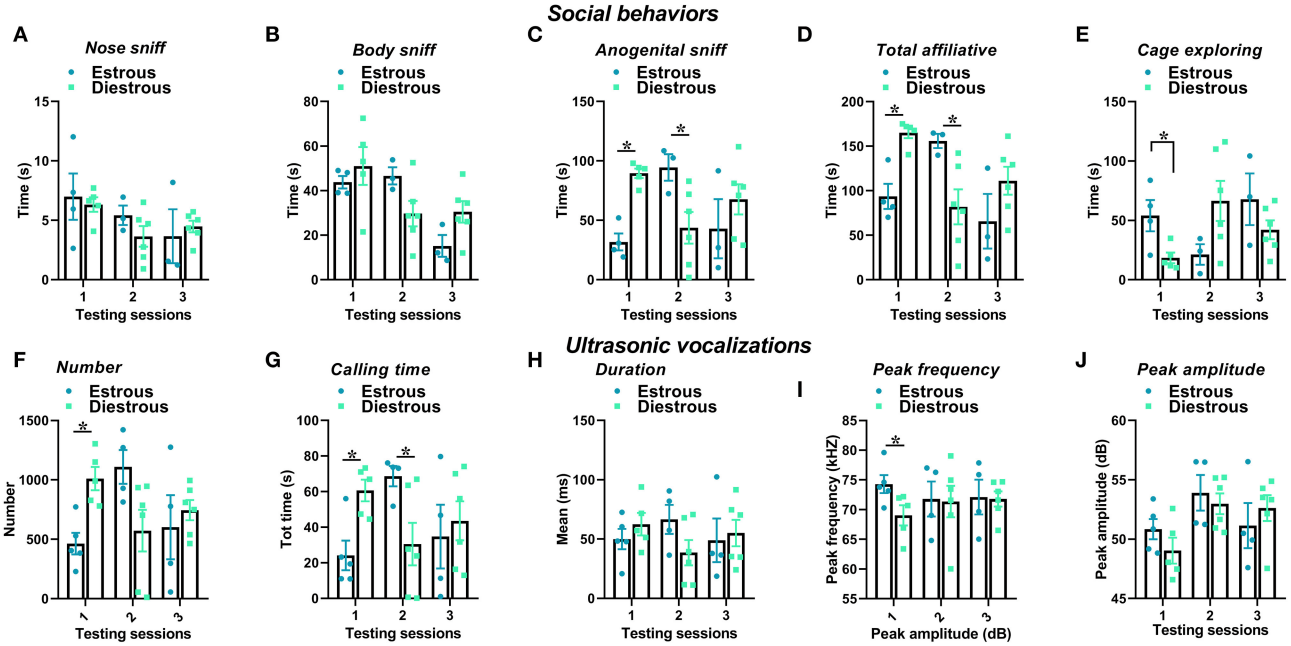


FIGURE 5 | Effects of the estrous phase of the resident female mice on social behaviors and USVs in experiment 1 (same pre-testing isolation). The impact of the estrous phase of the experimental female B6 subjects was investigated on each testing day on social behaviors **(A–E)** and USVs' characteristics **(F–J)**. All the B6 mice were isolated for 3 days in the testing cage before each social encounter. For the exact number of mice in each estrous phase, refer to **Table 1**. Data are mean ± SEM. Total $N = 10$ before exclusion of statistical outliers by Grubb's test for small samples.*$p < 0.05$.

two independent cohorts, in order to evaluate the replicability of female phenotypes, we analyzed the female dataset by an additional ANOVA with experiment as the between-subject factor and testing session as the within-subject variable.

Concerning social behaviors, the results on session-dependent changes were similar between the experiments, as shown by lack of the interaction experiment × session (all effects, n.s.). Independently of the experiment, the levels of affiliative behaviors tended to decrease with the testing sessions while those of cage exploration increased [session effect, respectively: $F_{(2,32)} = 10.31$ and $F_{(2,34)} = 10.47$, $p < 0.001$; **Figures 2**, **6**]. Independently of the experiments, all the USV parameters did not significantly change across the testing sessions (**Figures 2**, **6**), and the expression of the different call types (**Figures 3**, **4**, **7**, **8**; session effect and its interaction with experiment: all ns). Significant interaction experiment × session was only found in peak amplitude that tended to increase from the first to the second session but only in the first experiment [$F_{(2,34)} = 5.5$, $p < 0.05$; **Figure 2J**].

The effects of the estrous phase of the resident female mice were instead statistically different between the two experiments, both on social behaviors and USVs parameters, as expected because of the presence of estrous phase effects in the first but not in the second experiment. A significant interaction experiment × estrous phase was found in the time spent in anogenital sniffing, affiliative behaviors, and cage exploration (**Figures 5**, **9**) in session 1 [$F_{(1,15)} = 8.62$, 9.28, 5.27; $p < 0.05$] and in affiliative time in session 2 [$F_{(1,15)} = 6.14$; $p < 0.05$]. Concerning the USV parameters (**Figures 5**, **9**), a significant interaction experiment × estrous phase was found for session 1 in the number of USVs and calling time [$F_{(1,16)} = 6.13$, 10.2; $p < 0.05$].

DISCUSSION

Our findings provide convincing evidence for sex differences in ultrasonic communication and social interaction in the C57BL/6J mouse strain. In the context of male-female vs. female-female interactions, differences in ultrasonic communication between sexes were mainly qualitative, while those in social behaviors were both quantitative and qualitative. Sex differences were highly similar between the two experiments, i.e., their detection was not substantially affected by differences in pre-testing isolation of the male mice. Nonetheless, subtle differences in the social and ultrasonic profiles of the male mice emerged between the two experiments, suggesting an impact of pre-testing social isolation on the male mice. Sex differences were mostly stable across the three testing sessions with an unfamiliar intruder, although an overall tendency to social habituation occurred in both experiments and sexes.

The estrous cycle of the resident female mice altered the social behaviors and ultrasonic communication of the female mice, although these effects were significantly detected only in the first experiment. Here, both quantitative and qualitative

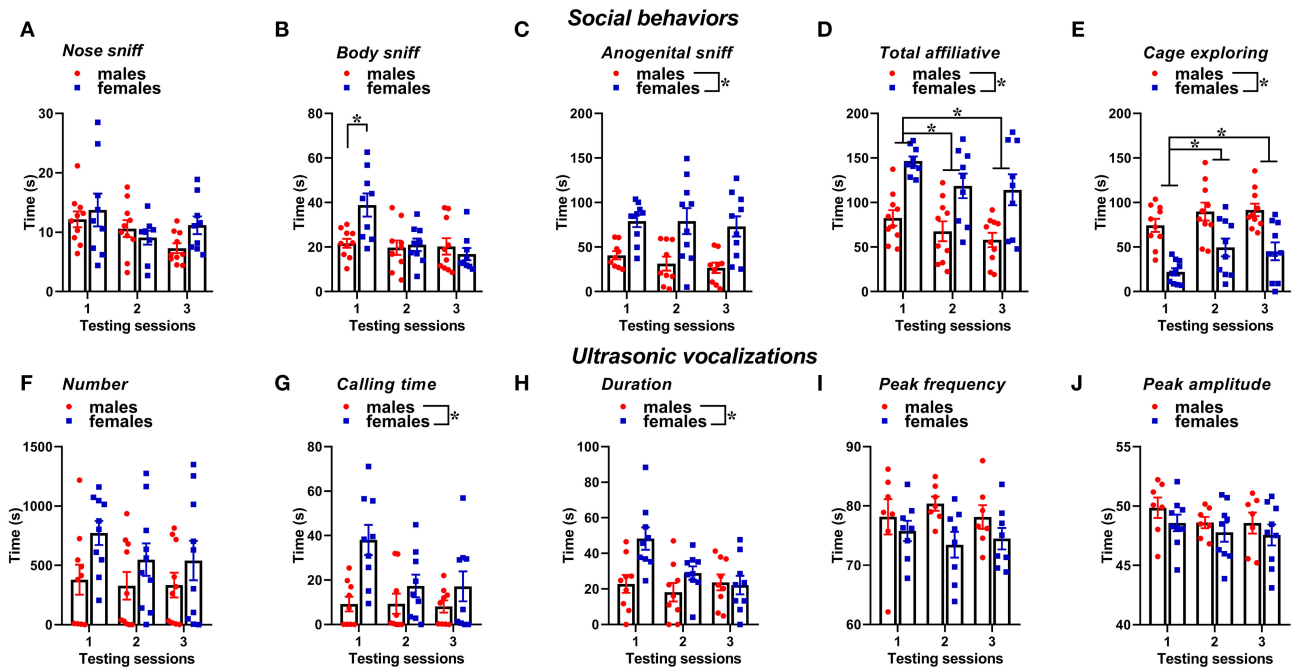


FIGURE 6 | Sex differences in social behaviors and USVs in experiment 2 (different pre-testing isolation). **(A–E)** Social behaviors and **(F–J)** USVs of the male and female experimental B6 subjects were assessed by analysis of the video recordings (Observer, Noldus) and spectrograms (Avisoft SASLab Pro), respectively. The B6 female mice were isolated for 3 days in the testing cage before each social encounter as in experiment 1, while the male mice were isolated for only 10 min pre-testing, as in most commonly used procedures. As in experiment 1, social behavior and USVs were analyzed during 3 testing sessions of 3 min each (inter-session interval: 7–10 days) using an adult female CD1 as stimulus. Data are mean ± SEM. $N = 10$ before exclusion of statistical outliers by Grubb's test for small samples.*$p < 0.05$. Sex differences are reported as * in each graph legend when a significant main effect of sex was detected in the absence of any interaction with testing session.

differences were indeed observed between receptive and non-receptive female residents, while the estrous phase of the intruder did not modulate any of the considered behavioral parameters.

Sex Differences and Isolation Effects (Comparison Between Experiments 1 and 2)

Sex differences in social behaviors and ultrasonic communication were overall highly comparable between the two experiments, with the female mice displaying more affiliative behaviors and less cage exploration than the male ones while emitting longer USVs and with less simple one-component calls (e.g., "short" and "up") but more complex calls (e.g., "step double" and "total complex"). Nonetheless, subtle additional sex differences were found only in experiment 1, including a male-specific decrease with testing sessions in anogenital sniffing and USV number, as well as an overall higher peak frequency of male USVs. Hence, our data suggest that the experimental settings used in our experiment 1, including a resident-intruder context with 72-h pre-testing isolation, may be the most suitable to detect both major and minor sex differences in social and ultrasonic behaviors. Since the female mice were tested under exactly the same experimental conditions in experiment 2, it is natural to infer that the differences emerging between our two experiments are due to corresponding differences in the behaviors of the male mice that were exposed to pre-testing isolation only for experiment 1. The behavior of the female mice was indeed highly comparable in our two experiments, as confirmed by the statistical comparison of the two female datasets, supporting the replicability of female social and ultrasonic behavioral profiles across the repeated testing sessions. Male behaviors appeared instead slightly different between the two experiments, although we could not conduct a statistical quantification of these differences because of the confounding effects of independent testing on social isolation.

Nonetheless, the visual comparison of **Figure 2** with **Figure 6** and **Figure 3** with **Figure 7** clearly shows that the male mice in experiment 1, i.e., with longer pre-testing isolation, displayed higher levels of anogenital sniffing, more USVs, and more one-component 'up" calls, suggesting a higher expression of these behaviors in territorial, i.e., isolated, male mice. The hypothesis of a territoriality effect of social isolation is further supported by the predominance of the differences between the two experiments during the first testing session, since this effect may be attenuated by repeated experience of social encounters with an intruder. Furthermore, it should be noted that a resident-intruder setting was employed only in experiment 1, while experiment 2 used a basically neutral testing environment as a consequence of the short pre-testing isolation (10 min). The reason for this

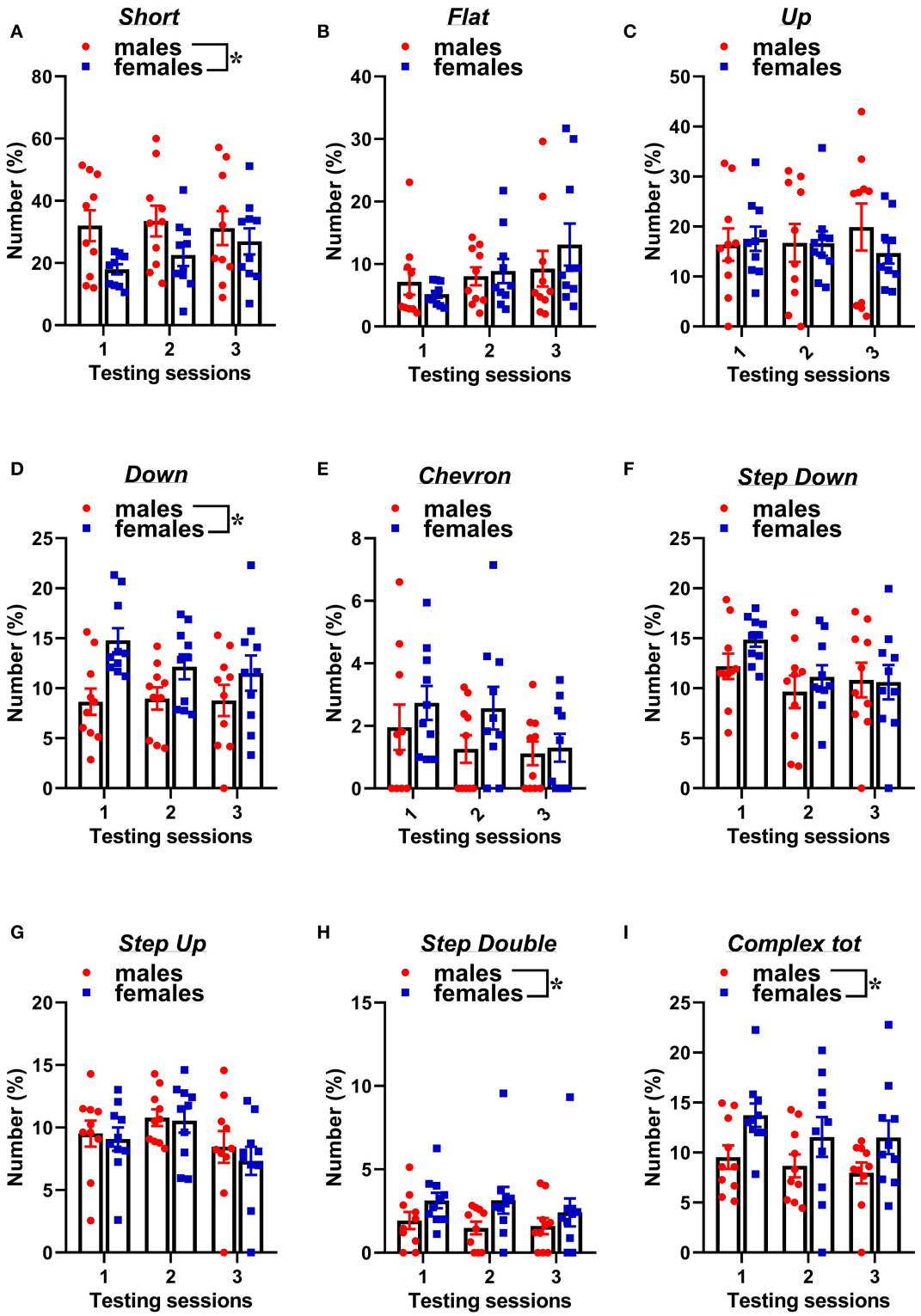


FIGURE 7 | Sex differences in ultrasonic call types in experiment 2 (different pre-testing isolation). **(A-I)** The different call types were automatically classified (refer to **Figure 1** for a detailed description). Complex tot = complex 3 + complex 4 + complex 5. Data are expressed as percentages over the total number of USVs for each sex and session. Data are mean ± SEM. $N = 10$ for each sex. *$p < 0.05$.* Refers to a nonsignificant tendency ($0.05 < p \leq 0.09$). Sex differences are reported as * in each graph legend when a significant main effect of sex was detected in the absence of any interaction with testing session.

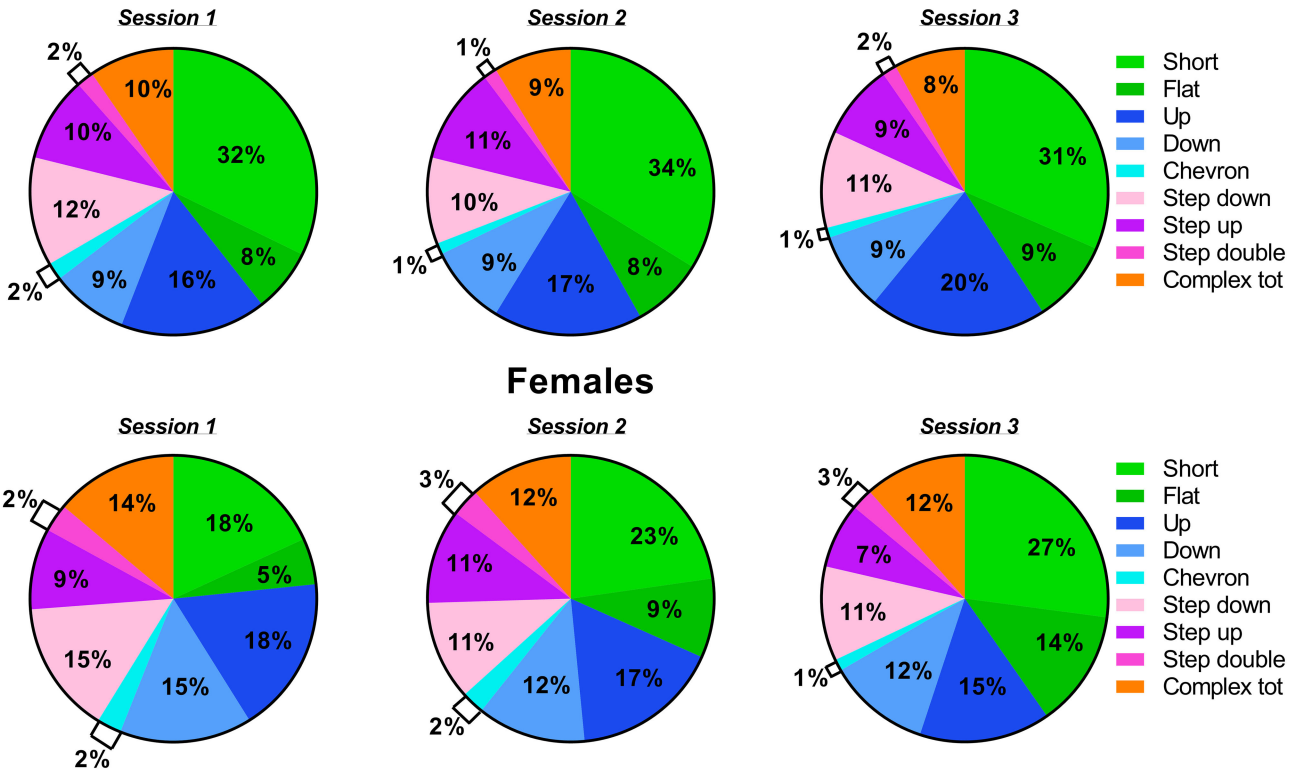


FIGURE 8 | Pie charts depicting sex differences in ultrasonic call types in experiment 2 (different pre-testing isolation). Distribution of call categories in each sex and testing session. Data are expressed as percentages over the total number of USVs for each sex and session. Data are mean ± SEM. $N = 10$ for each sex. Complex tot = complex 3 + complex 4+ complex 5 (refer also to **Figure 1**).

experiment design was intrinsically related to the major aim of our study, which was not to specifically investigate the effects of social isolation on ultrasonic communication in male and female mice, as previously conducted by others (e.g., Zhao et al., 2021; this also explains the lack of an additional female group with minimal pre-testing isolation in our design). Instead, our goal was to evaluate sex differences either under the exact same experimental conditions for male and female mice (i.e., in the resident-intruder paradigm) or under the experimental conditions most suitable and commonly used in research studies on USVs in ASD mouse models (i.e., in a resident-intruder setting for female mice and with a short habituation to the testing environment in the case of male subjects).

Interestingly, when the effects of 72-h social isolation were previously assessed, only subtle changes were described in male B6 mice (Zhao et al., 2021). These discrepancies may be due to the longer duration of the testing session used in the previous study (i.e., 30 vs. 3 min in ours): Zhao et al. indeed described no effects of isolation on the number of USVs emitted by a male mouse toward a female intruder when the entire session was considered, but they detected a significant increase in isolated vs. grouped male mice when only the first 5 min of the session was analyzed, and they described a higher first latency to USV emission in the isolated male mice (with an average value of ∼3 min, i.e., the duration of our testing session). Differences in the estrous cycle of the female intruder could also contribute to the discrepant outcomes of ours and Zhao's study on social non-vocal behaviors of isolated male mice: the authors reported a tendency, although not significant ($p = 0.08$) to an overall increase in the time spent in social interaction in isolated male mice compared to grouped ones that was accompanied by an increase in the occurrence of mounting behavior. The higher engagement in mounting of their isolated male mice could have attenuated the isolation effects on affiliative behaviors that we instead found in our study. We did not detect mounting in our tested male mice, and this is not surprising considering our short testing duration and the non-receptive estrous state of our female intruders (the estrous cycle was not assessed in Zhao's study). In conclusion, our results from the male mice and their comparison with previous findings suggest that 3 days of isolation of the male mice increases social affiliation and promotes USV emission during the initial phases of the social encounter with a female mouse. Nonetheless, the

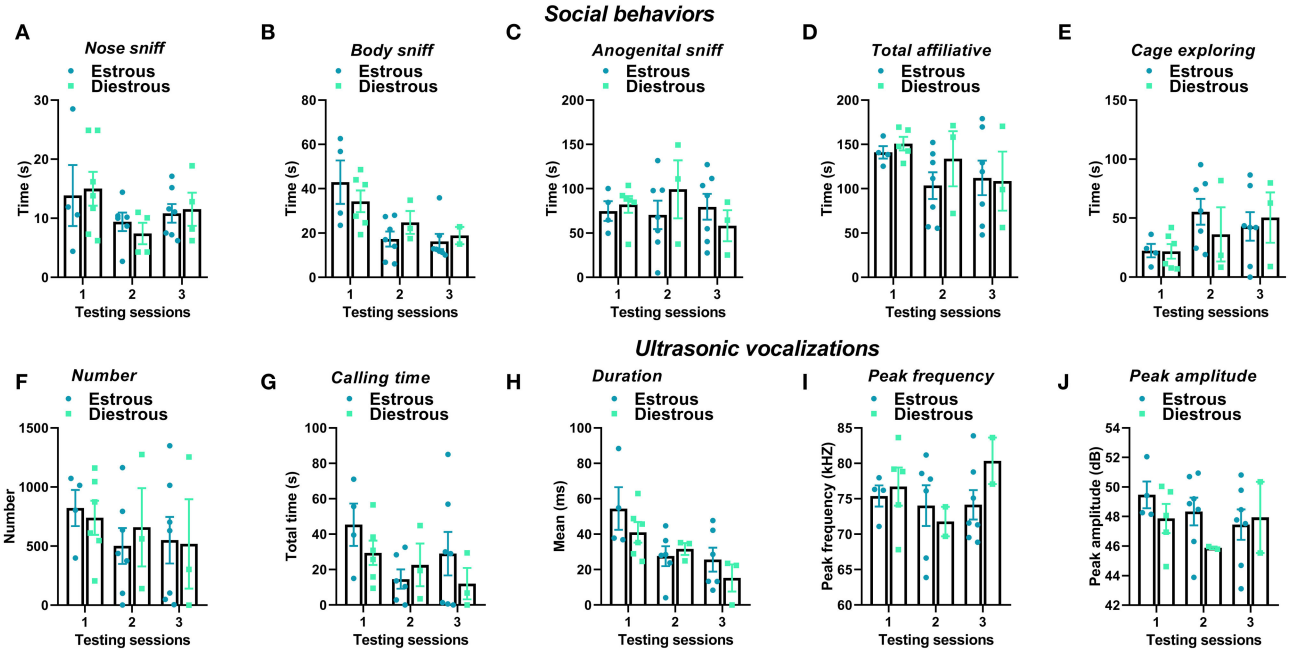


FIGURE 9 | Effects of the estrous phase of the resident female mice on social behaviors and USVs in experiment 2 (different pre-testing isolation). The effects of the estrous phase of the experimental female B6 subjects were investigated on each testing day on **(A–E)** social behaviors and **(F–J)** multiple USV characteristics. The B6 female mice were isolated for 3 days in the testing cage before each social encounter, as in experiment 1, while the male mice were isolated for only 10 min pre-testing, as in most standard protocols. Social behaviors and USVs were analyzed following otherwise the same procedures used for experiment 1 (3 testing sessions of 3 min each and at an interval of 7–10 days). A detailed distribution of the estrous phase in the female residents and intruders is provided in **Table 1**. Data are mean ± SEM. $N = 10$ before exclusion of statistical outliers by Grubb's test for small samples.

duration of the testing session and the estrous cycle of the intruders may critically influence the social effects of isolation, an issue that deserves to be specifically investigated in future studies.

Independently of the experiments, our findings suggest that the major sex differences affecting mouse ultrasonic communication are of qualitative nature (duration and call composition) rather than quantitative. While the presence of longer USVS in the female mice was in line with previous studies (e.g., von Merten et al., 2014), the lack of sex difference in the number of USVs that we found here is in disagreement with a previous study reporting that female mice emitted more USVs than male mice toward a female intruder (Hammerschmidt et al., 2012). This discrepancy may be due to the different substrain used in this study, since B6/N and B6/J are known to have markedly different ultrasonic profiles. Indeed, in B6/J mice, another study described no difference in the number of USVs (Matsumoto and Okanoya, 2018) but a reduced number of short calls and prevalence of complex calls in female mice. As short calls, together with simple calls in general, have been detected especially under territorial conditions, e.g., in male-male interactions (Matsumoto and Okanoya, 2018) and following long-term male isolation (Chabout et al., 2012), it is possible that male mice preferentially communicate using short calls. In contrast, complex calling bouts may be useful in maintaining the group structure necessary for female mice and promote interactions and cooperation (Matsumoto and Okanoya, 2018), in agreement with previous studies showing that this type of call is more attractive for female mice (Chabout et al., 2015). It is indeed increasingly accepted that ultrasonic calls from female mice facilitate proximity between animals in order to help residents to acquire relevant social information on intruders and promote group relationships (Moles et al., 2007).

Since the strain of the mice involved in the social encounter may play a role in their ultrasonic profile and potential related sex differences, it is important to underscore that our study employed different strains for the test subjects (B6) and the stimulus mice (CD1). Although mouse USVs are often analyzed during interactions within the same strain, our experimental setting is not unusual, as it has been employed in previous studies assessing the emission of USVs by resident female mice (Maggio and Whitney, 1985) and male mice (Sugimoto et al., 2011) during dyadic interactions. One study in particular (Sugimoto et al., 2011) demonstrated a lack of USVs emitted by the female stimulus (derived from the CD1 strain) also when the "devocalized" male was of a different background (B6), i.e., under conditions highly similar to ours. Furthermore, using stimuli of a different strain to induce emission of USVs by tested subjects is a typical procedure of several studies on urine-elicited USVs (Nyby et al., 1979, 1983; Nyby, 2010). The choice of the CD1 strain as stimulus enhances the applicability of our present data

to the research field of neurodevelopmental disorders. Indeed, several studies with mouse models of Autism and Fragile X syndrome obtained from the B6 background have performed social and ultrasonic testing (in both male and female mice) through interactions with female CD1 stimuli (Hebert et al., 2014; Pietropaolo et al., 2014; Oddi et al., 2015; Gaudissard et al., 2017; Gauducheau et al., 2017; Lemaire-Mayo et al., 2017; Fyke et al., 2021).

Estrous Cycle Effects on Social Behaviors and USVs (Experiments 1 and 2)

In both experiments, the estrous cycle of the intruders did not alter either the social behaviors or any parameter of ultrasonic communication. The estrous cycle of the residents instead modulated both behavioral domains, and these effects were more marked in experiment 1 (**Figure 5**) the female mice in estrous phase exhibited less affiliative behaviors in session 1 than those in diestrous phase, but this tendency inverted its direction in session 2 to return to the initial situation in session 3. These effects were mainly due to differences in anogenital sniffing. The effects of estrous cycle on social behavior followed those observed on USVs, since the number of USVs was also initially and finally lower in the estrous than in the diestrous female mice with a shift in session 2. USVs seemed more affected by the estrous cycle in session 1 when their number, call time, and peak frequency were all different between estrous and diestrous residents. The overall reduced social investigation and number of USVs of the estrous female mice are in agreement with previous reports (Moles et al., 2007) and fits with the reduced social interest of receptive female mice in a conspecific of the same sex. It has been suggested that oxytocin mediation of social processes is likely to play a role in the effects of the estrous phase, since it is known to be regulated by ovarian circulation (Choleris et al., 2003). Nonetheless, we failed to replicate the significant effects of the estrous cycle in experiment 2 despite the experimental testing conditions of the female mice being unchanged. It should be noted that the pattern of results of experiment 2 was still in line with what observed in experiment 1 (**Figure 9**), although the composition of the estrous vs. diestrous female in each session was less balanced than in experiment 1 (see **Table 1**). It is possible that the lower number of female mice, especially in the diestrous phase (3 in some testing sessions), may have limited the emergence of significant differences. In none of the experiments any effect of the estrous cycle was detected on the call types (**Supplementary Figures 2, 3**).

Effects of Testing Experience at the Group and Individual Levels (Experiments 1 and 2)

We observed several group differences in our study, both in social non-vocal and vocal behaviors. It is intriguing to question whether the group differences could be confirmed at the individual level; to this end, the visual evaluation of individual plots (**Supplementary Figures 4, 5**) supports interesting considerations. Concerning social non-vocal behaviors, most of the female individuals showed the expected reduction with testing sessions in the time spent performing body sniffing in both experiments (**Supplementary Figures 4B, 5B**), while a decrease in anogenital sniffing was confirmed in the male mice only in experiment 1 (**Supplementary Figure 4C**). In mice of both sexes in both experiments, the levels of affiliative behaviors tended to decrease with testing sessions while those of cage exploration increased (**Supplementary Figures 4D,E, 5D,E**).

Concerning individual trends in USV-related parameters, in experiment 1, we confirmed that all the USV parameters did not significantly change across the testing sessions, with the exception of peak amplitude that increased from the first to the second session in mice of both sexes (**Supplementary Figure 4J**) and the number of USVs that decreased across the testing sessions in the male mice only (**Supplementary Figure 4F**). None of the session differences appeared at the individual level in experiment 2 (**Supplementary Figures 5F–J**), when the male mice were not isolated before testing (confirming the lack of differences observed at the group level).

Concerning the types of ultrasonic calls, it is important to underscore that the stability of call compositions in each sex across the testing sessions was confirmed at the individual level (**Supplementary Figures 6, 7**) when we selected the individuals with the highest total calling rates (half of each sex for each experiment). The overall higher proportion of short calls in males mice and the lower proportion of complex calls ("step double" and "complex tot") were also evident in this subset of individuals.

CONCLUSIONS

Our findings provide novel evidence for marked sex differences in ultrasonic communication that are mirrored by differences in other social behaviors of adult B6 mice. The replication of the sex differences with and without pre-testing isolation in the male mice suggests their strong consistency and should be taken into account for designing future studies using male and female adult mice. This could be especially important for studies on genetic mouse models of ASD or other pathologies involving communication deficits where pre-testing isolation may induce undesirable confounding effects more marked than in their wild-type littermates (Pietropaolo et al., 2008). Furthermore, our results demonstrate that the sex differences observed in ultrasonic communication and social behaviors are not limited to the first testing session and represent a stable trait that seems independent of the novelty of the social stimulus. Finally, these data clearly show the importance of an extensive qualitative analysis of ultrasonic communication in adult mice, since this approach contributes to unravel the more complex structure of female vs. male calls.

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 Comité d'Ethique pour l'experimentation animale de Bordeaux (CE 50) and the French Ministry (Ministere de l'enseignement superieur de la recherché et de l'innovation).

AUTHOR CONTRIBUTIONS

MP conducted the experiments and behavioral analyses and participated in the writing of the manuscript. VP participated in the analysis of USVs, prepared all the figures, and contributed to the writing of the results. RB classified the call types. SP designed the experiments, provided technical support for all the experiments, and wrote the manuscript. SB contributed to manuscript writing. All authors reviewed and approved the manuscript.

FUNDING

SP received funding from Bordeaux University, CNRS, Association Autour de Williams and Fondation pour l'Audition (FPA-RD-2020-8).

ACKNOWLEDGMENTS

The authors thank Marion Piquemal-Langouilleat for her assistance in behavioral testing, Elodie Poinama for animal care, Thierry Lafon for informatics support, and Marie-Laure Rousseau for administrative assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnbeh.2022.883353/full#supplementary-material

REFERENCES

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

Caligioni, C. S. (2009). Assessing reproductive status/stages in mice. *Curr. Protoc. Neurosci.* 48:A-4. doi: 10.1002/0471142301.nsa04is48

Caruso, A., Ricceri, L., and Scattoni, M. L. (2020). Ultrasonic vocalizations as a fundamental tool for early and adult behavioral phenotyping of autism spectrum disorder rodent models. *Neurosci. Biobehav. Rev.* 116, 31–43. doi: 10.1016/j.neubiorev.2020.06.011

Chabout, J., Sarkar, A., Dunson, D. B., and Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Front. Behav. Neurosci.* 9, 76. doi: 10.3389/fnbeh.2015.00076

Chabout, J., Serreau, P., Ey, E., Bellier, L., Aubin, T., Bourgeron, T., et al. (2012). Adult male mice emit context-specific ultrasonic vocalizations that are modulated by prior isolation or group rearing environment. *PLoS ONE* 7, e29401. doi: 10.1371/journal.pone.0029401

Choleris, E., Gustafsson, J. A., Korach, K. S., Muglia, L. J., Pfaff, D. W., and Ogawa, S. (2003). An estrogen-dependent four-gene micronet regulating social recognition: a study with oxytocin and estrogen receptor-alpha and -beta knockout mice. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6192–6197. doi: 10.1073/pnas.0631699100

D'Amato, F. R., and Moles, A. (2001). Ultrasonic vocalizations as an index of social memory in female mice. *Behav. Neurosci.* 115, 834–840. doi: 10.1037/0735-7044.115.4.834

de Chaumont, F., Lemiere, N., Coqueran, S., Bourgeron, T., and Ey, E. (2021). LMT USV toolbox, a novel methodological approach to place mouse ultrasonic vocalizations in their behavioral contexts-a study in female and male C57BL/6J mice and in shank3 mutant females. *Front. Behav. Neurosci.* 15, 735920. doi: 10.3389/fnbeh.2021.735920

Demir, E., Li, K., Bobrowski-Khoury, N., Sanders, J. I., Beynon, R. J., Hurst, J. L., et al. (2020). The pheromone darcin drives a circuit for innate and reinforced behaviours. *Nature* 578, 137–141. doi: 10.1038/s41586-020-1967-8

Egnor, S. R., and Seagraves, K. M. (2016). The contribution of ultrasonic vocalizations to mouse courtship. *Curr. Opin. Neurobiol.* 38, 1–5. doi: 10.1016/j.conb.2015.12.009

Ferhat, A. T., Le Sourd, A. M., de Chaumont, F., Olivo-Marin, J. C., Bourgeron, T., and Ey, E. (2015). Social communication in mice–are there optimal cage conditions? *PLoS ONE* 10, e0121802. doi: 10.1371/journal.pone.0121802

Ferhat, A. T., Torquet, N., Le Sourd, A. M., de Chaumont, F., Olivo-Marin, J. C., Faure, P., et al. (2016). Recording mouse ultrasonic vocalizations to evaluate social communication. *J. Vis. Exp.* 5, e53871. doi: 10.3791/53871

Fyke, W., Premoli, M., Echeverry Alzate, V., Lopez-Moreno, J. A., Lemaire-Mayo, V., Crusio, W. E., et al. (2021). Communication and social interaction in the cannabinoid-type 1 receptor null mouse: implications for autism spectrum disorder. *Autism Res.* 14, 1854–1872. doi: 10.1002/aur.2562

Gaub, S., Fisher, S. E., and Ehret, G. (2016). Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. *Genes Brain Behav.* 15, 243–259. doi: 10.1111/gbb.12274

Gaudissard, J., Ginger, M., Premoli, M., Memo, M., Frick, A., and Pietropaolo, S. (2017). Behavioral abnormalities in the Fmr1-KO2 mouse model of fragile X syndrome: the relevance of early life phases. *Autism Res.* 10, 1584–1596. doi: 10.1002/aur.1814

Gauducheau, M., Lemaire-Mayo, V., D'Amato, F. R., Oddi, D., Crusio, W. E., and Pietropaolo, S. (2017). Age-specific autistic-like behaviors in heterozygous Fmr1-KO female mice. *Autism Res.* 10, 1067–1078. doi: 10.1002/aur.1743

Grimsley, J. M., Sheth, S., Vallabh, N., Grimsley, C. A., Bhattal, J., Latsko, M., et al. (2016). Contextual modulation of vocal behavior in mouse: newly identified 12 kHz "mid-frequency" vocalization emitted during restraint. *Front. Behav. Neurosci.* 10, 38. doi: 10.3389/fnbeh.2016.00038

Hammerschmidt, K., Radyushkin, K., Ehrenreich, H., and Fischer, J. (2012). The structure and usage of female and male mouse ultrasonic vocalizations reveal only minor differences. *PLoS ONE* 7, e41133. doi: 10.1371/journal.pone.0041133

Hanson, J. L., and Hurley, L. M. (2012). Female presence and estrous state influence mouse ultrasonic courtship vocalizations. *PLoS ONE* 7, e40782. doi: 10.1371/journal.pone.0040782

Hebert, B., Pietropaolo, S., Meme, S., Laudier, B., Laugeray, A., Doisne, N., et al. (2014). Rescue of fragile X syndrome phenotypes in Fmr1 KO mice by a BKCa channel opener molecule. *Orphanet J. Rare Dis.* 9, 124. doi: 10.1186/s13023-014-0124-6

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

Ivanenko, A., Watkins, P., van Gerven, M. A. J., Hammerschmidt, K., and Englitz, B. (2020). Classifying sex and strain from mouse ultrasonic vocalizations using deep learning. *PLoS Comput. Biol.* 16, e1007918. doi: 10.1371/journal.pcbi.1007918

Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K., and Okanoya, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PLoS ONE* 6, e17721. doi: 10.1371/journal.pone.0017721

Kim, H., Son, J., Yoo, H., Kim, H., Oh, J., Han, D., et al. (2016). Effects of the female estrous cycle on the sexual behaviors and ultrasonic vocalizations of male C57BL/6 and autistic BTBR T+ tf/J mice. *Exp. Neurobiol.* 25, 156–162. doi: 10.5607/en.2016.25.4.156

Lahvis, G. P., Alleva, E., and Scattoni, M. L. (2011). Translating mouse vocalizations: prosody and frequency modulation. *Genes Brain Behav*. 10:4–16. doi: 10.1111/j.1601-183X.2010.00603.x

Lefebvre, E., Granon, S., and Chauveau, F. (2020). Social context increases ultrasonic vocalizations during restraint in adult mice. *Anim. Cogn.* 23, 351–359. doi: 10.1007/s10071-019-01338-2

Lemaire-Mayo, V., Subashi, E., Henkous, N., Beracochea, D., and Pietropaolo, S. (2017). Behavioral effects of chronic stress in the Fmr1 mouse model for fragile X syndrome. *Behav. Brain Res.* 320, 128–135. doi: 10.1016/j.bbr.2016.11.051

Maggio, J. C., and Whitney, G. (1985). Ultrasonic vocalizing by adult female mice (*Mus musculus*). *J. Comp. Psychol.* 99, 420–436. doi: 10.1037/0735-7036.99.4.420

Mahrt, E. J., Perkel, D. J., Tong, L., Rubel, E. W., and Portfors, C. V. (2013). Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. *J. Neurosci.* 33, 5573–5583. doi: 10.1523/JNEUROSCI.5054-12.2013

Matsumoto, Y. K., and Okanoya, K. (2018). Mice modulate ultrasonic calling bouts according to sociosexual context. *R. Soc. Open Sci.* 5, 180378. doi: 10.1098/rsos.180378

Moles, A., Costantini, F., Garbugino, L., Zanettini, C., and D'Amato, F. R. (2007). Ultrasonic vocalizations emitted during dyadic interactions in female mice: a possible index of sociability? *Behav. Brain Res.* 182, 223–230. doi: 10.1016/j.bbr.2007.01.020

Moles, A., and D'Amato, F. R. (2000). Ultrasonic vocalization by female mice in the presence of a conspecific carrying food cues. *Anim. Behav.* 60, 689–694. doi: 10.1006/anbe.2000.1504

Neunuebel, J. P., Taylor, A. L., Arthur, B. J., and Egnor, S. E. (2015). Female mice ultrasonically interact with males during courtship displays. *Elife* 4, e06203. doi: 10.7554/eLife.06203.013

Nyby, J., Bigelow, J., Kerchner, M., and Barbehenn, F. (1983). Male mouse (*Mus musculus*) ultrasonic vocalizations to female urine: why is heterosexual experience necessary? *Behav. Neural Biol.* 38, 32–46. doi: 10.1016/S0163-1047(83)90354-0

Nyby, J., Wysocki, C. J., Whitney, G., Dizinno, G., and Schneider, J. (1979). Elicitation of male mouse (*Mus musculus*) ultrasonic vocalizations: I. urinary cues. *J. Comp. Physiol. Psychol.* 93, 957–975. doi: 10.1037/h0077623

Nyby, J. G. (2010). Adult house mouse (*Mus musculus*) ultrasonic calls: hormonal and pheromonal regulation. *Handb. Mammal. Vocal.* 19, 303–310. doi: 10.1016/B978-0-12-374593-4.00029-2

Oddi, D., Subashi, E., Middei, S., Bellocchio, L., Lemaire-Mayo, V., Guzman, M., et al. (2015). Early social enrichment rescues adult behavioral and brain abnormalities in a mouse model of fragile X syndrome. *Neuropsychopharmacology* 40, 1113–1122. doi: 10.1038/npp.2014.291

Pietropaolo, S., Goubran, M. G., Joffre, C., Aubert, A., Lemaire-Mayo, V., Crusio, W. E., et al. (2014). Dietary supplementation of omega-3 fatty acids rescues fragile X phenotypes in Fmr1-Ko mice. *Psychoneuroendocrinology* 49, 119–129. doi: 10.1016/j.psyneuen.2014.07.002

Pietropaolo, S., Guilleminot, A., Martin, B., D'Amato, F. R., and Crusio, W. E. (2011). Genetic-background modulation of core and variable autistic-like symptoms in FMR1 knock-out mice. *PLoS ONE* 6, e17073. doi: 10.1371/journal.pone.0017073

Pietropaolo, S., Sun, Y., Li, R., Brana, C., Feldon, J., and Yee, B. K. (2008). Limited impact of social isolation on Alzheimer-like symptoms in a triple transgenic mouse model. *Behav. Neurosci.* 123, 181. doi: 10.1037/a0013607

Premoli, M., Baggi, D., Bianchetti, M., Gnutti, A., Bondaschi, M., Mastinu, A., et al. (2021a). Automatic classification of mice vocalizations using machine learning techniques and convolutional neural networks. *PLoS ONE* 16, e0244636. doi: 10.1371/journal.pone.0244636

Premoli, M., Bonini, S. A., Mastinu, A., Maccarinelli, G., Aria, F., Paiardi, G., et al. (2019). Specific profile of ultrasonic communication in a mouse model of neurodevelopmental disorders. *Sci. Rep.* 9, 15912. doi: 10.1038/s41598-019-52378-0

Premoli, M., Memo, M., and Bonini, S. A. (2021b). Ultrasonic vocalizations in mice: relevance for ethologic and neurodevelopmental disorders studies. *Neural Regen Res* 16, 1158–1167. doi: 10.4103/1673-5374.300340

Simola, N., and Granon, S. (2019). Ultrasonic vocalizations as a tool in studying emotional states in rodent models of social behavior and brain disease. *Neuropharmacology* 159, 107420. doi: 10.1016/j.neuropharm.2018.11.008

Sugimoto, H., Okabe, S., Kato, M., Koshida, N., Shiroishi, T., Mogi, K., et al. (2011). A role for strain differences in waveforms of ultrasonic vocalizations during male-female interaction. *PLoS ONE* 6, e22093. doi: 10.1371/journal.pone.0022093

von Merten, S., Hoier, S., Pfeifle, C., and Tautz, D. (2014). A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (*Mus musculus* domesticus). *PLoS ONE* 9, e97244. doi: 10.1371/journal.pone.0097244

Whitney, G., Coble, J. R., Stockton, M. D., and Tilson, E. F. (1973). Ultrasonic emissions: do they facilitate courtship of mice. *J. Comp. Physiol. Psychol.* 84, 445–452. doi: 10.1037/h0034899

Wohr, M., Roullet, F. I., Hung, A. Y., Sheng, M., and Crawley, J. N. (2011). Communication impairments in mice lacking Shank1: reduced levels of ultrasonic vocalizations and scent marking behavior. *PLoS ONE* 6, e20631. doi: 10.1371/journal.pone.0020631

Zala, S. M., Reitschmidt, D., Noll, A., Balazs, P., and Penn, D. J. (2017). Sex-dependent modulation of ultrasonic vocalizations in house mice (*Mus musculus* musculus). *PLoS ONE* 12, e0188647. doi: 10.1371/journal.pone.0188647

Zhao, X., Ziobro, P., Pranic, N. M., Chu, S., Rabinovich, S., Chan, W., et al. (2021). Sex- and context-dependent effects of acute isolation on vocal and non-vocal social behaviors in mice. *PLoS ONE* 16, e0255640. doi: 10.1371/journal.pone.0255640

Conflict of Interest: RB is an employee of Metris B.V. and co-developer of the SONOTRACK Call Classification software.

The remaining 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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership