Research Topics

Entrainment and responses to rhythmic stimulation during development

Edited by
Stefanie Peykarjou, Aliette Lochy, Viola Macchi Cassia, Stefanie Hoehl
and Arnaud Leleu

Published in
Frontiers in Psychology
Frontiers in Neuroscience

ISSN 1664-8714
ISBN 978-2-8325-2411-4
DOI 10.3389/978-2-8325-2411-4

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

Entrainment and responses to rhythmic stimulation during development

Topic editors
Stefanie Peykarjou — Heidelberg University, Germany
Aliette Lochy — University of Luxembourg, Luxembourg
Viola Macchi Cassia — University of Milano-Bicocca, Italy
Stefanie Hoehl — University of Vienna, Austria
Arnaud Leleu — Université de Bourgogne, France

Citation
Peykarjou, S., Lochy, A., Cassia, V. M., Hoehl, S., Leleu, A., eds. (2023).
Entrainment and responses to rhythmic stimulation during development.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2411-4

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 28 April 2023
DOI 10.3389/fpsyg.2023.1189054

EDITED AND REVIEWED BY
Peter Klaver,
Interkantonale Hochschule für Heilpädagogik (HfH), Switzerland

*CORRESPONDENCE
Stefanie Peykarjou
✉ stefanie.peykarjou@psychologie.uni-heidelberg.de

RECEIVED 18 March 2023
ACCEPTED 10 April 2023
PUBLISHED 28 April 2023

CITATION
Peykarjou S, Hoehl S, Leleu A, Lochy A and Macchi Cassia V (2023) Editorial: Entrainment and responses to rhythmic stimulation during development. *Front. Psychol.* 14:1189054.
doi: 10.3389/fpsyg.2023.1189054

Editorial: Entrainment and responses to rhythmic stimulation during development

Stefanie Peykarjou[1]*, Stefanie Hoehl[2], Arnaud Leleu[3], Aliette Lochy[4] and Viola Macchi Cassia[5,6]

[1]Department of Psychology, Heidelberg University, Heidelberg, Germany, [2]Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria, [3]Development of Olfactory Communication and Cognition Laboratory, Centre des Sciences du Goût et de l'Alimentation, Université de Bourgogne, CNRS, INRAe, Institut Agro, Dijon, France, [4]Institute of Cognitive Science and Assessment (COSA), Department of Behavioral and Cognitive Sciences (DBCS), Faculty of Humanities, Social and Educational Sciences (FHSE), University of Luxembourg, Esch-sur-Alzette, Luxembourg, [5]Department of Psychology, University of Milano-Bicocca, Milan, Italy, [6]NeuroMI, Milan Center for Neuroscience, Milan, Italy

KEYWORDS

FPVS, frequency tagging, oscillations, SSVEPs, synchronization, development

Editorial on the Research Topic
Entrainment and responses to rhythmic stimulation during development

Rhythms are omnipresent in physiological processes, sensory inputs and social interactions. Humans respond to external rhythms with synchronization, which can be observed across multiple levels (e.g., behavior, brain, and physiology). Given their ubiquitous nature, utilizing rhythmic stimulation to characterize development is beneficial in many domains, including cognition, perception, and motor functioning.

This Research Topic brings together current research on behavioral and brain responses to rhythmic stimulation in development across these domains. This ultimately helps to broaden our understanding of neural entrainment (e.g., Bánki et al.), of the links between rhythms and cognitive or behavioral development (e.g., Frischen et al.; Yu et al.), and of the methods employing rhythmic stimulation to investigate early development (e.g., Peykarjou). Since rhythmic responses can be conceived as entrainment of endogenous oscillators to external rhythms (e.g., Notbohm et al., 2016), or periodic evoked responses (Capilla et al., 2011), "entrainment" is broadly defined here as responses elicited by a corresponding external driving rhythm.

Several contributions in this Research Topic focus on theoretical advancements regarding relevant concepts and relations of rhythms with development (Bánki et al.; Bowsher-Murray et al.; Frischen et al.). Bánki et al. provide a discussion of neural entrainment in contrast to stimulus tracking and argue that evidence in favor of entrainment is accumulating. However, this evidence largely comes from research with adults, and establishing the presence of neural entrainment in early development requires adapting current methods. The authors call for studies aligning the stimulation frequency to individual internal oscillations of participants and investigating how external rhythms modulate behavioral performance.

The relation between rhythm processing, cognitive and motor abilities is analyzed by Frischen et al. who discuss predictive processing (Friston, 2005) as the mechanism that might link these functions. Studies showing substantial correlations between rhythm processing and language ability, motor skills and executive functions are summarized. Predictive processing in beat perception, anticipating temporally ordered sound events, is suggested to stimulate prediction in other domains, providing a link explaining commonalities in development.

Lau et al. provide support for the view that early links between language (i.e. vocalizations) and cognition may be subserved by infants' perceptual sensitivity to rhythmic elements. Drawing on evidence from four-month-olds showing that language may facilitate categorization (Ferry et al., 2013), the authors employ supervised machine learning to show that rhythm-related features of acoustic stimuli and temporal envelope components may be vital for supporting cognition.

In their contribution, Bowsher-Murray et al. approach the role of rhythm in development by focusing on interpersonal synchrony, i.e., the "tendency for social partners to temporally coordinate their behavior when interacting". They provide an overview of components underpinning successful non-verbal interpersonal synchrony, including social orienting, action prediction and motor behavior, monitoring and error detection, pointing out that more work on the potentially bidirectional nature of behavioral and neural entrainment is required. Indeed, spanning these levels of analysis was accomplished across contributions in this Research Topic but should be tackled within studies in the future.

Behavioral entrainment is addressed in three studies investigating the development of rattle-shaking and drumming (Laudanska et al.; Rocha and Addyman; Yu et al.). The ability to coordinate drumming movements to match external rhythms was investigated in toddlers between 18 and 30 months of age (Rocha and Addyman; Yu et al.). The beginnings of rhythmic behavior are explored by Laudanska et al., who investigate coupling between infants' arm movements longitudinally from four to twelve months. They document infants' motivation for rhythmic behavior from early on, as participants similarly attempt rattle-shaking across all ages. With age, left and right arm movements become more coupled during rattle-shaking independent of age, demonstrating how synchronization diffuses throughout the body. Yu et al. demonstrate that by 24 months, children begin to slow down their spontaneous motor tempo toward the rhythm provided by a drumming partner, but successful synchronization is only achieved by 30 months. Rocha and Addyman employ a novel tool, deep learning-based OpenPose, to automatically estimate drumming behavior from home-recorded videos. Both studies show that toddlers between 24 and 30 months of age are strongly motivated to engage in synchronous movement, as they adapt their drumming tempo away from their intrinsic rate of movement to match either a social interaction partner, a robot (Yu et al.), or the video of a drumming hand (Rocha and Addyman).

The research by Frey et al. applies rhythmic stimulation as a tool for facilitating development. In a training study with preschool children, they demonstrate correlations between children‘s rhythmic and literacy skills and between rhythm synchronization and pen pressure. Training provided in their study effectively improves rhythmic abilities, however these improvements do not transfer to literacy or graphomotor development. Further research is needed to evaluate whether rhythmic training may causally support development in other domains.

Other studies assembled in this Research Topic employ responses to rhythmic stimulation coupled with neural measures to examine auditory and visual processing (Attaheri et al.; Bertels et al.; Cantiani et al.; Pescuma et al.; Peykarjou; Poncet et al.). In these studies, rhythmic stimulation is mainly employed as a tool for exploring perceptual and cognitive development without necessitating a behavioral response.

Auditory processing is investigated by Cantiani et al. who employ frequency tagging with complex music and speech stimuli. The authors report similar neural entrainment to auditory stimuli, particularly music, in eight-month-old infants and adults. They infer that neural synchronization and tempo flexibility develop before synchronization abilities become apparent in behavior. Attaheri et al. also compare neural synchronization to speech stimuli between adults and infants (four to eleven months) and find commonalities across development. Participants track sung speech in delta and theta bands, but specific peak-frequencies differ between the two populations. Together, these studies suggest that infant and adult brains utilize similar cortical mechanisms to track linguistic inputs, but the interplay between these mechanisms varies with age and language experience.

Four studies employ the Fast Periodic Visual Stimulation (FPVS) oddball paradigm (Liu-Shuang et al., 2014) with visual stimuli. Infants' developing ability to discriminate facial expressions from 3.5 to 7 months is investigated by Poncet et al.. Significant expression discrimination is observed from 3.5 months, with partly diverging topographies across facial expressions and age. Responses for emotional faces among neutral ones are recorded at right occipito-temporal sites, similar to adults (Poncet et al., 2019), suggesting that the discrimination of neutral vs. expressive faces is subserved by similar networks across development.

Bertels et al. investigate characteristics of a potentially inborn mechanism to detect snakes, a species posing an evolutionarily important threat to humans. EEG responses in six-to-eleven-month-olds reveal that they detect snakes among other animals, and this effect is not influenced by color. However, responses increase with infant age, which is interpreted as a specific visual development for evolutionary-relevant shapes in complex backgrounds.

Peykarjou presents a systematic literature review on frequency tagging EEG employing the FPVS oddball paradigm and examines the effects of analysis decisions in a large sample of infants. Based on the assembled evidence, recommendations regarding duration, number of conditions, sequence- and participant-retention are made. Overall, evidence indicates that analysis decisions should be tailored to the age-group.

Pescuma et al. use frequency tagging to investigate a developing cultural skill, morpheme identification. They adapt an established paradigm (Lochy et al., 2016a,b) to MEG and demonstrate morpheme identification in developing readers that is overall similar to adults. Neural responses are primarily present for word stems in children and word suffixes in adults, suggesting an influence of accumulated reading experience.

This Research Topic assembles a variety of findings demonstrating the importance of rhythms for development, and their usefulness for elucidating cognitive, visual, auditory, and motor functions. Moreover, it provides readers with an overview of methodological advancements for infant studies, such as wearable motion trackers (Laudanska et al.), an online version of a measure for drumming synchrony (Rocha and Addyman), or guidelines for frequency tagging EEG categorization tasks (Peykarjou). By integrating findings across the different contributions, several conclusions can be drawn. First, rhythms provide both a window into development across different domains and a linking concept for understanding the interplay of development and environmental influences. Second, across developmental domains and levels of analysis, rhythms play an important role that we only begin to understand. Third, at both conceptual and empirical levels, further work is required to leverage the potential of rhythmic stimulation and entrainment for developing a more comprehensive picture of how behavior and brain processing develop across domains.

Author contributions

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

Funding

This work was supported by a grant from the German Research Foundation (DFG, grant number: PE 3125/1-1) to SP, two FWF grants (grant numbers: P 33853 and I 4332) awarded to SH, the French National Research Agency (grant number: ANR-19-CE28-0009) awarded to ALe, and the Luxembourg National Research Fund (FNR grant number: C21/SC/16241557) to ALo.

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

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS ONE* 6, e14543. doi: 10.1371/journal.pone.0014543

Ferry, A. L., Hespos, S. J., and Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proc. Nat. Acad. Sci.* 110, 15231–15235. doi: 10.1073/pnas.1221166110

Friston, K. (2005). A theory of cortical responses. *Philosoph. Transact. Royal Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Liu-Shuang, J., Norcia, A. M., and Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* 52, 57–72. doi: 10.1016/j.neuropsychologia.10022

Lochy, A., Van Belle, G., and Rossion, B. (2016a). A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. *Neuropsychologia* 66, 18–31. doi: 10.1016/j.neuropsychologia.11007

Lochy, A., Van Reybroeck, M., and Rossion, B. (2016b). Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proc. Nat. Acad. Sci.* 113, 8544–8549. doi: 10.1073/pnas.1520366113

Notbohm, A., Kurths, J., and Herrmann, C. S. (2016). Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. *Front. Hum. Neurosci.* 3, 10. doi: 10.3389/fnhum.2016.00010

Poncet, F., Baudouin, J.-Y., Dzhelyova, M. P., Rossion, B., and Leleu, A. (2019). Rapid and automatic discrimination between facial expressions in the human brain. *Neuropsychologia* 129, 47–55. doi: 10.1016/j.neuropsychologia.2019.03.006

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fnins.2022.842447

Cortical Tracking of Sung Speech in Adults vs Infants: A Developmental Analysis

Adam Attaheri[1]*, Dimitris Panayiotou[1], Alessia Phillips[1], Áine Ní Choisdealbha[1], Giovanni M. Di Liberto[2,3], Sinead Rocha[1], Perrine Brusini[1,4], Natasha Mead[1], Sheila Flanagan[1], Helen Olawole-Scott[1] and Usha Goswami[1]

[1] Department of Psychology, Centre for Neuroscience in Education, University of Cambridge, Cambridge, United Kingdom, [2] School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland, [3] Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, Ecole Normale Supérieure, PSL Research University, Paris, France, [4] Institute of Population Health, University of Liverpool, Liverpool, United Kingdom

Here we duplicate a neural tracking paradigm, previously published with infants (aged 4 to 11 months), with adult participants, in order to explore potential developmental similarities and differences in entrainment. Adults listened and watched passively as nursery rhymes were sung or chanted in infant-directed speech. Whole-head EEG (128 channels) was recorded, and cortical tracking of the sung speech in the delta (0.5–4 Hz), theta (4–8 Hz) and alpha (8–12 Hz) frequency bands was computed using linear decoders (multivariate Temporal Response Function models, mTRFs). Phase-amplitude coupling (PAC) was also computed to assess whether delta and theta phases temporally organize higher-frequency amplitudes for adults in the same pattern as found in the infant brain. Similar to previous infant participants, the adults showed significant cortical tracking of the sung speech in both delta and theta bands. However, the frequencies associated with peaks in stimulus-induced spectral power (PSD) in the two populations were different. PAC was also different in the adults compared to the infants. PAC was stronger for theta- versus delta- driven coupling in adults but was equal for delta- versus theta-driven coupling in infants. Adults also showed a stimulus-induced increase in low alpha power that was absent in infants. This may suggest adult recruitment of other cognitive processes, possibly related to comprehension or attention. The comparative data suggest that while infant and adult brains utilize essentially the same cortical mechanisms to track linguistic input, the operation of and interplay between these mechanisms may change with age and language experience.

Keywords: EEG, language, neural oscillations, TRF, cortical tracking

Edited by:
Stefanie Hoehl,
University of Vienna, Austria

Reviewed by:
Peter Cariani,
Boston University, United States
Mireille Besson,
UMR7291 Laboratoire de Neurosciences Cognitives (LNC), France

Correspondence:
Adam Attaheri
aa974@cam.ac.uk

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *23 December 2021*
Accepted: *23 February 2022*
Published: *12 April 2022*

Citation:
Attaheri A, Panayiotou D, Phillips A, Ní Choisdealbha Á, Di Liberto GM, Rocha S, Brusini P, Mead N, Flanagan S, Olawole-Scott H and Goswami U (2022) Cortical Tracking of Sung Speech in Adults vs Infants: A Developmental Analysis. Front. Neurosci. 16:842447. doi: 10.3389/fnins.2022.842447

INTRODUCTION

The amplitude envelope of speech carries crucial low-frequency acoustic information that assists linguistic decoding at multiple time scales. According to "multi-time resolution" models of speech perception, linguistic decoding is achieved in part via the neural tracking of different temporal modulation patterns in speech at different timescales simultaneously (Poeppel, 2003; Greenberg, 2006; Luo and Poeppel, 2007; Ghitza and Greenberg, 2009; Chait et al., 2015). One core mechanistic proposal is that cortical oscillations entrain or phase-align their activity to modulations at corresponding timescales in the signal, thereby encoding the different energy patterns, and binding the information together into the final speech percept (Ghitza, 2011, 2012; Giraud and Poeppel, 2012; Poeppel, 2014). Neurophysiological signals are known to preferentially track the amplitude envelope of adult-directed speech (ADS) in the theta band, approximately 4–8 Hz (Luo and Poeppel, 2007; Ghitza, 2011; Ding and Simon, 2013, 2014; Gross et al., 2013; Zion Golumbic et al., 2013; Doelling et al., 2014; Di Liberto et al., 2015; Kösem and van Wassenhove, 2017). However, recent analyses of neural tracking of the speech amplitude envelope by infants shows preferential tracking in the delta band [∼0.5–4 Hz, see Attaheri et al. (2022)]. This developmental difference (infants primarily engage delta, adults primarily engage theta) may reflect the nature of the stimulus, as acoustic analyses of infant-directed speech (IDS) using a spectral-amplitude modulation phase hierarchy approach (S-AMPH, see Leong et al., 2017) reveal significantly greater modulation energy compared to ADS in an amplitude-modulation (AM) band centered on ∼2 Hz. Attaheri et al. (2022) were the first to compare infant cortical tracking in the delta and theta bands, as prior infant cortical tracking studies relied on the broadband speech envelope (Kalashnikova et al., 2018; Jessen et al., 2019). Accordingly, it could be argued that cortical tracking of IDS by delta-band neural signals may be key to initial language acquisition.

An outstanding question is whether the preferential delta-band tracking observed with infants (Attaheri et al., 2022) was driven by the IDS stimulus materials, and accordingly whether the adult brain would show similar preferential delta-band tracking to the same speech input. Note that delta-band tracking of speech input was also observed in the adult brain in studies that did not use IDS (Molinaro et al., 2016; sentences; Di Liberto et al., 2015, 2018; story listening, Kösem and van Wassenhove, 2017, for review). In order to ensure infant attention and engagement, Attaheri et al. (2022) used videos of a female singing nursery rhymes in a deliberately infant-directed manner (IDS), recording neural responses to this audio-visual input using 64-channel EEG. Nursery rhymes are inherently rhythmic, and neural responses in musical tasks are also typically strong in the delta band (Cirelli et al., 2016). Thus, the preferential delta-band tracking observed in the infants studied by Attaheri et al. (2022) might either reflect the use of IDS, or the fact that the speech input was sung, or both. Adult entrainment data may help to disentangle some of these possibilities, as the adult brain may respond differently to IDS compared to the infant brain.

Accordingly, here we report a replication of Attaheri et al. (2022) with adults. Identical stimuli, identical experimental procedures and identical analysis pipelines (bar removal of adult ocular artifacts not present in the infant data; see section "Materials and Methods" for full details) were adopted to allow a faithful comparison of neural tracking of sung speech in the infant versus fully-matured adult brain. Speech also contains acoustic information within its higher-frequency bands (beta, gamma) that is linguistically important. Low-frequency (delta, theta) phase and high-frequency (gamma) amplitudes track acoustic rhythms in the adult brain by operating together as an integrated representational mechanism, called phase-amplitude coupling, PAC (Gross et al., 2013; Hyafil et al., 2015; Lizarazu et al., 2019). In the adult brain, strongest PAC is observed for theta-gamma coupling (Gross et al., 2013). Attaheri et al. (2022) demonstrated PAC in infants at all ages studied (4, 7, and 11 months). However, they reported that delta acted as an equally strong carrier phase for higher-frequency amplitudes (beta and gamma) as theta phase. For infants, therefore, both delta and theta play important roles in the temporal organization of higher-frequency amplitudes during speech processing. Studying PAC to the same sung speech in adults enables clarification of whether the PAC previously observed with infants is stimulus-driven or development-driven.

On a stimulus-driven hypothesis, delta-based PAC should be as strong as theta-based coupling for adults when sung speech/IDS is the input. Alternatively, PAC in the adult brain may always favor theta-gamma organization. Also of interest was whether low-frequency phases (delta and theta) would act as equally strong carriers for both beta and gamma band information, as found for infants, or whether coupling to beta would be less important for adults (Hyafil, 2017). Attaheri et al. (2022) reported that whilst both gamma and beta high-frequency amplitudes coupled to delta and theta phases, it was amplitudes in the gamma band that produced the stronger coupling to both delta and theta phases in the infants. This effect did not change between 4 and 11 months of age. Again, it is currently unclear whether this is a developmental or a stimulus-driven phenomenon.

Theoretically, one key difference between adults and infants, when listening to the same speech input, is the accumulated years of prior experience with speech processing enjoyed by the adults. Infants are universal novices, and even by 11 months, are unlikely to yet comprehend the different nursery rhymes that they are hearing. They also cannot yet produce recognizable speech. In adult neural speech processing studies, delta-band tracking is frequently related to discourse-level parsing related to phrasing (Ding et al., 2016; Kösem and van Wassenhove, 2017), as well as to auditory attentional mechanisms linked to the automatic grouping of sounds (Boucher et al., 2019). While the pre-verbal infant brain may also utilize acoustic grouping mechanisms, which could be stimulus-driven, infants are unlikely to have learned discourse-level speech information by 11 months. Similarly, in adult studies theta band cortical tracking has been related to syllable parsing and speech intelligibility (Luo and Poeppel, 2007; Peelle et al., 2013; Millman et al., 2015; Baltzell et al., 2017). While there may be stimulus-driven cues for syllable parsing that infants can utilize automatically (such as the acoustic edges linked to syllable rise times, see Doelling et al., 2014; Lizarazu et al., 2021), speech intelligibility is unlikely to be a key factor in cortical tracking in infancy, as infants are still learning spoken language. Indeed, animal studies have also demonstrated cortical tracking of rhythmically-structured acoustic input in both delta and theta bands, possibly suggesting that these mechanisms are triggered by general auditory perceptual abilities conserved across mammalian species (Lakatos et al., 2005, 2016). The current study should throw light on these important developmental questions.

MATERIALS AND METHODS

The materials and methods used in the current analysis replicated those used in a prior study with infants (Attaheri et al., 2022). Any subtle adaptions required in adapting the experiment for adult participants are outlined below.

Participants

Total of 22 monolingual, English-speaking, participants (11M, 11F, aged 18–30, Mean age 21 years) were recruited from central Cambridge (United Kingdom) and surrounding areas. The study was reviewed by the Psychology Research Ethics Committee of the University of Cambridge and after a detailed explanation of the study, written consent was given by each participant. Each participant reported no history of language difficulties or dyslexia. One participant's data was excluded due to a technical error (no stimulus triggers were recorded) leaving 21 participants data remaining for analysis.

Stimuli

A selection of 18 typical English language nursery rhymes were chosen as the stimuli. Audio-visual stimuli of a singing head were recorded using a Canon XA20 video camera at 1,080p, 50fps and with audio at 4,800 Hz. A native female speaker of British English used infant directed speech to melodically sing (for example "Mary Mary Quite Contrary") or rhythmically chant (for nursery rhymes like "There was an old woman who lived in a shoe") the nursery rhymes whilst listening to a 120 BPM metronome. The beat was not present on the stimuli presented to the participants, but it ensured that a consistent quasi-rhythmic production was maintained throughout the 18 nursery rhymes. During recording of the video, the adult was singing to a real infant, with whom she was sharing mutual gaze.

EEG Data Collection

Participants were seated ∼650 mm away from the presentation screen within a sound-proof acoustic chamber. EEG data was recorded at a sampling rate of 1,000 Hz using a GES 300 amplifier connected to a correctly sized 128 channel electrode net (Geodesic Sensor Net, Electrical Geodesics Inc., Eugene, OR, United States). The sounds were presented at 60 dB (dBA, checked by a handheld sound level meter) from speakers (Q acoustics 2020i driven by a Cambridge Audio Topaz AM5 Stereo amplifier) placed either side of the screen. Participants were asked to attend to the screen whilst 18 nursery rhyme videos played sequentially, each repeated three times (54 videos, with a presentation time of 20′ 33″ in total). All participants included in analysis completed the full experiment. The stimulus period was followed by a 5 min resting state recording, in which the participants were asked to sit silently with their eyes open, whilst no sound or visual stimuli were present.

EEG Preprocessing

All analyses were conducted with custom-made scripts in Matlab 2017a (The MathWorks, Inc., Natick, MA, United States) incorporating the EEGLab toolbox (Delorme and Makeig, 2004). The analysis protocols were kept as consistent as possible to the previous infant analysis pipelines to allow faithful comparison to the previously reported infant results.

The EEG data, from the 128 channels, was first filtered (*pop_eegfiltnew* function of EEGLab toolbox) into a broadband signal (0.5–45 Hz) using zero-phase bandpass Hamming windowed FIR filters (transition band widths of 2 Hz with cutoff frequencies at −6 dB). The EEG data was down sampled to 100 Hz to reduce the computational load. Next, the *clean_asr* EEGLab function (Delorme and Makeig, 2004) was used to clean noise artifacts from the data by identifying and removing bad principal components *via* a modified PCA procedure (see **Supplementary Material** for more details). Further bad channels were identified *via* probability and kurtosis and were interpolated (*via* spherical interpolation), if they were 3 standard deviations away from the mean, before all channels were re-referenced to a 128-channel average reference. ICA (*runica*; EEGLab) was conducted to detect components containing ocular and ECG artifacts, which were visually identified and then removed from the data. Frequency bands of interest (0.5–4 Hz, 4–8 Hz or 8–12 Hz for the mTRF analysis) were acquired using a using zero-phase bandpass Hamming windowed FIR filters (transition band widths of 2 Hz with cutoff frequencies at −6 dB, 0–5 Hz, 3–9 Hz and 7–13 Hz, respectively).

EEG responses were epoched into trials aligned to the start and ending at the completion of a phrase (e.g., "Mary had a little lamb"), producing EEG responses to 83 phrases (*M length* ± *SD*: 4.23 s ± 0.88) which were repeated a maximum of 3 times in the experiment (249 epochs in total). This epoching procedure was used to keep consistency with the previous infant EEG study. To retain epochs where a single channel exhibited noise epoch by epoch channel interpolations were conducted. Per epoch, probability and kurtosis were used to identify bad channels and were interpolated (*via* spherical interpolation) if they were 3*SD* away from the mean.

Multivariate Temporal Response Function

TRFs are encoding models that can describe how an input and output of a system are related *via* linear convolution (Crosse et al., 2016). Here, we applied TRFs in a backward direction to assess how strongly a stimulus property, in this case the stimulus envelope, is encoded in the neural response. We chose backward TRF modeling as it uses information from all EEG channels at once to reconstruct the speech envelope, giving a low weighting to irrelevant channels whilst allowing the model to capture additional variance across channels (Crosse et al., 2016). The result is a single objective metric (i.e., the envelope decoding correlation). Backward TRF modeling has the advantage of producing larger correlation scores compared to forward TRF modeling, making it a good choice for analyzing the original infant EEG data, which is inherently noisy (Jessen et al., 2021).

After preprocessing, the epochs of EEG data in response to each nursery rhyme trial were averaged together to improve the signal to noise ratio of the data for the mTRF analysis (matching the infant analysis procedure). The mTRF analysis was conducted using the multivariate temporal response function (mTRF) toolbox v1.5 (Crosse et al., 2016) through Matlab 2017a (The MathWorks, Inc., Natick, MA, United States). The backward model can be expressed by the following formula in which the reconstructed stimulus envelope *s(t)* is created by a linear decoder, g(τ, n*)*, mapping the neural response, *r(t,n)*, back to the stimulus, *s(t)*. The TRF decoder was used to integrate the neural responses at multiple time lags, τ, between 0 and 250 ms (τ_{min} = 0 ms, τ_{max} = 250 ms). These "stimulus-relevant" time lags where selected in keeping with the previous literature (Ding and Simon, 2014; Di Liberto et al., 2015; Crosse et al., 2016).

$$\widehat{s}(t) = \sum_{n} \sum_{\tau} r(t + \tau, n)\, g(\tau, n),$$

The quality of the envelope tracking within each EEG frequency band was assessed by a "leave-one-out" cross-validation per participant. First the average trial EEG epochs (maximum of 83) were normalized *via* function *nt_normcol* (Noisetools[1]). Normalization, decreased the range of values that were necessary for the regularization parameter search in the mTRF toolbox, making the cross validation more efficient. Next, the normalized epoch trials were rotated M-1 times, each serving once as the "test set" with the remainder of the trials being the TRF "training set." For each rotation, the resultant M-1 training models were averaged to create one average model from the training set. The average model was subsequently convolved with the test data to reconstruct the stimulus. Pearson's correlation (r) was used to validate how well the reconstructed stimulus correlated to the original. This process was repeated for the M-1 rotations. To avoid overfitting the model to a specific trial, an average *r* value was taken from the 83 r validation values. This process was repeated at 12 ridge regressions (λ values, 1×10^{-3}:1×10^{8}) with the lowest λ value, where any increase gave no further improvement to the average *r* value, was taken (Crosse et al., 2016). Choosing the correct lambda value here again mitigated the potential overfitting of the TRF model. This average *r* value, at the optimal λ, was used for all further analysis.

mTRF Auditory Stimuli Preprocessing

The envelope of the auditory signal was extracted by taking the absolute value of the analytic signal generated by the Hilbert transform (Matlab). As the envelope of the lower frequencies is linearly relatable to the EEG signal (Pasley et al., 2012; Zion Golumbic et al., 2013; O'Sullivan et al., 2015) the envelope of the stimuli was filtered between 0.5 and 15 Hz (lowpass; 6th order Butterworth filter. Highpass; 9th order Butterworth filter). The resultant envelopes were normalized using *nt_normcol* (NoiseTools[1]). Finally, the stimulus envelopes were down-sampled to 100 Hz to match the EEG signal.

mTRF Random Permutation Statistics

Random permutation statistics were created for each participant to measure the average stimulus reconstruction (r) that could be obtained by chance. The random permutation procedure was conducted per participant for each frequency band producing a paired chance stimulus reconstruction (r). To obtain a random permutation of the data, whilst maintaining phase integrity, each of the stimulus envelopes were first reversed and a random circular shift was applied. Next, the mTRF cross-validation was ran in the same way as the real data (see above for details), to give a stimulus reconstruction (r) value. This procedure was iterated 100 times to create a null distribution and the average of these 100 iterations were used as that participant's random stimulus reconstruction (r) value.

Spectral Analysis (Periodogram Power Spectral Density Estimate)

All remaining epochs after preprocessing were concatenated back into one continual EEG signal. A one-sided PSD estimate was conducted separately for each electrode channel using the periodogram function (Matlab). The length of the participants data was zero padded to ensure the size of the rectangular window used was equal in length to the number of discrete Fourier transform (DFT) points, ensuring the correct FFT across participants. This resulted in 52,834 equal spaced frequency bins from 0 to 50 Hz.

The periodogram can be defined by the following formula. In which the EEG signal, x_n, is sampled at 100Hz, with Δt as the sampling interval.

$$\hat{P}(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-j2\pi f\ \Delta t\ n} \right|^2, -1/2\Delta t < f \leq 1/2\Delta t,$$

To achieve the one-sided periodogram output reported in **Figure 1**, values at all frequencies (except 0 and the Nyquist, 1/2Δt), were multiplied by two to conserve the total power.

Phase Amplitude Coupling

The same concatenated data sets created for the PSD analysis were also used for the Phase Amplitude Coupling (PAC) analysis. A modified version of the WinPACT plugin (EEGLab) (Delorme and Makeig, 2004) was used to acquire normalized modulation index (nMI) values (Özkurt and Schnitzler, 2011), a measure adapted from Canolty et al.'s (2006) modulation index (MI) (Canolty et al., 2006). The normalized version of the MI calculation (nMI) was selected as differences in low-frequency power have been shown to adversely affect the PAC calculation (Canolty et al., 2006; Özkurt and Schnitzler, 2011; Aru et al., 2015). The MI method combines the amplitude envelope time series A1(t + τ) of a high-frequency with the phase time series φ2(t) of a specified low-frequency, creating a composite complex-valued signal z(t, τ). The resulting value is a widely validated metric of the coupling strength and preferred phase between two frequencies. For each participant's data, low-frequency phase (LFP) and high-frequency amplitude (HFA) were extracted with a zero-phase FIR filter (Hamming window), separately for all 128 electrodes. LFP center frequencies were filtered from 2 to 8 Hz, in 1 Hz steps with a 2 Hz bandwidth, and HFA center frequencies were filtered from 17.5 to 42.5 Hz, in 5 Hz steps with a 5 Hz bandwidth. A sliding 5 s analysis window was implemented,

---

[1] http://audition.ens.fr/adc/NoiseTools/

with 2.5 s overlaps, with a mean vector length calculated per window. Next, 200 surrogate statistical iterations were created for each PAC calculation window. The statistically normalized MI estimate was obtained for each analysis window by subtracting the mean and dividing by the standard deviation obtained from a Gaussian fit of surrogate MI estimates (nMI = (Canolty's MI—surrogate MI Mean) / surrogate MI Std). This statistical procedure was first suggested by Canolty et al. (2006) and implemented in the winPACT plugin based on code adapted from Özkurt and Schnitzler (2011). Each iteration of the surrogate data was created by shuffling the high-frequency amplitude time series *via* circular rotation. A nMI estimate was obtained for each of the 200 surrogate data iterations, from which a 95% confidence interval was calculated using normcdf.m. This step accounted for the mean and standard deviation of the surrogate data set, thus creating an appropriate threshold for the frequency band analyzed (see winPACT_precompute.m subscript in the winPACT toolbox for complete code, implemented in our analysis script). Finally, generalized family-wise error rate correction was implemented to correct for the multiple PAC calculation windows. The remaining statistically significant nMI windows were averaged per channel for each of the PAC pairs (i.e., each LFP and HFA step) separately for each participant. The frequency bands of interest were defined as follows, delta 2–4 Hz, theta 4–8 Hz, beta 15–30 Hz and gamma 30–45 Hz). The channel exhibiting the strongest nMI, within predefined phase and amplitude band groupings (delta/beta, delta/gamma, theta/beta, theta/gamma), was taken forward for the LMEM and for the group level grand average plots.

RESULTS

The detailed analyses outlined in the Methods were aimed firstly at identifying peaks where the nursery rhymes induced increases in EEG spectral power. The second step was to investigate whether the observed oscillations in the lower frequency bands tracked the envelope of speech. Finally, we were interested to see to what extent the phase of these low frequency (delta and theta) oscillations coupled to the amplitudes of the higher frequency oscillations (beta and gamma).

Power Spectral Density Response to Resting State and Stimulus Periods

The distribution of low-frequency neural signals within our data was established using spectral decomposition of the signal, achieved using the periodogram power spectral density (PSD) estimate (**Figure 1**). After preprocessing, PSD was obtained for each of the remaining 128 electrodes in response to audio-visually presented nursery rhymes and during a 5-min period of silence. A grand average across both conditions (stimulus and resting state) and all channels revealed three prominent frequency peaks centered around ∼1.25, ∼8.54, and ∼10.04 Hz (**Supplementary Figure 1**). The peak visible in **Figure 1** at 0.5 Hz was not considered for analysis as its location at the broadband filter boundary means we are unable to discount its occurrence due to a filtering artifact. Due to the prevalence of low-frequency noise in some of the recording sessions, outlier analysis (*isoutlier "quartiles"* function, Matlab) was conducted to remove extreme data points. Four participant's data points were identified as outliers and removed from the PSD analysis leaving 17 PSD analysis data sets.

To establish whether the PSD peaks in the nursery rhyme stimulus period (NR) were significantly different to the resting state (RS) we used a repeated measures ANOVA. The ANOVA had 2 levels, a level of "condition" (NR versus RS) and a level of "band" (∼1.25, ∼8.54, and ∼10.04 Hz). The band level was included to investigate whether any RS-NR difference was significantly higher throughout all the observed peaks or not. The dependent variable was the maximum PSD value, averaged across channels, taken per participant from a 1 Hz window centered around 1.25, 8.54, and 10.04 Hz, for both the stimulus period and the resting state. Greenhouse-Geisser corrected results showed statistically significant main effects of condition [$F(1, 16) = 24.372$, $p = 1.49 \times 10^{-4}$], and band; [$F(1.140,18.239) = 17.046$, $p = 4.2 \times 10^{-4}$] and a significant interaction between condition and peak of the PSD values [$F(1.112, 17.797) = 18.308$, $p = 3.336 \times 10^{-4}$]. Due to the significant interaction, *post hoc*, Bonferroni corrected, simple main effects analysis was conducted and showed that the stimulus induced PSD was significantly larger than the corresponding peaks in the resting state at 1.25 Hz [$F(1, 16) = 22.427$, $p = 2.24 \times 10^{-4}$; mean ± SEM, RS = 4.454 ± 0.952, NR = 15.990 ± 3.123] and 8.54 Hz [$F(1, 16) = 7.791$, $p = 0.013$; mean ± SEM, RS = 1.671 ± 0.351, NR = 2.984 ± 0.482] but not at 10.04 Hz [$F(1, 16) = 0.904$, $p = 0.356$; mean ± SEM, RS = 3.081 ± 0.870, NR = 2.364 ± 0.273].

In summary, the data show stimulus-induced PSD at ∼1.25 and ∼8.54 Hz. A further peak was observed at 10.04 Hz, however, this was not significantly different from resting state.

Power Spectral Density Comparison to Infant Experiment

No peaks in the PSD spectrum were observed at the corresponding peak frequencies reported in the infant analysis (2.20 and 4.37 Hz). This may imply that the infant neural response was more stimulus-driven, as there were clear modulation peaks in the modulation spectrum of the nursery rhymes at these two frequencies (please see **Supplementary Figures 2–4**). To compare whether the nursery rhyme stimuli induced power increases in the same regions in the adult EEG data as in the infant EEG data, a two-way repeated measures ANOVA was conducted using the infant PSD peak values as the dependent variable. The maximum PSD value was taken per participant in a 1 Hz window centered around 2.20Hz and 4.37Hz for both the stimulus period and the resting state (band × condition). A statistically significant main effect of condition [$F(1, 16) = 41.601$, $p = 8.0 \times 10^{-6}$], band; [$F(1,16) = 24.498$, $p = 1.45 \times 10^{-4}$] and a significant interaction between condition and peak of the PSD values [$F(1, 16) = 17.230$, $p = 7.52 \times 10^{-4}$] was observed. Due to the significant interaction, *post hoc*, Bonferroni corrected, simple main effects analysis was conducted. This showed that the nursery rhyme stimuli induced a PSD increase from resting state at both ∼2.20Hz [$F(1, 16) = 29.453$, $p = 5.60 \times 10^{-5}$; mean ± SEM,

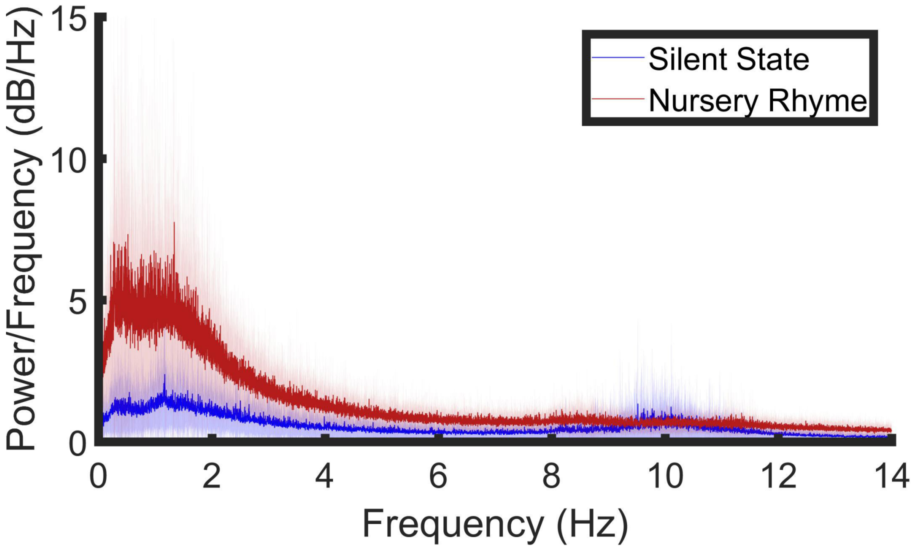


FIGURE 1 | Spectral decomposition of the EEG signal (0.5–14 Hz) in response to nursery rhyme stimulation. A periodogram was used to obtain a power spectral density (PSD) estimate separately for the resting state (blue line) and the nursery rhyme stimulus (red line) periods. Bold lines indicate the mean values and pale shading plots the standard deviation of the data. Outlier analysis was also conducted to remove extreme data points leaving, resting state, $N = 17$; Nursery Rhyme stimulus period $N = 17$.

RS = 3.241 ± 0.696, NR = 10.839 ± 1.771) and ~4.37 Hz [$F(1, 16) = 94.241$, $p = 4.14 \times 10^{-8}$; mean ± SEM, RS = 1.636 ± 0.410, NR = 3.741 ± 0.495] in the adult data. Accordingly, although no visible peaks in activity were observed (see **Figure 1**), significant stimulus-related PSD power was present in the adult data at the matched infant frequencies.

Delta and Theta EEG Frequency Bands Track Nursery Rhyme Envelopes

To investigate the presence and strength of cortical tracking, backward mTRFs (**Figure 2A**) were employed. The models were trained with either delta (0.5–4 Hz), theta (4–8 Hz) or alpha (8–12 Hz) EEG signals extracted from the EEG recorded to the nursery rhymes. The quality of the stimulus reconstruction was then compared to randomly permuted data (see section "Materials and Methods"). To recap briefly, a backward TRF decoding model was fit separately to the Hilbert envelope of each of the 83 nursery rhyme trials separately for each participant, using a leave-one-out cross-validation procedure (see section "Materials and Methods"). Pearson's correlation (r) was used to test the quality of the reconstruction (**Figure 2B**) providing an objective metric of envelope tracking at the individual level. To test the correlation (r) values against chance, random permutation statistics were created for each participant ($N = 100$ permutations).

To examine whether significant cortical tracking was present in each band, a two-way repeated measures ANOVA was employed. This utilized the factors of condition (real mTRF r values vs randomly permuted mTRF r values) and band (whether these values differed significantly between the analysis bands of 0.5–4 Hz, 4–8 Hz, 8–12 Hz). the data (Greenhouse-Geisser corrected) showed significant main effects of condition [$F(1, 20) = 20.481$, $p = 2.063 \times 10^{-4}$], band; [$F(1.250, 24.992) = 53.748$, $p = 2.47 \times 10^{-8}$] and a significant interaction between condition and band [$F(1.267, 25.340) = 7.772$, $p = 0.007$]. Due to the significant interaction, *post hoc*, Bonferroni corrected, simple main effects analysis was employed. This showed that the real mTRF r values were significantly larger than chance level (randomly permuted mTRF r value) in delta [$F(1, 20) = 12.900$, $p = 0.0005$; mean ± SEM, Rand = 0.019 ± 0.0001, Real = 0.037 ± 0.0003] and theta [$F(1, 20) = 21.171$, $p = 0.0036$; mean ± SEM, Rand = 0.012 ± 0.0001, Real = 0.022 ± 0.002] bands. The alpha band showed a trend toward significance (p = 0.0656), but fell outside our chosen alpha level of $p = 0.05$, [$F(1, 20) = 6.212$, $p = 0.0656$; mean ± SEM, Rand = 0.006 ± 0.0001, Real = 0.009 ± 0.003].

Bayesian, related samples t-test, statistics were also conducted to further investigate the relative effect size within each frequency band. This approach enables a stronger test of the alpha band result, as Bayes Factors indicate the strength of evidence for rejecting the null hypothesis. The Bayesian analyses showed that there was very strong evidence for above chance delta cortical tracking (BF10 = 57.84, so the evidence for the "significant tracking" hypothesis is fifty seven times stronger than the evidence for the null hypothesis), decisive evidence for above chance theta cortical tracking (BF10 = 987.54) but only moderate evidence for above chance alpha cortical tracking (BF10 = 3.31; in Bayesian terms this value means that we have only moderate to anecdotal evidence to reject the null hypothesis). Taken together the adult stimulus reconstruction analyses showed that cortical delta and theta neural signals significantly tracked the envelopes of the nursery rhyme stimuli (**Figure 3**), but there was only moderate evidence for alpha tracking.

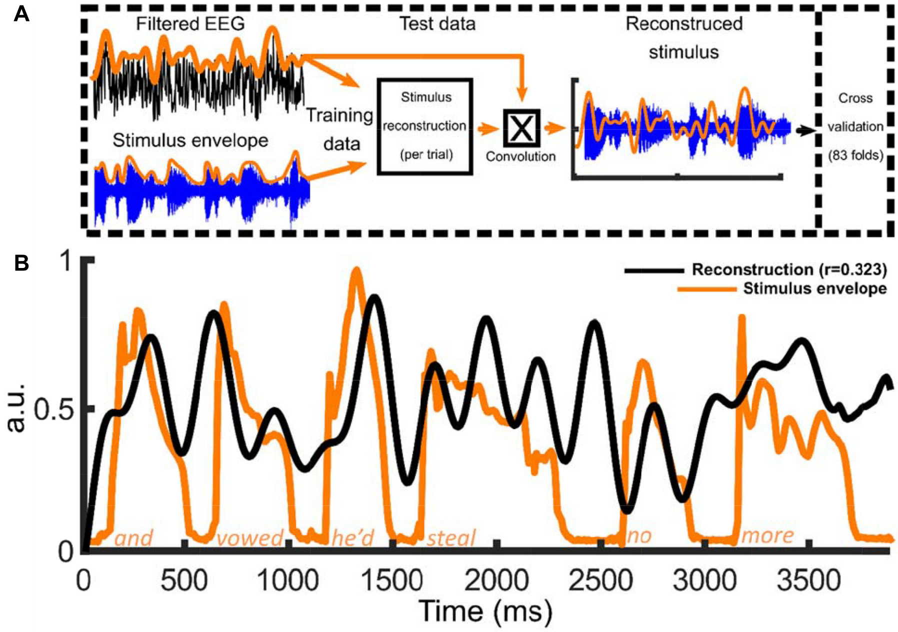


FIGURE 2 | Overview of method to reconstruct the low frequency stimulus envelopes in nursery rhyme phrases using the multivariate temporal response function (mTRF) approach. Panel **(A)** provides a schematic of the stimulus reconstruction model along with a summary of the mTRF analysis pipeline. The EEG signal and the stimulus envelope (absolute value of the Hilbert envelope) were submitted to the mTRF stimulus reconstruction. For the cross validation procedure, 83 nursery rhyme trials were rotated M-1 times each serving once as the "test set" with the remainder of the trials being the "training set." The process was repeated at 12 lambda values (λ values, 1×10^{-3}:1×10^{8}) with the average model convolved with the test data to reconstruct the stimulus envelope at the optimal λ. Panel **(B)** Example of one of the 83 mTRF stimulus reconstructions (this example trained on 0.5:4 Hz data) for one participant along with the original acoustic stimulus envelope. The black line depicts the reconstruction (in arbitrary units; a.u.) and the orange line illustrating the absolute value of the Hilbert envelope of the nursery rhyme phrase, *"and vowed he'd steal no more"* (in arbitrary units; a.u.).

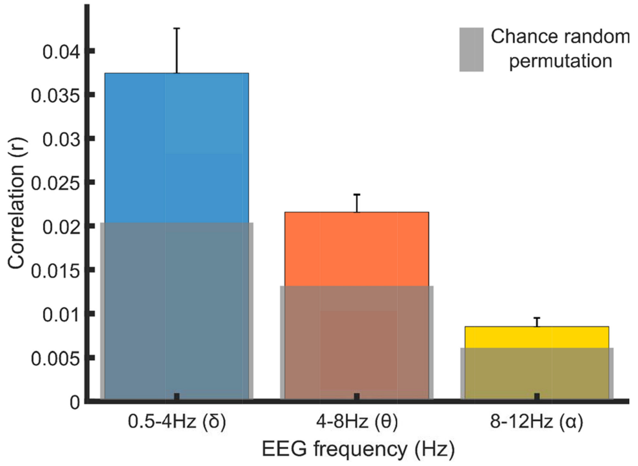


FIGURE 3 | Grand average correlations between stimulus envelope waveforms and their stimulus reconstructions generated by the EEG data from selected frequency bands. Blue, orange and yellow bars show the average correlation value (Pearson's r; mTRF correlation) and standard error, across the 21 participants. The gray bars show the average random permutation *r* values within each frequency band.

A Bonferroni corrected, simple main effects analysis showed that the real mTRF *r* values, in all bands, were significantly different to each other band (**Table 1**). The finding that delta band values were significantly greater than values in the theta band matches the infant data. However, alpha tracking was not significant in the infant data nor in the adult data, although it approached significance in the latter ($p = 0.0656$), as shown in **Figure 3**.

Accordingly, the mTRF stimulus reconstruction data broadly match the results found with infants using the same experimental

TABLE 1 | Pairwise differences between the mTRF *r* values within each frequency band.

*0.01 **0.00001	Delta	Theta	Alpha
Delta		−0.016*	−0.029**
Theta	0.016*		−0.13**
Alpha	0.029**	0.13**	

The numbers show the relative difference between the cortical tracking values (r value) between the different frequency bands column heading, in white text, minus the row heading, in black text).
*Bonferroni corrected p values are denoted by *0.01 and **0.00001.*

procedure and stimulus reconstruction analyses (Attaheri et al., 2022). In both infants and adults, delta and theta EEG responses track the acoustic envelope of sung speech. Regarding the differences between infants and adults observed for the alpha band, it is notable that the stimulus-induced PSD peak at 8.54 Hz only shows anecdotal evidence of a mechanistic role in tracking the envelope of the sung speech. Accordingly, the observed increase in low alpha power (**Figure 1**) may be linked to increased attentional processing (Pichora-Fuller et al., 2016; Dimitrijevic et al., 2017), the recruitment of other language-related processes by the adults (Weisz et al., 2011), or other factors. It is also interesting that the Bayesian analyses run for the adult data showed strong evidence for delta band cortical tracking (BF10 = 20.00) but decisive (i.e., extremely strong) evidence for theta band cortical tracking (BF10 = 166.67). This may imply that theta band tracking increases in its importance in adult speech processing.

Phase Amplitude Coupling

Finally, we explored whether the phase of low-frequency oscillations act to modulate high-frequency amplitudes in similar ways in infants and adults. PAC was calculated using a composite complex-valued signal $z(t, \tau)$ combining the amplitude envelope $A_1(t + \tau)$ of a high-frequency with the phase $\varphi_2(t)$ of a filtered low-frequency signal (Canolty et al., 2006; Özkurt and Schnitzler, 2011). The length of the resulting vector was used as a measure of coupling strength between two frequencies. As differences in low-frequency power have been shown to adversely affect the PAC calculation (Özkurt and Schnitzler, 2011; Aru et al., 2015; Jensen et al., 2016), a statistically normalized version of the modulation index (nMI) was calculated (see section "Materials and Methods"). For each participant low-frequency phases from 2 to 8 Hz (1 Hz steps) and high-frequency amplitudes from 15 to 45 Hz (5 Hz steps) were extracted from the EEG signal from each of the 128 electrode channels. For each of these PAC pairing steps, multiple nMI values were calculated per infant *via* a 5-s sliding window. The significant windows were identified if they exceeded the 95% confidence interval calculated from a surrogate data set made up of 200 statistical iterations of the same analysis window (see section "Materials and Methods" for full procedure).

In order to examine whether a similar pattern of PAC would be exhibited by adults and infants, a two-way repeated measures ANOVA was conducted with two levels, carrier phase and high-frequency amplitude. The aim was to examine whether the nMI values were significantly different when delta versus theta was the low-frequency carrier phase (low-frequency phase; levels of delta and theta) and when beta versus gamma was the high-frequency amplitude (high-frequency amplitude; beta or gamma).

The analysis showed a significant main effect of low-frequency phase [$F(1, 20) = 16.499$, $p = 6.09 \times 10^{-4}$], because theta phases (mean ± SEM, 3.530 ± 0.048) produced significantly higher coupling than delta phases (3.252 ± 0.046). There was no significant effect of high-frequency amplitude [$F(1, 20) = 0.104$, $p = 0.751$], and no significant interaction was observed between low-frequency phases and high-frequency amplitudes [$F(1, 20) = 0.705$, $p = 0.411$].

To further investigate the observed significant effect of phase, Bayesian related samples *t*-tests were conducted to explore the relative effect size within each PAC pairing. One way Bayesian repeated measures ANOVA's were conducted separately comparing delta and theta phase coupling with either gamma or beta high-frequency amplitudes. There was strong evidence that theta/gamma coupling was greater than delta/gamma coupling (BF10 = 21.564), however, there was only anecdotal evidence that theta/beta coupling was greater than delta/beta coupling (BF10 = 1.896).

Overall, the PAC analyses suggest that theta is a more dominant carrier phase than delta when coupling with gamma amplitudes in the adult brain. This is different to infants, where both delta and theta show equal PAC (nMI) with gamma amplitudes when rhythmic speech was the input.

DISCUSSION

Here we replicated with adults a study of cortical tracking to sung speech originally conducted with infants aged 4, 7, and 11 months (Attaheri et al., 2022). Our aim was to explore whether neural responses to rhythmic inputs differ between infants and adults. Accordingly, the same PSD, mTRF stimulus reconstruction and PAC analyses were applied to EEG data recorded from adults, who had experienced the exact same paradigm used previously with infants. It was expected that developmental differences in both cortical tracking and PAC might be observed. Our methods were designed to enable us to distinguish between developmental effects and stimulus-driven phenomena.

Regarding cortical tracking, we found that the mTRF stimulus reconstruction data broadly matched the patterns found with infants (Attaheri et al., 2022), suggestive of no developmental differences. In both infants and adults, delta and theta EEG responses tracked the acoustic envelope of sung speech. Despite significant cortical tracking in both groups, the stimulus-induced PSD peaked at 1.25 Hz in the adult brain, whereas the infant PSD peaks were observed at 2.20 and 4.37 Hz. Only the infant PSD peaks corresponded to the prominent modulation spectrum peaks in the averaged nursery rhyme speech envelopes, suggestive of more stimulus-driven processing (**Supplementary Figure 2**). The lack of peaks at 2.20 and 4.37 Hz for adults may imply a more stimulus-driven response in the infant brain, which appears primarily to track the prominent modulation peaks in the nursery rhyme

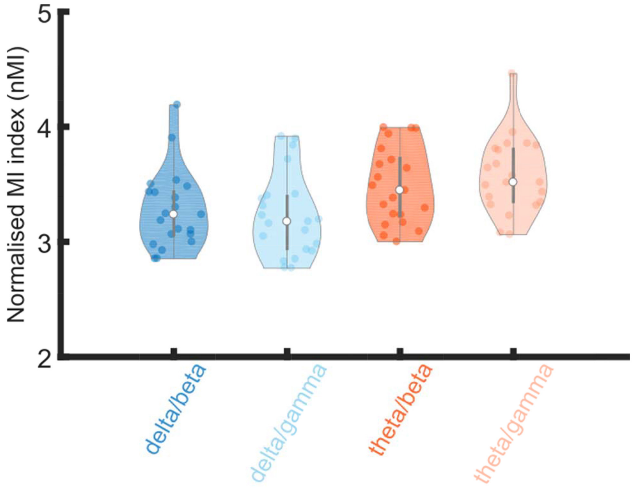


FIGURE 4 | Violin plot of distributions of normalized modulation index (nMI) as measures of phase amplitude coupling (PAC). The PAC bands of interest are given on the X axis (delta/beta, delta/gamma, theta/beta and theta/gamma). Shades of blue denotes PAC pairs with delta as the low frequency phase and shades of orange denotes when theta was the low frequency phase. The nMIs were averaged together (from all significant analysis windows) for each participants data separately for each low frequency phase and high frequency amplitude pairing. The PAC pairing with the maximum nMI, per participant, from within the pre-defined frequency bands of interest; delta 2–4 Hz, theta 4–8 Hz, beta 15–30 Hz and gamma 30–45 Hz), were included in the grand average violin plot.

speech envelopes (Attaheri et al., 2022). It may also be due to the difference in the number of participants between the studies (∼60 infants vs 21 adults). Nevertheless, when the adult data were analyzed using the same center frequencies as the infant PSD peaks (2.20 and 4.37 Hz), significant increases in PSD power compared to resting state were observed for adults also.

The stimulus reconstruction analyses also showed a trend toward significant envelope tracking in the alpha band for the adults, an effect not found for infants. In the adult data, stimulus-induced PSD power also peaked at ∼8.54 and ∼10.04 Hz, both of these peaks occurring in the alpha band. Statistical analysis showed that only the peak at 8.54 Hz was significantly different to adult resting state data. Accordingly, the observed increase in low alpha power may be linked to the formation of additional linguistic processing mechanisms by the adults, for example related to comprehension or attention (Weisz et al., 2011). Bayesian analyses showed very strong evidence for delta band cortical tracking and decisive evidence for theta band cortical tracking. This may suggest that the relative importance of theta band tracking increases with development. Whilst most cortical tracking studies in adults report theta band tracking, there is a now growing body of literature reporting a strong role for delta band tracking in speech processing at lexically and semantically coherent word, phrase and sentence levels (see Ding and Simon, 2014; Kösem and van Wassenhove, 2017 for full review). This raises the possibility that delta band tracking has different functions for infants and adults, however, the current data do not enable any conclusions to be drawn. For our IDS stimuli, the stimulus reconstruction analyses showed that cortical delta and theta neural signals tracked the envelopes of the nursery rhyme stimuli, matching the findings with infants. The delta and theta band entrainment observed here appear to be mainly stimulus-driven effects, as they were also observed in pre-verbal infants. The main developmental difference observed was the relative increase in the strength of the theta band tracking in adults, which could be related to mechanisms important for comprehension and speech intelligibility (Ghitza, 2011). Nevertheless, both frequency bands showed significant cortical tracking in both populations.

The beta and gamma high frequency amplitudes utilized here (15–30 Hz and 30–45 Hz) showed significant coupling to the delta (2–4 Hz) and theta (4–8 Hz) low frequency phases in both infants and adults. However, clear developmental differences in the pattern of phase amplitude coupling were observed. Infants showed significant differences in the high frequency amplitudes used but not in the low frequency phases. Adults showed significant differences in the low frequency phases used but not the high frequency amplitudes. The Bayesian analyses indicated strong evidence that theta/gamma coupling was greater than delta/gamma coupling for adults. This may also relate to the TEMPO model proposed by Ghitza (2011), who has suggested that specific theta-gamma couplings are required when speech is processed for meaning. This pattern differed from prior findings with infants, as the infant brain showed similar levels of PAC to gamma when either delta or theta was the low frequency carrier (the interested reader is invited to compare Attaheri et al., 2022; **Figure 4**, with **Figure 4** here). However, the infant brain did show a significantly higher nMI when gamma was the high frequency amplitude rather than beta. Given previous adult data (Hyafil, 2017), beta was also expected to be less important than gamma for adult PAC. In the adult data, nMI values were greater when gamma rather than beta was coupling to theta phases, with evidence that theta/gamma coupling was stronger than delta/gamma coupling. Overall, the PAC analyses suggest that theta is a more dominant carrier phase than delta when coupling with high frequency amplitudes in the adult brain. This is different to PAC in infants, where both delta and theta showed equal normalized PAC to gamma amplitudes when sung speech was the input.

The current work has a number of limitations. Ideally, children at different ages (i.e., after 11 months) as well as adults should experience the sung speech paradigm while EEG is recorded. This would provide a better assessment of developmental trajectories. The number of adult participants (21) utilized here was also much smaller than the number of infants studied by Attaheri et al. (2022) (∼60 each at ages 4, 7, and 11 months), hence the number of participants could be increased. It could also be interesting to compare adults with and without musical training using the same materials, in order to see whether the use of sung speech reveals any effects of musical expertise on adult cortical tracking. Finally, it could be interesting to record EEG while adults listen to sung nursery materials in unknown languages. This would allow the effects of familiarity and comprehension of the sung speech to be disentangled from the fact that the speech is sung, equating adults and infants for non-maturational factors related to differences in linguistic expertise.

In conclusion, by matching exactly the experimental paradigm and the analysis methods used to investigate cortical tracking of sung speech in infants versus adults, we have revealed more developmental similarities than developmental differences in entrainment. The main differences related to increasing prominence of theta-band mechanisms with age and language experience. While both delta and theta band tracking were observed in the adult brain, the Bayesian analyses showed "decisive" evidence for theta band cortical tracking but "very strong" evidence for delta band cortical tracking. There was also "moderate" evidence for alpha band tracking. Further, there was strong evidence that theta/gamma coupling was stronger than delta/gamma coupling for adults. In the case of infants, both delta and theta showed equally strong coupling to gamma. Whilst we can only speculate regarding the mechanisms underpinning our observed phase/amplitude alignments, previous literature suggests the alignments may be caused by both modulatory and evoked effects (Lakatos et al., 2009; Gross et al., 2013). The increasing role for theta signals in speech processing by adults could reflect many factors, including neural maturation, increased language experience, better language comprehension, and even learning written language, as phase locking in the theta band to rhythmic speech is known to increase in children in line with their reading ability (Power et al., 2012).

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: https://osf.io/s9ezd/.

ETHICS STATEMENT

The study was reviewed by the Psychology Research Ethics Committee of the University of Cambridge, and after a detailed explanation of the study, written consent was given by each participant.

AUTHOR CONTRIBUTIONS

AA: EEG paradigm development, EEG preprocessing, investigation—data curation, formal analysis—design, creation and implementation of analysis, and writing—original draft. DP and AP: predominant data collection. ÁC: data curation—writing—review and editing. GD: formal analysis, writing—review and editing. SR: writing—review and editing. PB: EEG paradigm development and investigation. NM: investigation—data curation. SF: Analysis: modulation spectrum analysis. HO-S: investigation—data curation. UG: conceptualization—methodology, funding acquisition, supervision, project administration and writing—original draft. All authors contributed to the article and approved the submitted version.

FUNDING

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 694786).

ACKNOWLEDGMENTS

We would like to thank all the adult participants, families and infants who kindly donated their time to this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.842447/full#supplementary-material

REFERENCES

Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., et al. (2015). Untangling cross-frequency coupling in neuroscience. *Curr. Opin. Neurobiol.* 31, 51–61. doi: 10.1016/j.conb.2014.08.002

Attaheri, A., Choisdealbha, ÁN., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage* 247:118698. doi: 10.1016/j.neuroimage.2021.118698

Baltzell, L. S., Srinivasan, R., and Richards, V. M. (2017). The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. *J. Neurophysiol.* 118, 3144–3151. doi: 10.1152/jn.00023.2017

Boucher, V. J., Gilbert, A. C., and Jemel, B. (2019). The role of low-frequency neural oscillations in speech processing: revisiting delta entrainment. *J. Cogn. Neurosci.* 31, 1205–1215. doi: 10.1162/jocn_a_01410

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313, 1626–1628. doi: 10.1126/science.1128115

Chait, M., Greenberg, S., Arai, T., Simon, J. Z., and Poeppel, D. (2015). Multi-time resolution analysis of speech: evidence from psychophysics. *Front. Neurosci.* 9:214. doi: 10.3389/fnins.2015.00214

Cirelli, L. K., Spinelli, C., Nozaradan, S., and Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: effects of music background. *Front. Neurosci.* 10:229. doi: 10.3389/fnins.2016.00229

Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10:604. doi: 10.3389/fnhum.2016.00604

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. *ENeuro* 5:ENEURO.0084-18.2018. doi: 10.1523/ENEURO.0084-18.2018

Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465. doi: 10.1016/j.cub.2015.08.030

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Front. Hum. Neurosci.* 11:88. doi: 10.3389/fnhum.2017.00088

Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–35. doi: 10.1523/JNEUROSCI.5297-12.2013

Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Front. Psychol.* 3:238. doi: 10.3389/fpsyg.2012.00238

Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126. doi: 10.1159/000208934

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063

Greenberg, S. (2006). "A multi-tier framework for understanding spoken language," in *Listening to Speech: An Auditory Perspective*, eds S. Greenberg and W. A. Ainsworth (Mahwah: Lawrence Erlbaum Associates), 411–433. doi: 10.4324/9780203933107-25

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752

Hyafil, A. (2017). Disharmony in neural oscillations. *J. Neurophysiol.* 118, 1–3. doi: 10.1152/jn.00026.2017

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A.-L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *ELife* 4:e06213. doi: 10.7554/eLife.06213

Jensen, O., Spaak, E., and Park, H. (2016). Discriminating valid from spurious indices of phase-amplitude coupling. *Eneuro* 3:ENEURO.0334-16.2016. doi: 10.1523/ENEURO.0334-16.2016

Jessen, S., Fiedler, L., Münte, T. F., and Obleser, J. (2019). Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *Neuroimage* 202:116060. doi: 10.1016/j.neuroimage.2019.116060

Jessen, S., Obleser, J., and Tune, S. (2021). Neural tracking in infants – An analytical tool for multisensory social processing in development. *Dev. Cogn. Neurosci.* 52:101034. doi: 10.1016/j.dcn.2021.101034

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., and Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* 8:13745. doi: 10.1038/s41598-018-32150-6

Kösem, A., and van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Lang. Cogn. Neurosci.* 32, 536–544. doi: 10.1080/23273798.2016.1238495

Lakatos, P., Barczak, A., Neymotin, S. A., McGinnis, T., Ross, D., Javitt, D. C., et al. (2016). Global dynamics of selective attention and its lapses in primary auditory cortex. *Nat. Neurosci.* 19, 1707–1717. doi: 10.1038/nn.4386

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005

Leong, V., Kalashnikova, M., Burnham, D., and Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind* 1, 78–90. doi: 10.1162/OPMI_a_00008

Lizarazu, M., Lallier, M., and Molinaro, N. (2019). Phase-amplitude coupling between theta and gamma oscillations adapts to speech rate. *Ann. N. Y. Acad. Sci.* 1453, 140–152. doi: 10.1111/nyas.14099

Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., and Molinaro, N. (2021). Impaired neural response to speech edges in dyslexia. *Cortex* 135, 207–218. doi: 10.1016/j.cortex.2020.09.033

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004

Millman, R. E., Johnson, S. R., and Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *J. Cogn. Neurosci.* 27, 533–545. doi: 10.1162/jocn_a_00719

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., and Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Hum. Brain Mapp.* 37, 2767–2783. doi: 10.1002/hbm.23206

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Özkurt, T. E., and Schnitzler, A. (2011). A critical note on the definition of phase–amplitude cross-frequency coupling. *J. Neurosci. Methods* 201, 438–443. doi: 10.1016/j.jneumeth.2011.08.014

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251

Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr. Opin. Neurobiol.* 28, 142–149. doi: 10.1016/j.conb.2014.07.005

Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Front. Psychol.* 3:216. doi: 10.3389/fpsyg.2012.00216

Weisz, N., Hartmann, T., Müller, N., and Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Front. Psychol.* 2:73. doi: 10.3389/fpsyg.2011.00073

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a 'Cocktail Party'. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

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

frontiers | Frontiers in Psychology

OPINION
published: 06 May 2022
doi: 10.3389/fpsyg.2022.878984

Neural Entrainment vs. Stimulus-Tracking: A Conceptual Challenge for Rhythmic Perceptual Stimulation in Developmental Neuroscience

Anna Bánki[1]*, *Alicja Brzozowska*[1], *Stefanie Hoehl*[1] *and Moritz Köster*[2,3]

[1] Faculty of Psychology, University of Vienna, Vienna, Austria, [2] Institute of Psychology, University of Regensburg, Regensburg, Germany, [3] Faculty of Education and Psychology, Freie Universität Berlin, Berlin, Germany

Keywords: neural entrainment, stimulus-tracking, rhythmic perceptual stimulation, steady-state response (SSR), frequency-tagging, cognitive development

OPEN ACCESS

Edited by:
Elena Nava,
University of Milano-Bicocca, Italy

Reviewed by:
Kaili Clackson,
University of Cambridge,
United Kingdom

Correspondence:
Anna Bánki
anna.banki@univie.ac.at

Specialty section:
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: *18 February 2022*
Accepted: *11 April 2022*
Published: *06 May 2022*

Citation:
Bánki A, Brzozowska A, Hoehl S and Köster M (2022) Neural Entrainment vs. Stimulus-Tracking: A Conceptual Challenge for Rhythmic Perceptual Stimulation in Developmental Neuroscience.
Front. Psychol. 13:878984.
doi: 10.3389/fpsyg.2022.878984

INTRODUCTION

Rhythmic perceptual stimulation, i.e., the presentation of periodic stimuli via sensory input pathways (e.g., auditory or visual; Thut et al., 2011; Calderone et al., 2014) induces resonant brain responses at the presentation frequencies. Since rhythmic perceptual stimulation does not perturbate the neural system beyond its normal operating range (Obleser and Kayser, 2019) and is non-invasive, it has also become a promising tool for neuromodulation in cognitive and developmental neuroscience. In particular, recent work focused on applying rhythmic perceptual stimulation to uncover the functional relevance of neural oscillations at specific frequencies, and their links with cognitive processes in early development, including attention (Christodoulou et al., 2018) and learning (Köster et al., 2019a).

However, an open debate in the adult literature (e.g., Keitel et al., 2014; Haegens, 2020; Meyer et al., 2020; Doelling and Assaneo, 2021; van Bree et al., 2022) emphasizes the need to ascertain whether rhythmic perceptual stimulation directly alters (i.e., stimulates) intrinsic brain oscillations, a phenomenon referred to as *entrainment*, or rather elicits a series of perceptually evoked potentials independent of endogenous oscillatory activity, referred to as *stimulus-tracking* (Capilla et al., 2011; Notbohm et al., 2016).

There is a growing body of research with adults showing that perceptual entrainment of endogenous neural oscillations is indeed possible (Herrmann, 2001; Haegens and Zion Golumbic, 2018; Lakatos et al., 2019; Obleser and Kayser, 2019), as demonstrated by its effects on behavioral outcomes such as memory performance (Köster et al., 2019b) and temporal predictions (Daume et al., 2021), and its interplay with individual intrinsic frequencies (Notbohm et al., 2016; Gulbinaite et al., 2019). Yet, conclusive evidence for entrainment is still lacking, especially from developmental populations.

Here, we bring forward the challenges of interpreting results from rhythmic perceptual stimulation studies with infants and children and discuss how insights from the adult literature can help us adequately examine the entrainment hypothesis in developmental research. We argue that neural entrainment is indeed possible in the developing brain, but that further critical evidence is needed to pin down the underlying neural mechanisms. We discuss the implications of distinguishing entrainment from stimulus tracking for the application of rhythmic perceptual stimulation in developmental neuroscience and suggest potential avenues for future research.

ENTRAINMENT VS. STIMULUS-TRACKING

Oscillations can be modified through synchronization with an external periodic stimulus, a phenomenon called entrainment. By definition, during entrainment, endogenous neural oscillations align with the temporal structure of the exogenous stimulus (Thut et al., 2011; Obleser and Kayser, 2019). It has been proposed that entrainment may facilitate information sampling and sensory selection, underlying various cognitive and perceptual processes (Lakatos, 2008).

Stimulus-tracking refers to the occurrence of frequency-following brain responses to a rhythmic stimulus that, unlike entrainment, show no direct interference with ongoing internal oscillations and related cognitive and perceptual processes (Keitel et al., 2014, 2019; Haegens, 2020). The main difference between the two phenomena is that during stimulus-tracking, a rhythmic external stimulus elicits a frequency following neural sensory response, but ongoing internal brain rhythms are not perturbated. Whereas during entrainment, the stimulus "hijacks" the ongoing internal rhythms at the stimulation frequency, which then become altered and align with the external stimulus (**Figure 1**).

The phenomenon of stimulus-tracking forms the basis of frequency-tagging, a method commonly applied in visual perception research: tagging multiple, simultaneously presented stimuli with different frequencies elicits steady-state responses (SSRs), thus the visual processing of each element can be assessed separately (Müller et al., 2003). Importantly, frequency-tagging considers SSRs independent from the preferred oscillatory frequency of the involved sensory networks (Keitel et al., 2014).

Analytically distinguishing between entrainment and stimulus-tracking solely based on brain activity during rhythmic perceptual stimulation is not possible, as both are characterized by the synchronization between an exogenous stimulus and neural rhythms (Thut et al., 2011; Obleser and Kayser, 2019; Haegens, 2020). However, differentiating between the two phenomena is crucial for studies aiming to cause neuromodulation effects or to test hypotheses about the functionality of endogenous brain rhythms through applying the method of rhythmic perceptual stimulation. There are also implications for paradigms that rely on this methodology to assess attention and perception processes without a specific focus on the underlying oscillatory activity.

Some have postulated that for entrainment to be differentiated from phase-locked responses, it would require that endogenous neural oscillations decouple from the external stimulus and outlive it, particularly in the narrow frequency band in which these oscillations operate (Capilla et al., 2011; Haegens and Zion Golumbic, 2018), but it is not a universally accepted criterion (see e.g., Doelling and Assaneo, 2021). Nonetheless, alterations of endogenous oscillations would likely manifest on the behavioral level and thus could be experimentally tested (Lakatos, 2008).

In the adult literature, there have been two significant ways in particular to establish entrainment: by the alteration of behavioral performance, and by the alignment/matching between the stimulation frequency and the individual internal oscillations

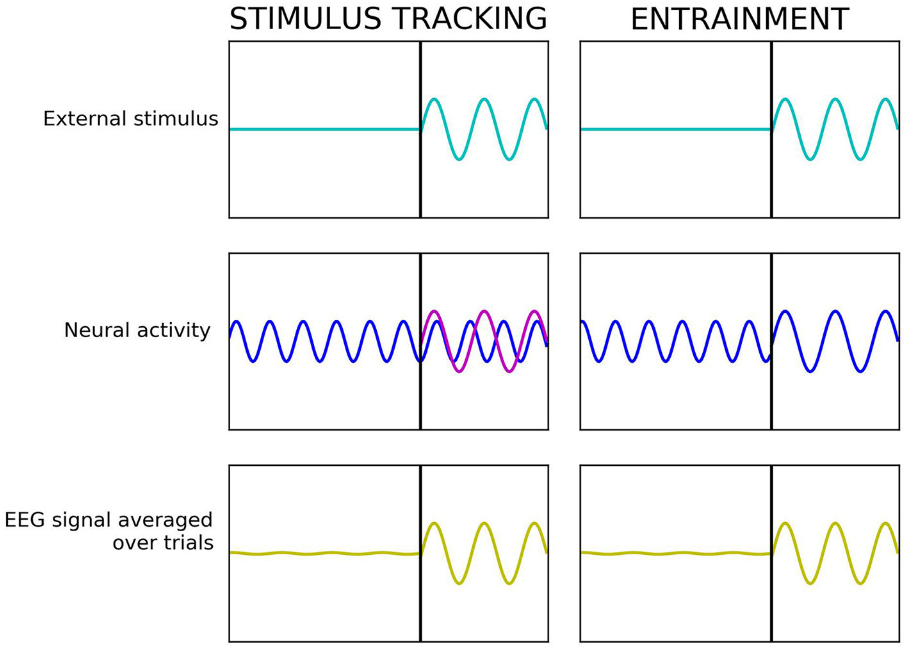


FIGURE 1 | Stimulus-tracking vs. entrainment—simplified example. In all the panels, the *x*-axis reflects the time and the *y*-axis—the amplitude of the signal. In stimulus-tracking (left), rhythmic external stimulation (light blue) causes a frequency-following neural sensory response (magenta), but the internal rhythms (dark blue) are not perturbated. In entrainment (right), rhythmic external stimulation (light blue) causes a resonant response at the frequency of the external stimulus by "hijacking" ongoing internal rhythms (dark blue). In both cases, the resulting electroencephalogram (EEG) signal (yellow) looks the same.

of participants. In the following, we discuss both methods and outline associated challenges as well as future directions for entrainment research with infants and children.

BEHAVIORAL EVIDENCE FOR ENTRAINMENT

Endogenous brain rhythms at different frequency bands have differing functional correlates, e.g., the theta band has been linked to memory processes (Klimesch, 1999; Friese et al., 2013), while the alpha band has been associated with attentional processes (Klimesch, 2012). If rhythmic perceptual stimulation is able to entrain these endogenous brain rhythms, then we should be able to observe distinct behavioral effects of stimulation at external frequencies corresponding to different EEG frequency bands (Herrmann et al., 2016).

Good examples of this line of research supporting this notion are studies examining memory-enhancing effects of rhythmic sensory stimulation at the theta frequency, which have accumulated in the recent years (see Hanslmayr et al., 2019 for a review). For example, it has been shown that presenting audiovisual and visual stimuli at the theta, but not the alpha frequency improves later memory of these stimuli (Clouter et al., 2017; Köster et al., 2019b). What is important, this effect is dependent on the phase synchrony between the external stimulus and the brain response (Clouter et al., 2017; Wang et al., 2018), as well as the power of the response (Köster et al., 2019b), which further suggests that the behavioral effects of rhythmic perceptual stimulation are caused by neural entrainment. Interestingly, the behavioral effects of entrainment of the theta rhythm seem to persist beyond the period of stimulation, as Roberts et al. (2018) found improved source memory in participants exposed to rhythmic sensory stimulation (content-wise irrelevant to the memorized stimuli) between learning and test phase. To conclude, research with adults provides accumulating evidence for the modulation of cognitive processes through rhythmic perceptual stimulation.

ENTRAINMENT AT INDIVIDUAL STIMULATION FREQUENCIES

It has been argued that an important feature of entrainment is that it should occur for external rhythms close to the neural oscillator's intrinsic rate (so-called "eigenfrequency"; Obleser and Kayser, 2019; Haegens, 2020). Indeed, Notbohm et al. (2016) demonstrated that rhythmic visual flicker at frequencies closer to participants' intrinsic frequencies of the alpha rhythm caused a more pronounced phase coupling between the external rhythm and the neural response. Moreover, a number of studies have found that the behavioral effects (i.e., performance impairment) of stimulation in the alpha band on participants' attention are predicted by how close the stimulation frequencies are to participants' endogenous alpha peaks (de Graaf et al., 2013; Gulbinaite et al., 2017). Finally, Köster et al. (2019b) successfully entrained memory functions by stimulating participants' brain activity at their individual theta vs. alpha frequencies.

However, more studies examining other frequency bands and an improved understanding of the generators behind the intrinsic brain rhythms are needed. Ideally, future studies would also examine the behavioral effects of individually optimized vs. not individually optimized stimulation frequencies within the same frequency bands. In developmental populations, EEG frequency ranges change with age by moving toward higher frequencies (Marshall et al., 2002), thus we would also expect individual frequencies to shift toward the higher ends of the spectra with age. Observing behavioral effects of rhythmic sensory stimulation at frequencies optimized (vs. non-optimized) to specific age groups would support the notion of entrainment effects.

ENTRAINMENT IN DEVELOPMENTAL RESEARCH

Previous studies with infants and children applied rhythmic auditory and/or visual stimulation as a tool to assess attention (e.g., Robertson et al., 2012; Köster et al., 2017; Christodoulou et al., 2018), face perception (e.g., de Heering and Rossion, 2015; Peykarjou et al., 2017), speech perception and language learning (e.g., Telkemeyer et al., 2009; Attaheri et al., 2022), and related developmental disorders such as dyslexia (e.g., Colling et al., 2017), as well as musical rhythm perception (e.g., Cirelli et al., 2016). In these studies, stimulation frequencies were chosen either arbitrarily, to index brain responses to stimuli of interest (e.g., a face appearing at 1.2 Hz among a stream of images updated at 6 Hz in the study by de Heering and Rossion, 2015), or were motivated by hypotheses about naturally occurring frequencies in infants' environment (e.g., sounds with spectral properties resembling naturalistic speech in the study by Telkemeyer et al., 2009). Another approach aimed to capture infants' stimulus-tracking by reconstructing features of natural stimuli based on brain responses or by predicting brain activity from stimulus features (e.g., Jessen et al., 2021). However, very few developmental studies to date applied rhythmic visual stimulation with the explicit goal to entrain endogenous rhythms (e.g., Köster et al., 2019a), an approach that can help to uncover the functionality of brain oscillations in early cognitive and perception processes.

Recent debates in the developmental literature challenged the interpretation of findings from rhythmic perceptual stimulation studies since the underlying mechanisms of entrainment are still not well-understood. As an example, Köster et al. (2019a) used this method to assess the functional role of infants' theta (4 Hz) and alpha (6 Hz) oscillations in the cognitive processing of unexpected events. Findings revealed that infants' visually entrained theta, but not alpha oscillations sharply increased for unexpected vs. expected events, in line with evidence on the critical role of theta oscillations in early learning (Begus and Bonawitz, 2020; Köster et al., 2020). However, in a commentary, using simulated data, Keitel et al. (2021) argued that the results from Köster et al. (2019a) could potentially be explained by a stimulus-tracking account: unexpected events could enhance the negative central (Nc) component of the event related potential (ERP) (Kayhan et al., 2019), leading to a difference in ERPs between conditions without the involvement of functionally relevant oscillations elicited by the stimulus. Although the simulated data did not closely reflect the observed data (see Köster et al., 2021b), this debate highlights the need to differentiate between entrainment and stimulus-tracking when interpreting data from rhythmic perceptual stimulation studies.

AVENUES FOR FUTURE RESEARCH

Here, we argued that further evidence is needed on how perceptually entrained rhythms interact with ongoing oscillatory dynamics in the developing brain, so that a comprehensive theoretical and analytical framework could be established for developmental entrainment research. Studies in the field need to investigate the effects of rhythmic perceptual stimulation by including behavioral outcome measures, and by using individually optimized stimulation frequencies. However, commonly applied paradigms in adult entrainment research need to be adjusted for developmental studies (Wass et al., 2022).

Measuring altered behavior of infants following rhythmic perceptual stimulation—though challenging—can be achieved by subsequent memory paradigms that assess learning and memory performance via preferential looking (Begus et al., 2015), imitation (Köster et al., 2021a), or habituation and dishabituation (Choi et al., 2020). In case of older children, behavioral outcomes of entrainment such as attention, learning and memory effects can be tested by incorporating subsequent playful tasks into the experimental design. Observed behavioral effects linked to sensory stimulation would provide causal evidence for neural oscillatory entrainment. Applying stimulation at individualized frequencies could be another non-invasive way to ascertain entrainment in developmental samples, by testing the alignment of endogenous oscillations following sensory stimulation with infants' or children's individual vs. other frequencies. The two approaches can be combined to contrast behavioral changes following entrainment to individualized vs. other frequencies.

Taken together, we need a deeper understanding of the effects of rhythmic perceptual stimulation on endogenous oscillatory activity to establish if entrainment occurs in the developing brain, and if so, whether it can be considered as a tool for neuromodulation to test the functional aspects of specific neural rhythms in early cognitive development.

AUTHOR CONTRIBUTIONS

ABá and ABr wrote the manuscript. SH and MK contributed to revising and editing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by an FWF grant awarded to SH (Grant number: P 33853) and a D-A-CH research grant awarded to MK and SH by the DFG and FWF jointly (Grant numbers: KO 6028/1-1; I 4332). We are grateful for the open access funding provided by the University of Vienna.

REFERENCES

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage* 247, 118698. doi: 10.1016/j.neuroimage.2021.118698

Begus, K., and Bonawitz, E. (2020). The rhythm of learning: theta oscillations as an index of active learning in infancy. *Dev. Cogn. Neurosci.* 45, 100810. doi: 10.1016/j.dcn.2020.100810

Begus, K., Southgate, V., and Gliga, T. (2015). Neural mechanisms of infant learning: differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biol. Lett.* 11, 20150041. doi: 10.1098/rsbl.2015.0041

Calderone, D. J., Lakatos, P., Butler, P. D., and Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cogn. Sci.* 18, 300–309. doi: 10.1016/j.tics.2014.02.005

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS ONE* 6, e14543. doi: 10.1371/journal.pone.0014543

Choi, D., Batterink, L., Black, A. K., Paller, K., and Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: evidence from neural entrainment. *Psychol. Sci.* 31, 1161–1173. doi: 10.1177/0956797620933237

Christodoulou, J., Leland, D. S., and Moore, D. S. (2018). Overt and covert attention in infants revealed using steady-state visually evoked potentials. *Dev. Psychol.* 54, 803–815. doi: 10.1037/dev0000486

Cirelli, L. K., Spinelli, C., Nozaradan, S., and Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: effects of music background. *Front. Neurosci.* 10, 229. doi: 10.3389/fnins.2016.00229

Clouter, A., Shapiro, K. L., and Hanslmayr, S. (2017). Theta phase synchronization is the glue that binds human associative memory. *Curr. Biol.* 27, 3143–3148. doi: 10.1016/j.cub.2017.09.001

Colling, L. J., Noble, H. L., and Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: an EEG study. *Front. Neurosci.* 11, 360. doi: 10.3389/fnins.2017.00360

Daume, J., Wang, P., Maye, A., Zhang, D., and Engel, A. K. (2021). Non-rhythmic temporal prediction involves phase resets of low-frequency delta oscillations. *NeuroImage* 224, 117376. doi: 10.1016/j.neuroimage.2020.117376

de Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., and Thut, G. (2013). Alpha-band rhythms in visual task performance: phase-locking by rhythmic sensory stimulation. *PLoS ONE* 8, e60035. doi: 10.1371/journal.pone.0060035

de Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *Elife* 4, e06564. doi: 10.7554/eLife.06564

Doelling, K. B., and Assaneo, M. F. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biol.* 19, e3001234. doi: 10.1371/journal.pbio.3001234

Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., and Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *NeuroImage*. 66, 642–647. doi: 10.1016/j.neuroimage.2012.11.002

Gulbinaite, R., Roozendaal, D. H. M., and VanRullen, R. (2019). Attention differentially modulates the amplitude of resonance frequencies in the visual cortex. *NeuroImage* 203, 116146. doi: 10.1016/j.neuroimage.2019.116146

Gulbinaite, R., van Viegen, T., Wieling, M., Cohen, M. X., and VanRullen, R. (2017). Individual alpha peak frequency predicts 10 Hz flicker effects on selective attention. *J. Neurosci.* 37, 10173–10184. doi: 10.1523/JNEUROSCI.1163-17.2017

Haegens, S. (2020). Entrainment revisited: a commentary on Meyer, Sun, and Martin (2020). *Lang. Cogn. Neurosci.* 35, 1119–1123. doi: 10.1080/23273798.2020.1758335

Haegens, S., and Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: a critical review. *Neurosci. Biobehav. Rev.* 86, 150–165. doi: 10.1016/j.neubiorev.2017.12.002

Hanslmayr, S., Axmacher, N., and Inman, C. S. (2019). Modulating human memory via entrainment of brain oscillations. *Trends Neurosci.* 42, 485–499. doi: 10.1016/j.tins.2019.04.004

Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. doi: 10.1007/s002210100682

Herrmann, C. S., Strüber, D., Helfrich, R. F., and Engel, A. K. (2016). EEG oscillations: from correlation to causality. *Int. J. Psychophysiol.* 103, 12–21. doi: 10.1016/j.ijpsycho.2015.02.003

Jessen, S., Obleser, J., and Tune, S. (2021). Neural tracking in infants – An analytical tool for multisensory social processing in development. *Dev. Cogn. Neurosci.* 52, 101034. doi: 10.1016/j.dcn.2021.101034

Kayhan, E., Meyer, M., O'Reilly, J. X., Hunnius, S., and Bekkering, H. (2019). Nine-month-old infants update their predictive models of a changing environment. *Dev. Cogn. Neurosci.* 38, 100680. doi: 10.1016/j.dcn.2019.100680

Keitel, C., Keitel, A., Benwell, C. S. Y., Daube, C., Thut, G., and Gross, J. (2019). Stimulus-driven brain rhythms within the alpha band: the attentional-modulation conundrum. *J. Neurosci.* 39, 3119–3129. doi: 10.1523/JNEUROSCI.1633-18.2019

Keitel, C., Obleser, J., Jessen, S., and Henry, M. J. (2021). Frequency-specific effects in infant electroencephalograms do not require entrained neural oscillations: a commentary on Köster et al. (2019). *Psychol. Sci.* 32, 966–971. doi: 10.1177/09567976211001317

Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: Entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Köster, M., Castel, J., Gruber, T., and Kärtner, J. (2017). Visual cortical networks align with behavioral measures of context-sensitivity in early childhood. *NeuroImage* 163, 413–418. doi: 10.1016/j.neuroimage.2017.08.008

Köster, M., Kayhan, E., Langeloh, M., and Hoehl, S. (2020). Making Sense of the World: Infant Learning from a Predictive Processing Perspective. *Perspect. Psychol. Sci.* 15, 562–571. doi: 10.1177/1745691619895071

Köster, M., Langeloh, M., and Hoehl, S. (2019a). Visually entrained theta oscillations increase for unexpected events in the infant brain. *Psychol. Sci.* 30, 1656–1663. doi: 10.1177/0956797619876260

Köster, M., Langeloh, M., Kliesch, C., Kanngiesser, P., and Hoehl, S. (2021a). The value of subsequent memory paradigms in uncovering neural mechanisms of early social learning. *NeuroImage* 234, 117978. doi: 10.1016/j.neuroimage.2021.117978

Köster, M., Langeloh, M., Michel, C., and Hoehl, S. (2021b). Young infants process prediction errors at the theta rhythm. *NeuroImage* 236, 118074. doi: 10.1016/j.neuroimage.2021.118074

Köster, M., Martens, U., and Gruber, T. (2019b). Memory entrainment by visually evoked theta-gamma coupling. *NeuroImage* 188, 181–187. doi: 10.1016/j.neuroimage.2018.12.002

Lakatos, P. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 5872. doi: 10.1126/science.1154735

Lakatos, P., Gross, J., and Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Curr. Biol.* 29, 890–905. doi: 10.1016/j.cub.2019.07.075

Marshall, P. J., Bar-Haim, Y., and Fox, N. A. (2002). Development of the EEG from 5 months to 4 years of age. *Clin. Neurophysiol.* 113, 1199–1208. doi: 10.1016/S1388-2457(02)00163-3

Meyer, L., Sun, Y., and Martin, A. E. (2020). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Lang. Cogn. Neurosci.* 35, 1089–1099. doi: 10.1080/23273798.2019.1693050

Müller, M. M., Malinowski, P., Gruber, T., and Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature.* 424, 309–312. doi: 10.1038/nature01812

Notbohm, A., Kurths, J., and Herrmann, C. S. (2016). Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. *Front. Hum. Neurosci.* 10, 10. doi: 10.3389/fnhum.2016.00010

Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* 23, 913–926. doi: 10.1016/j.tics.2019.08.004

Peykarjou, S., Hoehl, S., Pauen, S., and Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* 7, 12526. doi: 10.1038/s41598-017-12760-2

Roberts, B. M., Clarke, A., Addante, R. J., and Ranganath, C. (2018). Entrainment enhances theta oscillations and improves episodic memory. *Cogn. Neurosci.* 9, 181–193. doi: 10.1080/17588928.2018.1521386

Robertson, S. S., Watamura, S. E., and Wilbourn, M. P. (2012). Attentional dynamics of infant visual foraging. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11460. doi: 10.1073/pnas.1203482109

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733. doi: 10.1523/JNEUROSCI.1246-09.2009

Thut, G., Schyns, P. G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Front. Psychol.* 2, 170. doi: 10.3389/fpsyg.2011.00170

van Bree, S., Alamia, A., and Zoefel, B. (2022). Oscillation or not—Why we can and need to know (commentary on Doelling and Assaneo, 2021). *Eur. J. Neurosci.* 55, 201–204. doi: 10.1111/ejn.15542

Wang, D., Clouter, A., Chen, Q., Shapiro, K. L., and Hanslmayr, S. (2018). Single-trial phase entrainment of theta oscillations in sensory regions predicts human associative memory performance. *J. Neurosci.* 38, 6299–6309. doi: 10.1523/JNEUROSCI.0349-18.2018

Wass, S., Perapoch Amadó, M., and Ives, J. (2022). Oscillatory entrainment to our early social or physical environment and the emergence of volitional control. *Dev. Cogn. Neurosci.* 54, 101102. doi: 10.1016/j.dcn.2022.101102

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fpsyg.2022.894405

Rhythm May Be Key to Linking Language and Cognition in Young Infants: Evidence From Machine Learning

Joseph C. Y. Lau[1,2,3]*, *Alona Fyshe*[4] *and Sandra R. Waxman*[1,2]

[1] Department of Psychology, Northwestern University, Evanston, IL, United States, [2] Institute for Policy Research, Northwestern University, Evanston, IL, United States, [3] Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, [4] Department of Computing Science and Psychology, University of Alberta, Edmonton, AB, Canada

Rhythm is key to language acquisition. Across languages, rhythmic features highlight fundamental linguistic elements of the sound stream and structural relations among them. A sensitivity to rhythmic features, which begins *in utero*, is evident at birth. What is less clear is whether rhythm supports infants' earliest links between language and cognition. Prior evidence has documented that for infants as young as 3 and 4 months, listening to their native language (English) supports the core cognitive capacity of object categorization. This precocious link is initially part of a broader template: listening to a non-native language from the same rhythmic class as (e.g., German, but not Cantonese) and to vocalizations of non-human primates (e.g., lemur, *Eulemur macaco flavifrons*, but not birds e.g., zebra-finches, *Taeniopygia guttata*) provide English-acquiring infants the same cognitive advantage as does listening to their native language. Here, we implement a machine-learning (ML) approach to ask whether there are acoustic properties, available on the surface of these vocalizations, that permit infants' to identify which vocalizations are candidate links to cognition. We provided the model with a robust sample of vocalizations that, from the vantage point of English-acquiring 4-month-olds, either support object categorization (English, German, lemur vocalizations) or fail to do so (Cantonese, zebra-finch vocalizations). We assess (a) whether supervised ML classification models can distinguish those vocalizations that support cognition from those that do not, and (b) which class(es) of acoustic features (including rhythmic, spectral envelope, and pitch features) best support that classification. Our analysis reveals that principal components derived from rhythm-relevant acoustic features were among the most robust in supporting the classification. Classifications performed using temporal envelope components were also robust. These new findings provide *in principle* evidence that infants' earliest links between vocalizations and cognition may be subserved by their perceptual sensitivity to rhythmic and spectral elements available on the surface of these vocalizations, and that these may guide infants' identification of candidate links to cognition.

Keywords: infant cognition, language, rhythm, machine learning, non-human vocalizations

Edited by:
Stefanie Peykarjou,
Heidelberg University, Germany

Reviewed by:
Courtney Hilton,
Harvard University, United States
Katharina Menn,
Max Planck Institute for Human Cognitive and Brain Sciences, Germany

***Correspondence:**
Joseph C. Y. Lau
josephcylau@northwestern.edu

Specialty section:
This article was submitted to Developmental Psychologys, a section of the journal Frontiers in Psychology

Received: *11 March 2022*
Accepted: *03 May 2022*
Published: *26 May 2022*

Citation:
Lau JCY, Fyshe A and Waxman SR (2022) Rhythm May Be Key to Linking Language and Cognition in Young Infants: Evidence From Machine Learning. Front. Psychol. 13:894405. doi: 10.3389/fpsyg.2022.894405

1. INTRODUCTION

The link between human language and cognition affords us exceptional communicative and representational power. By virtue of this link, we are able establish mental representations, ones that permit us to move beyond the present to consider the past and imagine the future, and to communicate these representations with others (Miller, 1990). Within the developmental sciences, considerable attention has been devoted to discovering how, and how early, this language-cognition link is established (Perszyk and Waxman, 2018 for a recent review). Considerable evidence has been derived from behavioral experiments measuring infant object categorization. Object categorization is a fundamental building block of cognition. In human infants, as in adults, categorization is supported by language (Gelman, 2004; Waxman and Gelman, 2009). Indeed, language supports infants' ability to form categories (Waxman and Markow, 1995; Perszyk and Waxman, 2018).

The evidence comes from a simple, yet robust object categorization task (Ferry et al., 2010, 2013; Perszyk and Waxman, 2019; Woodruff Carr et al., 2021b). During a familiarization phase, infants view a series of distinct objects, all members of the same object category (e.g., images of 8 different dinosaurs). Next, during the test phase, infants view two novel objects—one a member of the familiarized category (e.g., another dinosaur) and the other a member of a different category (e.g., a fish). The logic is straightforward: if infants detected the category-based commonalities among the familiarization objects, then they should distinguish the novel from familiar test object (as indexed by a reliable looking preference, i.e., longer looking time, for the novel object); if infants failed to detect the familiarization category, then they should fail to distinguish between the novel and familiar images. This task also allows the effect of auditory signals on object categorization to be examined: with infants viewing the same visual images in the same paradigm, the sounds paired with the familiarization images can be systematically manipulated. With this uniform design, the effect of different auditory signals on object categorization can be directly compared, even across studies by comparison of effect sizes (Woodruff Carr et al., 2021b).

Studies utilizing this task offer compelling evidence that infants' categorization is influenced by listening to language. For infants as young as 3- and 4-months, listening to their native language boosts their performance in object categorization, and does so in a way that carefully-matched acoustic signals (sine-wave tone sequences, backward speech) do not (Ferry et al., 2010, 2013). Moreover, this link to cognition is shaped by infant's own language experience. For 3- to 4-month-old infants acquiring English, listening to either English or German (a "typological cousin" to their native English with similar rhythmic properties) facilitates object categorization in the same task. In contrast, listening to Cantonese (a language typologically and rhythmically distant from English) fails to support object categorization in this task (Perszyk and Waxman, 2019). Apparently, then, infants' increasingly precise perceptual tuning to their native language (Werker and Tees, 1984; Kuhl and Rivera-Gaxiola, 2008; Peña et al., 2010; Werker, 2018) has powerful downstream consequences beyond perception alone; it also sets boundaries on which other language(s) support infant cognition.

Surprisingly, however, infants' earliest link is not restricted to language alone. Even at 4 months, as infants are narrowing the range of human languages they link to cognition, during the same object categorization task, listening to the vocalizations of non-human primates (e.g., blue-eyed black lemur, *Eulemur macaco flavifrons*) confers the same cognitive advantage as does listening to their native language (Ferry et al., 2013; Woodruff Carr et al., 2021b). Importantly, however, this link may be restricted to vocalizations of primates, our nearest evolutionary relations; It is not so broad as to include vocalization of birds (e.g., zebra-finches, *Taniopygia guttata*).

Taken together, these findings raise a compelling new question. Which acoustic features, if any, are available on the surface of human and non-human vocalizations to support very young infants in identifying which vocalizations might serve as candidate links to cognition (Ferry et al., 2013; Woodruff Carr et al., 2021b)? Focusing on the language side of this link, researchers have discovered that a strong sensitivity to rhythm, available *in utero*, is essential in the acquisition of language from the start (May et al., 2011; Langus et al., 2017; Minai et al., 2017; Gervain et al., 2021). *In utero*, the womb and other maternal tissues act as a low-pass acoustic filter, permitting lower frequency cues including rhythm and stress patterns, but not the higher-frequency cues that encode segmental detail, to be transmitted (Lecanuet and Granier-Deferre, 1993). Within hours of their birth, infants distinguish languages from the same rhythmic class as their native language, preferring them over languages with non-native rhythmical patterns (Mehler et al., 1988; Nazzi et al., 1998). Within the first year of life, rhythm continues to be instrumental (Christophe et al., 2001), enabling infants to segment the continuous speech stream into words (Johnson and Jusczyk, 2001) and to discover other structural linguistic properties including word order (Gervain and Werker, 2013) and syntactic structures (Nazzi et al., 2000). There is also strong neurophysiological evidence for the importance of speech rhythm. Neurons in the auditory cortex oscillate at frequencies that entrain speech rhythm. This entrainment, which enables infants to extract hierarchical information, including lexical stress, syllabic structure and syntactic patterns (Goswami, 2019), evident already at 4 months and develops throughout infancy (Attaheri et al., 2022), continues to support language processing in adulthood (Poeppel and Assaneo, 2020).

Thus, the power of rhythm is clear: Rhythm supports language acquisition from the start and continues to support language processing into adulthood (Gleitman and Wanner, 1982; Morgan and Demuth, 1996; Hilton and Goldwater, 2021).

Because rhythm is crucial in early language acquisition, there is reason to suspect that it may be instrumental in guiding infants to identify candidate links to cognition. There is strong evidence that speech rhythm (including that of their native language and others from the same rhythmic class), engage infant attention robustly (Jusczyk et al., 1993; Sansavini et al., 1997; Höhle et al., 2009; Räsänen et al., 2018). Finally, rhythmic properties that occur in both human language and mammalian vocalizations have been documented (Kotz et al., 2018; Ravignani et al., 2019).

This behavioral evidence of infants' sensitivity to rhythm, coupled with cross-species findings on shared rhythmic features across language and non-human vocalizations, leads to an intriguing hypothesis: that there are rhythmic properties present at the surface of languages and non-human vocalizations, that would, in principle, support 3- to 4-month-old infants in identifying them as candidate links to cognition.

Here, we provide the first test of this hypothesis. Implementing a supervised machine-learning (ML) approach, we trained a series of ML models, using acoustic features derived from a collection of audio samples of human languages and non-human vocalizations, to classify samples into classes of signals that either support infant cognition (i.e., English, German, and lemur vocalizations) or do not (i.e., Cantonese, and zebra finch vocalizations), from the vantage point of 3- to 4-month-old English-acquiring infants (Ferry et al., 2010, 2013; Perszyk and Waxman, 2016, 2019; Woodruff Carr et al., 2021b). With these models, we tested the hypothesis that rhythmic features, present at the surface of the input, support the training of the distinguishing signals that do, and do not, support cognition. As comparison, we also tested two other fundamental classes of vocal properties, namely (1) spectral envelope features and (2) pitch features. Spectral envelope features are associated with vocal configurations that differ across languages segmentally (e.g., in terms of consonant and vowel repertoire) and across species (e.g., laryngeal vs. syringeal vocalizations) (Mogran et al., 2004; Cheng et al., 2012; Andén and Mallat, 2014; Fedurek et al., 2016). Pitch features represent fundamental vocal properties across species (Belin, 2006), as well as speech intonation, another fundamental aspect of prosody central to infant language acquisition (Nooteboom, 1997).

2. METHODS

2.1. Materials: Vocalization Dataset

Our modeling dataset consisted of a total of 3,197 audio samples (**Table 1**) of human languages and non-human vocalizations for which links to cognition (or the lack thereof) have been attested behaviorally thus far in 4-month-old infants (Ferry et al., 2010, 2013; Perszyk and Waxman, 2016, 2019; Woodruff Carr et al., 2021b).

Language audio samples were utterance-length recordings produced by multiple female native speakers of English, German, and Cantonese, in their respective languages, using an infant directed speech (IDS) register in interactions with a young child. These audio samples were high-quality recordings from three different publicly available or private IDS corpora. Samples of American English were parts of a multilanguage corpora collected for the purpose of examining aspects of universality of IDS across cultures and societies (Hilton et al., 2022). Samples of German were from the Konstanz Prosodically Annotated InfantDirected Speech (KIDS) Corpus (Zahner et al., 2016), collected from a semi- structured mother-infant play situation where mothers were given a picture book and some other toys, which they could use according to their infant's interest. Cantonese samples were from the dataset of a study examining functions of acoustic-phonetic modifications in IDS (Wang et al., 2021). The Cantonese IDS samples were collected from a semi-structured caregiver-child interaction task, where various toys were given to the female caregiver to elicit keywords of interest, while she played with the child.

Audio samples of non-human vocalizations consisted of lemur and zebra finch vocalizations. Samples of lemur vocalizations were from a private collection of lemur vocalizations collected for a sound art project (Mercer, 2012), collected from single semi-free-range lemurs from the lemur habitat in the Duke University Lemur Center. Zebra finch vocalization samples were from a publicly available database of zebra finch songs (Laboratory of Vocal Learning at Hunter College, 2015), which have also been analyzed in prior acoustic studies (Tchernichovski et al., 2001; Isomura et al., 2019).

Descriptive statistics of our vocalization dataset are presented in **Table 1**.

2.2. Acoustic Feature Extraction

A series of multivariate acoustic features were extracted from each of the vocalization samples, to serve as input in subsequent ML classification. Before feature extraction, all audio samples were first normalized in intensity (80 dB) and resampled to a sampling rate of 22,050 Hz. Since the duration of each vocalization sample varies, the duration of each vocalization was normalized by repeating the audio samples until it reaches 9.54 s (i.e., samples), the maximum duration among all vocalization samples. Next, from each time-normalized vocalization sample, we extracted three series of acoustic features that have been shown to primarily represent *rhythmic*, *spectral envelope*, or *pitch information*, respectively (e.g., Hilton et al., 2022). The three series of acoustic features.

First, for *rhythmic* features: four types of acoustic features were derived from all vocalization samples to comprehensively capture aspects of rhythm, namely:

1. The speech envelope spectrum (ENV) represents temporal regularities correlating to rhythmic properties of the signal (Tilsen and Johnson, 2008; Poeppel and Assaneo, 2020; Hilton and Goldwater, 2021). For each vocalization sample, the vocalic energy amplitude envelope was first derived. To derive the envelope, the raw time series was first chunked into consecutive bins of 1 s. Following Tilsen and Arvaniti (2013), the time series of each chunk was filtered with a passband of 400–4,000 Hz to de-emphasize non-vocalic energy such as glottal energy (including the f0) and obstruent noise. The bandpass-filtered signal was then low-pass filtered with a cutoff of 10 Hz to represent the envelope. The frequency decomposition of the envelope was then computed. First, the envelope was downsampled by a factor of 100 and windowed using a Tukey window ($r = 0.1$) to aid further spectral analyses. The envelope was then normalized by subtracting the mean and rescaled to have minimum and maximum values of −1 and 1, respectively. A fast Fourier transform was first applied to the normalized envelope which was also zero-padded to a 2,048-sample window. The spectra across all 1-s chunks were then averaged to form the envelope spectrum of the vocalization sample and included as features.

2. The intrinsic mode functions (IMFs) were further computed from the time-varying speech envelope (as described above) using empirical mode decomposition (EMD), representing syllabic (IMF1) and supra-syllabic-level (IMF2) fluctuations relevant to speech rhythm (Tilsen and Arvaniti, 2013). The frequency decompositions of IMF1 and IMF2 (i.e., the averaged power spectrum density of 1–10 Hz from the frequency decomposition all IMF1s and IMF2s across all 1-s envelope bins of each vocalization sample) were included as features. We selected a bin duration of 1 s to maximally eliminate the representations of slower prosodic information (e.g., intonation) and mixtures of tempos and variations in rhythmicity not relevant to the syllabic and supra-syllabic rhythm (Tilsen and Arvaniti, 2013).
3. The temporal modulation spectrum (TMS) is the frequency decomposition of the temporal envelope of a signal that reflects how fast sound intensity fluctuates over time (Ding et al., 2017). Temporal modulation of lower frequencies (<32 Hz) is a primary acoustic correlate of perceived rhythm in speech (Greenberg et al., 2003; Goswami and Leong, 2016), which contributes to speech intelligibility (Elliott and Theunissen, 2009). For each vocalization sample, the raw time series was first chunked into consecutive bins of 1 s. The TMS of each 1-s bin was then computed using the procedure and MATLAB script from Ding et al. (2017). In the procedure, the sound signal in each bin was first decomposed into narrow frequency bands using a cochlear model and then from each band the temporal envelope was extracted. The extracted envelopes were rescaled using a logarithmic function, and were then converted into the frequency domain by the Discrete Fourier Transform (DFT). The TMS was the root-mean-square of the DFT of all narrowband power envelopes. The TMS features of each vocalization sample were taken as the average TMS of all bins.
4. The wavelet time scattering (WTS) representations are low-variance representations of time-frequency properties of sounds including amplitude and frequency modulations of acoustic signals (Andén and Mallat, 2014; Andén et al., 2015). The WTS is resistant to time-warping deformations, and is therefore advantageous to be used in machine learning since as class discriminability is not sacrificed in the transformation. The WTS has been used in machine-learning work in phoneme recognition and music genre classification (Andén and Mallat, 2014), and more recently, in the detection of speech impairments based on speech signals (Lauraitis et al., 2020). WTS representations of each vocalization sample were computed using the `scatteringTransform` function on MATLAB, averaged across WTS transformations on consecutive 1-s chunks of the raw time series. In the WTS transformation, the acoustic signal was decomposed by filtering the time series signal using a constant-Q wavelet filter bank. Different layers of wavelet convolution transform the signal into scattering coefficients consistent of multiple orders. The second-order scattering coefficients (WTS2), representing larger-scale acoustic structures like amplitude and frequency modulation (Andén and Mallat, 2014), were taken as features.

Second, for *spectral envelope* features: two types of acoustic features were derived from all vocalization samples to comprehensively capture acoustic properties representing vocal configurations:

1. The mel-frequency cepstral coefficients (MFCC) are cepstral representations of the audio sample that concisely describe the overall shape of a spectral envelope as perceived by human. While the MFCC has been the state-of-the-art of speech recognition, representing configurations of the vocal tract in speech, it has also been used to represent configurations of the vocal tract across other mammalian species, including primates (Fedurek et al., 2016). The MFCC is also a good representation of the syringeal properties of birds (Cheng et al., 2012). We derived the MFCC using the `mfcc` function of the Audio Toolbox in MATLAB, with analysis windows that spanned 50 ms and overlapped with adjacent analysis windows for 25 ms. This function first took the spectrum of the data in each analysis window using the Fourier transform, and then filtered the powers of the spectrum through a mel filter bank, linearly spaced across the first 10 triangular filters and logarithmically spaced in the remaining filters. The amplitude of the discrete cosine transform of the logged mel-transformed spectral powers were taken as the MFCC, and concatenated across all analysis windows for each vocalization sample.
2. The fist-order scattering coefficients (WTS1) were features derived from WTS representations described in the previous section; these capture the spectral envelope of sounds which are related to segmental features (i.e., consonants and vowels) (Andén and Mallat, 2014).

Third, for *pitch* features, fundamental frequency (f0) contour for each vocalization sample were derived to represent how pitch varies across the duration of the vocalization. For each vocalization sample, a raw f0 contour was first derived using the `pitch` function of the Audio Toolbox in MATLAB. f0 values of the contour were estimated using a Normalized Correlation Function algorithm (Atal, 1972), with analysis windows that spanned 50 ms and overlapped with adjacent analysis windows for 25 ms, and were taken as pitch features.

2.3. Machine Learning Classification Pipeline

A total of four sets of classification models were performed, each designed to classify vocalizations that do (+cognition) and do not (−cognition) support object categorization, from the vantage point of 4-month-old English-acquiring infants (Ferry et al., 2010, 2013; Perszyk and Waxman, 2016, 2019; Woodruff Carr et al., 2021b). We first performed classifications using all classes of features combined together in a single inclusive model (*full* model). Performance of the full model will identify whether these acoustic properties distinguish vocalizations that support infant cognition from those that do not. We then performed three more specific classifications, each using one of the three feature classes (i.e., spectral envelope, rhythmic, or pitch features). Performance of these models will identify which classes of acoustic features, if any, successfully distinguish vocalizations that support cognition from those that do not.

Since the number of vocalizations varied across types (see **Table 1**), a Monte Carlo cross-validation (MCCV) procedure was performed to avoid imbalanced classification. The MCCV involved an undersampling procedure which randomly selected 120 vocalization samples each from those that do (English, German, and Lemur vocalizations) and do not (Cantonese and Zebra Finch vocalizations) support cognition (i.e., a total of 240 samples). Each type of vocalizations was represented equally in the two classes (i.e., 60 Cantonese, 60 Zebra Finch vocalization, 40 English, 40 German, and 40 lemur vocalization samples). The 240 samples were then split into training and testing sets with stratified sampling in a 75:25 ratio. The MCCV also allowed us to minimize optimistic bias in the classification (Raschka, 2018) so as to objectively evaluate its performance. 100 iterations of MCCV were performed.

In each iteration of MCCV, a principal component analysis (PCA) was first performed on the input acoustic features of the particular model, so as to reduce the dimensionality of the data. PCA was performed only on the training set to avoid data leakage. Principal component (PC) scores that collectively explain 95% of total variance of the training set was selected as training features for subsequent classification, whereas acoustic features from the test set were separately transformed into PC scores using the transformation matrix of the PCA results.

Classification was then performed using an ensemble modeling approach of ML, which selected the optimal classifier for the particular MCCV sub-sample, out of a classifier array of: (1) LASSO, (2) decision tree (DT), (3) support vector machine (SVM), (4) ridge regression (Ridge), and (5) Naïve Bayes (NBC). The selection of the optimal classifier was performed using a nested four-fold cross-validation procedure, which further divided the test set into four-folds. The five classifiers were trained using three out of the four-folds of the data to classify vocalizations that do and do not support cognition, while being blind to the actual specific vocalization type (i.e., language or non-human species). The training of these classifiers was then validated on the remaining fold. The process was repeated four times until all four-folds were validated. To maximize classification performance, hyper-parameter tuning for each classifier was also performed during the same nested cross-validation procedure using a grid search approach, which repeated the training and validation using all combinations of the following hyper-parameters: (1) LASSO (λ:{0.1,1,10,100}); (2) DT (minimum leaf size, 10 intervals in the log-scaled range between 1 and 67); (3) SVM (C: {0.01,0.1,1,10}; Kernel: {linear, rbf}); (4) Ridge (λ:{0.1,1,10,100}); (5) NBC: normal, kernel NBC). The combination of classifier and hyper-parameters which achieved the highest accuracy on the validation across the four-folds were selected as optimal. The optimal classifier and hyper-parameters were then used for training on the whole training set, and were then used to predict the labels of the test set. Based on such prediction, metrics of classification performance were computed, namely (1) Area Under the Curve (AUC) of a Receiver Operating Characteristics curve, (2) prediction accuracy (ACC), (3) sensitivity, and (4) specificity. Overall performance of each model was computed by averaging the AUC, ACC, sensitivity, and specificity values of all 100 MCCV iterations.

TABLE 1 | Descriptive statistics of dataset for vocalizations that do (+) and do not (–) support object categorization, from the vantage point of 4-month-old English-acquiring infants.

	Vocalization	Label	*n*	Duration (s): Mean (SD)
Human	English	+	703	1.23 (0.78)
	German	+	369	2.62 (1.95)
	Cantonese	–	1,634	1.94 (0.99)
Non-human	Lemur	+	122	1.55 (0.48)
	Finch	–	369	9.54 (4.59)

Schema of the MCCV and nested cross-validation procedure is visualized in **Figure 1**.

Performance of each model was further evaluated using a permutation approach, which involved randomizing the classification labels (+cognition vs. −cognition) while repeating the classification 1,000 times in each of the 100 MCCV sub-samples. The percentage of AUC values across all 100,000 permutations (1,000 randomizations × 100 MCCV iterations) which was equal to or higher than the actual mean AUC value was taken as the *p*-value of the model.

All machine learning procedures were performed in MATLAB, using classifier and hyperparameter tuning, and PCA functions provided by the *Statistics and Machine Learning Toolbox*.

3. RESULTS

Classification metrics are presented in **Table 2**. **Figure 2** presents the confusion matrices on the percentage of each type of vocalizations being classified as those which do and do not support object categorization across the four models, from the vantage point of 4-month-old English-acquiring infants.

The *full* model performed successful classifications, achieving an AUC of 0.9030, ACC of 0.8913, sensitivity of 0.8937, and specificity of 0.8890. It also achieved statistical significance, as per the permutation test on AUC ($p < 0.001$). This is consistent with the possibility that there are acoustic properties, present at the surface among human language and non-human vocalizations, that contribute to the identification of candidate links to cognition.

The *rhythmic* model achieved robust classifications, with an AUC of 0.9939, ACC of 0.9682, sensitivity of 0.9717, and specificity of 0.9647. It was statistically significant, as per the permutation test on AUC ($p < 0.001$).

Classifications in the *spectral envelope* model were also robust, achieving an AUC of 0.9955, ACC of 0.9807, sensitivity of 0.9827, and specificity of 0.9787. Its AUC value also achieved statistical significance ($p < 0.001$).

These results may suggest that both rhythmic and spectral envelope features are among acoustic

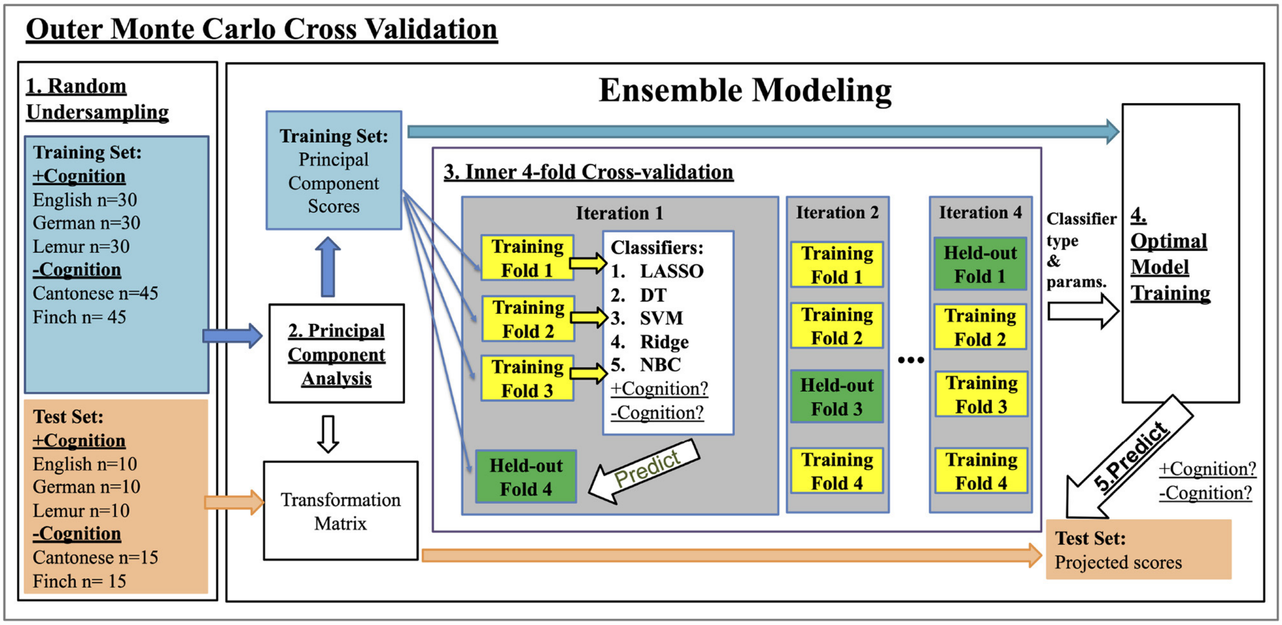


FIGURE 1 | Nested Cross Validation Schema: training classifiers to classify +cognition vs. -cognition vocalizations. Step 1: In each of the 100 outer Monte Carlo Cross Validation iterations, we used an undersampling procedure to randomly select a total of 240 vocalizations from those that do (+cognition) and do not (-cognition) support cognition. Each type of vocalization was represented equally within the two classes (i.e., 60 Cantonese, 60 Zebra Finch vocalization, 40 English, 40 German, and 40 lemur vocalization samples). The 240 samples were then split into training and test sets with stratified sampling in a 75:25 ratio. Step 2: A principal component analysis (PCA) was performed on the input acoustic features (see Section 2.2) of the training set. Principal components (PCs) with scores that collectively explain 95% of total variance of the training set were selected as training features. Acoustic features from the test set were transformed into PC scores using the PCA transformation matrix calculated on the training data only. Step 3. An inner four-fold cross-validation procedure was performed to selected the optimal classifier type and parameters. We divided the training set into four-folds, and trained different combinations of classifier type and parameters using three out of the four folds of the data. The resulting models were validated on the remaining held-out fold. The process was repeated for four iterations with a different heldout fold each time. The combination of classifier type and parameters that achieved the highest accuracy in the inner cross-validation were selected as optimal. Step 4: The optimal classifier and parameters were then used for training on the whole training set. Step 5. The model from Step 4 was validated by predicting the +cognition or -cognition labels of the test set from step 1, after being transformed in step 2. The numbers calculated in Step 5 are reported in **Table 2**.

TABLE 2 | Classification results, expressed as median area-under-the-curve (AUC) values, Sensitivity, Specificity, and Accuracy for each model.

Model	AUC	Sensitivity	Specificity	Accuracy
Full	0.9030***	0.8937	0.8890	0.8913
Rhythmic	0.9939***	0.9717	0.9647	0.9682
Spectral envelope	0.9955***	0.9827	0.9787	0.9807
Pitch	0.6287***	0.6703	0.5093	0.5898

****$p < 0.001$ in permutation test.*

properties shared by human languages and non-human vocalizations which may be identified as candidate links to cognition.

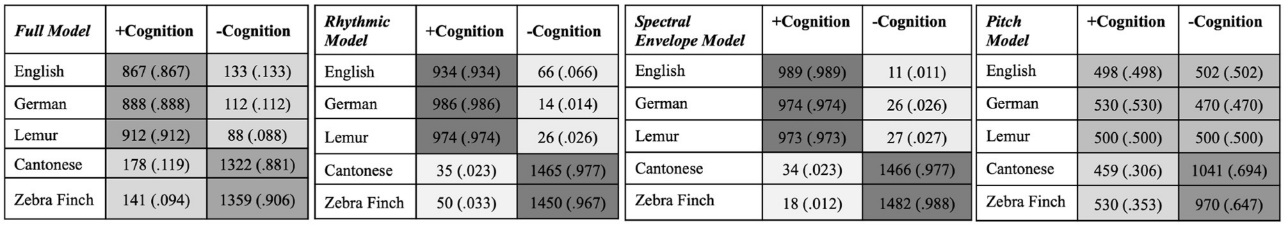
In contrast, although the pitch model also achieved statistical significance ($p < 0.001$), its AUC of 0.6287 is indicative of "poor classification" (Hosmer et al., 2013). Its poor performance is also indicated by its near chance-level ACC (0.5898) and specificity (0.5093), although its sensitivity (0.6703) is slightly above chance. These results challenge our prediction that pitch may also play a role in identifying candidate links to cognition among human languages and non-human vocalizations.

4. DISCUSSION

The current study was designed to harness the power of a supervised ML approach to address a fundamental developmental question: Which acoustic features, if any, are available on the surface of human and non-human vocalizations to support very young infants in identifying which vocalizations might serve as candidate links to cognition. Focusing on three classes of acoustic information (rhythmic, spectral envelope, and pitch), we asked (a) whether ML models could be trained to perform classifications that reliably distinguish vocalizations that support cognition from those that do not, and (b) whether rhythm or other any other class(es) of acoustic information was sufficient to support that classification.

4.1. Full Model

Consider first, the performance of full model. This model, which used rhythmic, spectral envelope, and pitch features combined, successfully classified vocalizations that support infant cognition from those that do not. This success held up both for human languages and non-human vocalizations. It should be noted that our models were supervised to utilize just *some* of the acoustic features, if any, that are common among



Full Model	+Cognition	-Cognition
English	867 (.867)	133 (.133)
German	888 (.888)	112 (.112)
Lemur	912 (.912)	88 (.088)
Cantonese	178 (.119)	1322 (.881)
Zebra Finch	141 (.094)	1359 (.906)

Rhythmic Model	+Cognition	-Cognition
English	934 (.934)	66 (.066)
German	986 (.986)	14 (.014)
Lemur	974 (.974)	26 (.026)
Cantonese	35 (.023)	1465 (.977)
Zebra Finch	50 (.033)	1450 (.967)

Spectral Envelope Model	+Cognition	-Cognition
English	989 (.989)	11 (.011)
German	974 (.974)	26 (.026)
Lemur	973 (.973)	27 (.027)
Cantonese	34 (.023)	1466 (.977)
Zebra Finch	18 (.012)	1482 (.988)

Pitch Model	+Cognition	-Cognition
English	498 (.498)	502 (.502)
German	530 (.530)	470 (.470)
Lemur	500 (.500)	500 (.500)
Cantonese	459 (.306)	1041 (.694)
Zebra Finch	530 (.353)	970 (.647)

FIGURE 2 | Confusion Matrices: classification of English, German, Lemur, Cantonese, and Zebra Finch vocalizations into classes of vocalizations that do (+Cognition) and do not (-Cognition) support object categorization.

vocalizations that support cognition to perform classification. Therefore, the success of the classification does not implicate that English and German resemble lemur vocalizations more than Cantonese *overall* acoustically. Instead, this successful classification, especially in the face of the considerable acoustic variability across these vocalizations, suggests there are indeed *some* common acoustic features, available on the surface of human and non-human vocalizations which support very young infants in identifying which vocalizations might serve as candidate links to cognition.

We turn next to test which class(es) of acoustic properties might best signal these candidate links.

4.2. Rhythmic Model

The rhythmic model, like the full model, achieved robust classification, successfully distinguishing vocalizations that do, and do not, support cognition from the vantage point of a 4-month-old English-acquiring infant (i.e., English, German, and lemur vocalizations vs. Cantonese and zebra finch vocalizations).

This outcome is consistent with robust evidence of the importance of rhythmic properties in human languages and non-human vocalizations. It also mirrors the behavioral evidence regarding infants' earliest links to cognition (Perszyk and Waxman, 2019).

Especially intriguing is that the new evidence, reported here, is consistent with proposals of parallels between rhythmic features instrumental to both human and non-human vocalizations (Ramus et al., 2000; Tincoff et al., 2005; Ravignani et al., 2019). From an acoustic perspective, non-human animals' sensitivity to rhythm is well-documented (Ravignani et al., 2019). Moreover, parallels in "babbling" of infant bats and humans suggest that rhythmic motor activity may be foundation for basic rhythmic structures across mammalian vocalizations (Knörnschild et al., 2006; Ravignani et al., 2019). In addition, human and non-human animals alike demonstrate neural entrainment to rhythm in vocalizations (Patel et al., 2009; Schachner et al., 2009). In humans, these neural oscillations are essential to identifying linguistic structure (Poeppel and Assaneo, 2020). Neural entrainment appears to be subserved by the frontostriatal brain circuitry in both humans and non-human animals, suggesting that it is not language specific (Kotz et al., 2018). This observation raises the intriguing possibility that for infants as young as 3- or 4-months of age, who cannot yet parse individual words from the ongoing sound stream, rhythm provides an entry point for identifying candidate links to cognition by establishing an early template according to infant's native rhythmic properties. Auditory signals that may conform to this native rhythmic template, such as speech from rhythmically similar foreign languages or even non-human vocalizations, may therefore be initially linked to cognition.

Indeed, we suspect that this early native rhythmic template may engage attentional mechanisms in such a way as to support infants' precocious language-cognition link. There is considerable evidence that rhythm engages infant attention (Jusczyk et al., 1993; Sansavini et al., 1997; Höhle et al., 2009; Räsänen et al., 2018) and that attention to speech rhythm is crucial to the acquisition of language (Gervain et al., 2021), highlighting distinct linguistic elements and relations among them (Soderstrom, 2007; Spinelli et al., 2017). But even more to the point, listening to their native language and to lemur vocalizations engages infants' attention neurally, as indexed by 4–9 Hz neural oscillatory activities (Woodruff Carr et al., 2021a). This rhythm-sensitive heightened attention may be a mechanism that supports infants' identification of which signals are candidate links to cognition. Additional work is required to clarify how attentional mechanisms and rhythmic properties guide infants as they discover the language-cognition link.

4.3. Spectral Envelope Model

The results of the ML model reported here suggest that information in the spectral envelope also yielded robust classifications. This outcome, although unanticipated, suggests that spectral envelope properties successfully classified vocalizations that support infant cognition from those that do not. This is interesting because spectral envelope features richly represent acoustic properties of speech segments (Mogran et al., 2004; Andén and Mallat, 2014) that young infants may not yet represent. Infants' sensitivity to spectral properties appears to emerge later than their sensitivity to rhythmic features (Kuhl and Rivera-Gaxiola, 2008; Werker, 2018). Thus, the current ML results may best be interpreted to suggest that spectral envelope features, whenever they do become available to infants, may be among those infants use to identify candidate links to cognition.

The success of the spectral envelope model in classifying the non-human vocalizations is not unexpected. Spectral envelope features represent vocal configurations across species (Mogran et al., 2004; Fedurek et al., 2016). For example, the physiologic distinction laryngeal (human and non-human primates) and syringeal (birds) vocalizations may be represented acoustically in spectral envelope features in the model. This raises an intriguing possibility: that infants' earliest links to cognition reflect an evolutionarily ancestral route, one that confers cognitive advantage through primate-general attentional mechanisms (Perszyk and Waxman, 2018).

The surprising success of the spectral envelope model certainly opens new avenues for investigation. For example, in future work, it will be important to assess whether lemur vocalizations have the same facilitative effect on categorization in infants acquiring languages, like Cantonese, with both segmental inventories (hence spectral envelope features) and speech rhythm that differ systematically from those of English. Meanwhile, it remains an open question whether there are other aspects of spectral envelope properties potentially common between Cantonese and lemur vocalizations, both as mammalian laryngeal vocalizations. One intriguing possibility is that lemur vocalizations do confer some cognitive advantage to Cantonese-acquiring infants, but perhaps less robustly without lemur vocalizations conforming to the rhythmic template of Cantonese. Delineating these possibilities would further shed light on the mechanistic nature of the pathway that enables infant's earliest links to cognition.

4.4. Pitch Model

Pitch features, which like rhythm are also related to prosody, yielded surprisingly low classification performance. This suggests that there may be few, if any, surface pitch-relevant acoustic properties that distinguish between vocalizations that do and do not support cognition, despite that pitch is one of the most prominent features of infant-direct speech (Hilton et al., 2022) also known to engage infant's attention (Sullivan and Horowitz, 1983). There are several possible accounts for this outcome. First, it may be related to the broad acoustic variability in our corpus. After all, lemur vocalizations have higher average pitch and broader pitch range than human vocalizations (Woodruff Carr et al., 2021b). Alternatively, this may reflect a limitation more particular to our corpus. We were only able to capture the f0 contour to represent speech intonation and pitch properties of vocalization in the current models. As a result, we may have failed to capture the more dynamic intonational properties of these signals. The limited amount of information represented in the f0 contour as compared to rhythmic and spectral envelope features may also have hindered classification performance from a computational perspective. Addressing this question will require additional work that incorporates a broader and more dynamic set of measures that tap into more fine-grained vocalic properties of both human and non-human vocalizations.

4.5. Limitations and Future Directions

The ML approach invoked here suggests that there are indeed certain acoustic properties, present in the surface of human and non-human vocalizations, that are available, *in principle*, to support infants' identification of which vocalizations link to cognition.

This outcome, important in itself, raises new questions for future work. For example, it will be important to discover whether, as infants forge their earliest links to cognition, they use the same mechanisms, or different ones, in identifying candidate human languages and non-human vocalizations. There is reason to suspect that there may be two distinct routes, one governing the links from language and another governing the candidate links from non-human vocalizations (Owren et al., 2011; Perszyk and Waxman, 2018). First, cross-species neurophysiological work has identified two neural pathways in response to human vocal communication: a subcortical pathway shared among human and non-human primates for affective vocalizations, and another cortical pathway that appears to be specific to humans for speech (Owren et al., 2011; Ackermann et al., 2014). Second, neural and behavioral evidence from 4- to 6-month-old English-acquiring infants is consistent with the possibility that there are two distinct routes (Ferry et al., 2013; Perszyk and Waxman, 2019; Woodruff Carr et al., 2021b). To examine this hypothesis, it will fruitful for future studies to apply a ML approach separately on human languages and non-human vocalizations, as well as modeling from the vantage points of infants acquiring a language other than English. Doing so will not only merely require a larger database of human and non-human vocalizations, but crucially broader empirical behavioral evidence delineating natural classes of human and non-human vocalizations that do and do not support cognition, from the vantage points of infants acquiring different varieties of languages.

While future modeling would benefit from an expansion of empirical evidence, results of the current model nevertheless shed light on future directions of empirical studies on infants' language-cognition link. Indeed, the features identified in the current models may not represent veridically the acoustic features actually utilized by infants as they evaluate candidate links to cognition. Nevertheless, future studies could target rhythm and spectral envelope features to manipulate in the stimuli in object categorization experiments (e.g., testing with low-pass filtered vocalizations or speech chimera), so as to pinpoint acoustic properties infants utilize to evaluate candidate links to cognition empirically. Further, by testing vocalizations of a larger variety of mammalian and non-human primate species, future studies could also shed light on the extent to which the link governing non-human vocalizations and cognition in young infants is modulated by the etiological distance between the animal and humans, so as to examine the hypothesis that the link governing non-human vocalizations and cognition is an ancestral pathway that reflects the residue of evolution (Perszyk and Waxman, 2018).

5. CONCLUSION

The current results offer support for the proposal that rhythmic and spectral envelope features, available in the input of human language and of non-human linguistic vocalizations, may guide infants in identifying which signals are candidate links to cognition. This in principle evidence, important in itself, is also consistent with the possibility that infants' earliest links to cognition may be subserved by their sensitivity to rhythmic and spectral envelope properties of sounds.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JL and SW conceptualized the study. JL and AF design machine learning model. JL implemented the model. All authors interpreted the data, wrote the paper, and approved it for publication.

FUNDING

This work was supported by National Institute of Health (NIH) Grant R01HD083310 to SW.

ACKNOWLEDGMENTS

We thank Patrick Wong, Luchang Wang, Samuel Mehr, Chris Mercer, and Baby Speech Lab at the University of Konstanz for providing us with vocalization samples. We also thank Kali Woodruff Carr and Elena Luchkina for providing comments on the study.

REFERENCES

Ackermann, H., Hage, S. R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci*. 6, 529–546. doi: 10.1017/S0140525X13003099

Andén, J., Lostanlen, V., and Mallat, S. (2015). "Joint time-frequency scattering for audio classification," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (Boston, MA: IEEE), 1–6.

Andén, J., and Mallat, S. (2014). Deep scattering spectrum. *IEEE Trans. Signal Process*. 62, 4114–4128. doi: 10.1109/TSP.2014.2326991

Atal, B. S. (1972). Automatic speaker recognition based on pitch contours. *J. Acoust. Soc. Am*. 52, 1687–1697. doi: 10.1121/1.1913303

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022). Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *Neuroimage* 247, 118698. doi: 10.1016/j.neuroimage.2021.118698

Belin, P. (2006). Voice processing in human and non-human primates. *Philos. Trans. R. Soc. B Biol. Sci*. 361, 2091–2107. doi: 10.1098/rstb.2006.1933

Cheng, J., Xie, B., Lin, C., and Ji, L. (2012). A comparative study in birds: call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. *Bioacoustics* 21, 157–171. doi: 10.1080/09524622.2012.669664

Christophe, A., Mehler, J., and Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy* 2, 385–394. doi: 10.1207/S15327078IN0203_6

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev*. 81, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Elliott, T. M., and Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput. Biol*. 5, e1000302. doi: 10.1371/journal.pcbi.1000302

Fedurek, P., Zuberbühler, K., and Dahl, C. D. (2016). Sequential information in a great ape utterance. *Sci. Rep*. 6, 1–11. doi: 10.1038/srep38226

Ferry, A. L., Hespos, S. J., and Waxman, S. R. (2010). Categorization in 3-and 4-month-old infants: an advantage of words over tones. *Child Dev*. 81, 472–479. doi: 10.1111/j.1467-8624.2009.01408.x

Ferry, A. L., Hespos, S. J., and Waxman, S. R. (2013). Nonhuman primate vocalizations support categorization in very young human infants. *Proc. Natl. Acad. Sci. U.S.A*. 110, 15231–15235. doi: 10.1073/pnas.1221166110

Gelman, S. A. (2004). Psychological essentialism in children. *Trends Cogn. Sci*. 8, 404–409. doi: 10.1016/j.tics.2004.07.001

Gervain, J., Christophe, A., and Mazuka, R. (2021). "Prosodic bootstrapping," in *The Oxford Handbook of Language Prosody*, eds C. Gussenhoven and A. Chen (New York, NY: Oxford University Press), 563–573. doi: 10.1093/oxfordhb/9780198832232.013.36

Gervain, J., and Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nat. Commun*. 4, 1–6. doi: 10.1038/ncomms2430

Gleitman, L., and Wanner, E. (1982). "The state of the state of the art," in *Language Acquisition: The State of the Art*, eds E. Wanner and L. Gleitman (New York, NY: Cambridge University Press), 3–48.

Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Ann. N. Y. Acad. Sci*. 1453, 67–78. doi: 10.1111/nyas.14137

Goswami, U., and Leong, V. (2016). "Speech rhythm and temporal structure: converging perspectives?" in *Linguistic Rhythm and Literacy* (John Benjamins), 111–132. doi: 10.1075/tilar.17.06gos

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech–a syllable-centric perspective. *J. Phonet*. 31, 465–485. doi: 10.1016/j.wocn.2003.09.005

Hilton, C. B., and Goldwater, M. B. (2021). Linguistic syncopation: meter-syntax alignment affects sentence comprehension and sensorimotor synchronization. *Cognition* 217, 104880. doi: 10.1016/j.cognition.2021.104880

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., et al. (2022). Acoustic regularities in infant-directed speech and song across cultures. *bioRxiv*. doi: 10.1101/2020.04.09.032995

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev*. 32, 262–274. doi: 10.1016/j.infbeh.2009.03.004

Hosmer, D. W. Jr., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression, Vol. 398*. Hoboken, NJ: John Wiley and Sons. doi: 10.1002/9781118548387

Isomura, T., Parr, T., and Friston, K. (2019). Bayesian filtering with multiple internal models: toward a theory of social intelligence. *Neural Comput*. 31, 2390–2431. doi: 10.1162/neco_a_01239

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang*. 44, 548–567. doi: 10.1006/jmla.2000.2755

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of english words. *Child Dev*. 64, 675–687. doi: 10.1111/j.1467-8624.1993.tb02935.x

Knörnschild, M., Behr, O., and von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). *Naturwissenschaften* 93, 451–454. doi: 10.1007/s00114-006-0127-9

Kotz, S., Ravignani, A., and Fitch, W. (2018). The evolution of rhythm processing. *Trends Cogn. Sci*. 22, 896–910. doi: 10.1016/j.tics.2018.08.002

Kuhl, P., and Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu. Rev. Neurosci*. 31, 511–534. doi: 10.1146/annurev.neuro.30.051606.094321

Laboratory of Vocal Learning at Hunter College. (2015). *Zebra finch song library 2015*. Available online at: http://ofer.hunter.cuny.edu (accessed September 23, 2019).

Langus, A., Mehler, J., and Nespor, M. (2017). Rhythm in language acquisition. *Neurosci. Biobehav. Rev*. 81, 158–166. doi: 10.1016/j.neubiorev.2016.12.012

Lauraitis, A., Maskeliūnas, R., Damaševičius, R., and Krilavičius, T. (2020). Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features. *IEEE Access*. 8, 96162–72 doi: 10.1109/ACCESS.2020.2995737

Lecanuet, J. -P., and Granier-Deferre, C. (1993). "Speech stimuli in the fetal environment," in *Developmental Neurocognition: Speech and Face*

Processing in the First Year of Life (Dordrecht: Springer), 237–248. doi: 10.1007/978-94-015-8234-6_20

May, L., Byers-Heinlein, K., Gervain, J., and Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Front. Psychol.* 2, 222. doi: 10.3389/fpsyg.2011.00222

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition* 29, 143–178. doi: 10.1016/0010-0277(88)90035-2

Mercer, C. (2012). *The Audible Phylogeny of Lemurs*.

Miller, G. A. (1990). The place of language in a scientific psychology. *Psychol. Sci.* 1, 7–14. doi: 10.1111/j.1467-9280.1990.tb00059.x

Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., and Sereno, J. (2017). Fetal rhythm-based language discrimination: a biomagnetometry study. *Neuroreport* 28, 561. doi: 10.1097/WNR.0000000000000794

Mogran, N., Bourlard, H., and Hermansky, H. (2004). "Automatic speech recognition: An auditory perspective," in *Speech Processing in the Auditory System*. (New York, NY: Springer), 309–338. doi: 10.1007/0-387-21575-1_6

Morgan, J. L., and Demuth, K. (1996). "Signal to syntax: an overview," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah, NJ: Lawrence Erlbaum Associates), 13–34. doi: 10.4324/9781315806822

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756. doi: 10.1037/0096-1523.24.3.756

Nazzi, T., Nelson, D. G. K., Jusczyk, P. W., and Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: effects of prosodic well-formedness. *Infancy* 1, 123–147. doi: 10.1207/S15327078IN0101_11

Nooteboom, S. (1997). The prosody of speech: melody and rhythm. *Handb. Phonet. Sci.* 5, 640–673.

Owren, M. J., Amoss, R. T., and Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. *Am. J. Primatol.* 73, 530–544. doi: 10.1002/ajp.20913

Patel, A., Iversen, J., Bregman, M., and Schulz, I. (2009). Studying synchronization to a musical beat in nonhuman animals. *Ann. N. Y. Acad. Sci.* 1169, 459–469. doi: 10.1111/j.1749-6632.2009.04581.x

Peña, M., Pittaluga, E., and Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3823–3828. doi: 10.1073/pnas.0914326107

Perszyk, D. R., and Waxman, S. R. (2016). Listening to the calls of the wild: the role of experience in linking language and cognition in young infants. *Cognition* 153, 175–181. doi: 10.1016/j.cognition.2016.05.004

Perszyk, D. R., and Waxman, S. R. (2018). Linking language and cognition in infancy. *Annu. Rev. Psychol.* 69, 231–250. doi: 10.1146/annurev-psych-122216-011701

Perszyk, D. R., and Waxman, S. R. (2019). Infants' advances in speech perception shape their earliest links between language and cognition. *Sci. Rep.* 9, 1–6. doi: 10.1038/s41598-019-39511-9

Poeppel, D., and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nat. Rev. Neurosci.* 21, 322–334. doi: 10.1038/s41583-020-0304-4

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351. doi: 10.1126/science.288.5464.349

Räsänen, O., Kakouros, S., and Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising?-linking properties of ids to statistical learning and attention at the prosodic level. *Cognition* 178, 193–206. doi: 10.1016/j.cognition.2018.05.015

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*. doi: 10.48550/arXiv.1811.12808

Ravignani, A., Dalla Bella, S., Falk, S., Kello, C. T., Noriega, F., and Kotz, S. A. (2019). Rhythm in speech and animal vocalizations: a cross-species perspective. *Ann. N. Y. Acad. Sci.* 1453, 79. doi: 10.1111/nyas.14166

Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33, 3. doi: 10.1037/0012-1649.33.1.3

Schachner, A., Brady, T. F., Pepperberg, I. M., and Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836. doi: 10.1016/j.cub.2009.03.061

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27, 501–532. doi: 10.1016/j.dr.2007.06.002

Spinelli, M., Fasolo, M., and Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Dev. Rev.* 44, 1–18. doi: 10.1016/j.dr.2016.12.001

Sullivan, J. W., and Horowitz, F. D. (1983). The effects of intonation on infant attention: the role of the rising intonation contour. *J. Child Lang.* 10, 521–534. doi: 10.1017/S0305000900005341

Tchernichovski, O., Mitra, P. P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291, 2564–2569. doi: 10.1126/science.1058522

Tilsen, S., and Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *J. Acoust. Soc. Am.* 134, 628–639. doi: 10.1121/1.4807565

Tilsen, S., and Johnson, K. (2008). Low-frequency fourier analysis of speech rhythm. *J. Acoust. Soc. Am.* 124, EL34-EL39. doi: 10.1121/1.2947626

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., and Mehler, J. (2005). The role of speech rhythm in language discrimination: further tests with a non-human primate. *Dev. Sci.* 8, 26–35. doi: 10.1111/j.1467-7687.2005.00390.x

Wang, L., Kalashnikova, M., Kager, R., Lai, R., and Wong, P. C. (2021). Lexical and prosodic pitch modifications in cantonese infant-directed speech. *J. Child Lang.* 48, 1235–1261. doi: 10.1017/S0305000920000707

Waxman, S. R., and Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13, 258–263. doi: 10.1016/j.tics.2009.03.006

Waxman, S. R., and Markow, D. B. (1995). Words as invitations to form categories: evidence from 12-to 13-month-old infants. *Cogn. Psychol.* 29, 257–302. doi: 10.1006/cogp.1995.1016

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Appl. Psycholinguist.* 39, 703–728. doi: 10.1017/S0142716418000152

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Woodruff Carr, K., Perszyk, D. R., Norton, E. S., Voss, J. L., Poeppel, D., and Waxman, S. R. (2021a). Developmental changes in auditory-evoked neural activity underlie infants' links between language and cognition. *Dev. Sci.* 24, e13121 doi: 10.1111/desc.13121

Woodruff Carr, K., Perszyk, D. R., and Waxman, S. R. (2021b). Birdsong fails to support object categorization in human infants. *PLoS ONE* 16, e0247430. doi: 10.1371/journal.pone.0247430

Zahner, K., Schönhuber, M., Grijzenhout, J., and Braun, B. (2016). "Konstanz prosodically annotated infant-directed speech corpus (KIDS corpus)," in *Speech Prosody 2016* (Boston, MA: Speech Prosody), 562–566. doi: 10.21437/SpeechProsody.2016-115

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

frontiers | Frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 06 June 2022
doi: 10.3389/fpsyg.2022.897015

The Components of Interpersonal Synchrony in the Typical Population and in Autism: A Conceptual Analysis

Claire Bowsher-Murray[1,2], Sarah Gerson[2], Elisabeth von dem Hagen[1,2,3]† and Catherine R. G. Jones[1,2]*†*

[1] Wales Autism Research Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom, [2] Cardiff University Centre for Human Developmental Science, School of Psychology, Cardiff University, Cardiff, United Kingdom, [3] Cardiff University Brain Imaging Research Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom

Interpersonal synchrony – the tendency for social partners to temporally co-ordinate their behaviour when interacting – is a ubiquitous feature of social interactions. Synchronous interactions play a key role in development, and promote social bonding and a range of pro-social behavioural outcomes across the lifespan. The process of achieving and maintaining interpersonal synchrony is highly complex, with inputs required from across perceptual, temporal, motor, and socio-cognitive domains. In this conceptual analysis, we synthesise evidence from across these domains to establish the key components underpinning successful non-verbal interpersonal synchrony, how such processes interact, and factors that may moderate their operation. We also consider emerging evidence that interpersonal synchrony is reduced in autistic populations. We use our account of the components contributing to interpersonal synchrony in the typical population to identify potential points of divergence in interpersonal synchrony in autism. The relationship between interpersonal synchrony and broader aspects of social communication in autism are also considered, together with implications for future research.

Keywords: interpersonal synchrony, behavioural co-ordination, social interaction, synchronisation, social motor behaviour, autism

Edited by:
Stefanie Hoehl,
University of Vienna, Austria

Reviewed by:
Bahar Tuncgenc,
Nottingham Trent University,
United Kingdom
Stefania Conte,
University of South Carolina,
United States

****Correspondence:***
Claire Bowsher-Murray
bowsher-murraycj@cardiff.ac.uk
Catherine R. G. Jones
jonescr10@cardiff.ac.uk

† These authors share senior authorship

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *15 March 2022*
Accepted: *16 May 2022*
Published: *06 June 2022*

Citation:
Bowsher-Murray C, Gerson S, von dem Hagen E and Jones CRG (2022) The Components of Interpersonal Synchrony in the Typical Population and in Autism: A Conceptual Analysis. Front. Psychol. 13:897015. doi: 10.3389/fpsyg.2022.897015

INTRODUCTION

The tendency for social partners to temporally co-ordinate their behaviour, known as interpersonal synchrony (IS), is a common feature of social interactions (Bernieri et al., 1988; Delaherche et al., 2012). It is sometimes the product of conscious effort, such as when we shake hands, high five, or dance together. It may also arise spontaneously: social partners might fall into step (Zivotofsky and Hausdorff, 2007), align their postural positions (Shockley et al., 2003; Gaziv et al., 2017), or entrain their body movements (Hadar et al., 1984) or facial expressions (Louwerse et al., 2012). IS may display rhythmical properties (e.g., walking in step; nodding), but equally may be less structured in nature (e.g., sporadic gestures or postural adjustment). Although IS may arise *via* a broad range of behavioural processes including gesture, gaze, facial expression, speech, and vocalisation, the current review is focussed on the synchrony of non-verbal behaviours. Temporal co-ordination of social behaviour emerges shortly after birth (Condon and Sander, 1974; Dominguez et al., 2016) and becomes more temporally accurate, more complex, and less reliant on adult facilitation during infancy (Hilbrink et al., 2015; Meyer and Hunnius, 2020). Mother-child IS is believed to positively influence self-regulation and empathy, and to promote later cognitive, social and emotional development and secure attachment relationships (Harrist and Waugh, 2002; Feldman, 2007; Evans and Porter, 2009). Throughout the lifespan, IS serves as a social signifier and promotes various social outcomes, including increased affiliation (Hove and Risen, 2009; Tunçgenç et al., 2015), rapport (Vacharkulksemsuk and Fredrickson, 2011), bonding (Tarr et al., 2015; Tunçgenç and Cohen, 2016), helping (Tunçgenç and Cohen, 2018), and co-operation (Rabinowitch and Meltzoff, 2017). Such effects are present both when IS is spontaneous and when it is intentional, although there is mixed evidence as to whether they are enhanced when partners share an intention to co-ordinate (Reddish et al., 2013; Howard et al., 2021). The full range of social outcomes arising from IS has been documented in recent reviews (Rennung and Goritz, 2016; Vicaria and Dickens, 2016; Mogan et al., 2017; Cross et al., 2019; Hoehl et al., 2021).

Difficulties with social communication and social interaction, including with non-verbal communicative behaviour and building and maintaining relationships, are hallmarks of autism (American Psychiatric Association, 2013). As such, there has been particular interest in how autistic people engage in and experience IS (McNaughton and Redcay, 2020). Evidence indicates that IS is less accurate and/or less frequent in interactions involving autistic people, in both spontaneous (Marsh et al., 2013; Fitzpatrick et al., 2016; Kaur et al., 2018; Georgescu et al., 2020; Zampella et al., 2020) and intentional (Fitzpatrick et al., 2016) contexts. Several studies have found an association between lower levels of IS and higher levels of autistic traits (Brezis et al., 2017; Cheng et al., 2017; Fitzpatrick et al., 2017b; Zampella et al., 2020; Granner-Shuman et al., 2021; although cf. Kaur et al., 2018). There is also some evidence that the social significance of IS may be attenuated for many autistic people (Koehne et al., 2016).

IS emerges as a function of multiple mechanisms operating in concert with each other (Konvalinka et al., 2010; Delaherche et al., 2012; Mills et al., 2019; McNaughton and Redcay, 2020). However, research commonly focuses on the role of individual mechanisms, such as attention (e.g., Temprado and Laurent, 2004; Richardson et al., 2007), perceptual processing (e.g., Noel et al., 2018), motor behaviour (e.g., Hart et al., 2014; Monier and Droit-Volet, 2019) and social factors (e.g., Kirschner and Tomasello, 2009; Lumsden et al., 2012; Honisch et al., 2021). To better understand IS there is a need to synthesise findings from across the perceptual, sensorimotor, social and cognitive domains. Understanding how the component processes underlying IS operate together is also necessary for understanding why IS manifests differently in autism. In addition to core differences in social functioning, differences between autistic and typical populations have been observed across domains relevant to IS, including attention (Frazier et al., 2017; Hedger et al., 2020), temporal perception (e.g., Allman and Falter, 2015); perceptual processing (e.g., Feldman et al., 2018; Meilleur et al., 2020), and motor behaviour (e.g., Fournier et al., 2010). Therefore, characterising IS in autism necessarily requires a holistic understanding of how differences in functioning across multiple underlying processes operate together.

The first part of this conceptual analysis begins with a synthesis of the key component mechanisms that contribute to IS (**Figure 1**), including an exploration of how such processes interact, and the factors that may moderate their operation. Where useful, we additionally draw on the related concepts of imitation and joint action. Imitation, like IS, involves behavioural matching in form, albeit not in time (Hove and Risen, 2009; Catmur and Heyes, 2013). Joint action involves the conscious co-ordination of complimentary behaviour to achieve a shared goal (Sebanz et al., 2006). The parallels between these two phenomena and IS mean that their underlying processes can shed light on the role of equivalent processes in IS. The second part of this conceptual analysis describes how each of the identified component mechanisms operates in autism, and considers the extent to which differences in the functioning of these mechanisms may explain differences in IS. In considering each mechanism not only individually but also as part of a wider system, we aim to build an understanding of how relevant mechanisms collectively underpin reduced IS in autism.

PART 1: INTERPERSONAL SYNCHRONY IN THE TYPICAL POPULATION

Social Orienting

In order for an individual to synchronise their movements with an external stimulus, they must have perceptual access to the stimulus (Richardson et al., 2007; Oullier et al., 2008; Miyata et al., 2017; Oh Kruzic et al., 2020). Perception of the stimulus, in turn, requires that sufficient attentional resources are allocated to it (Aubin et al., 2021). In the case of IS, the relevant stimulus is one's interaction partner. During social interaction, the perception of one's partner and their movements is facilitated by an inherent tendency to orient toward social stimuli, that is, the inherent preferential allocation of visual attention to social rather than non-social cues (Fletcher-Watson et al., 2008; Gluckman and Johnson, 2013; Leppanen, 2016; Rösler et al., 2017). Experimental manipulation of the extent to which partners are oriented toward each other has demonstrated its importance in facilitating IS. Partners spontaneously synchronised their movements at above chance levels when looking directly at each other, but not when they had only peripheral visual access to each other (Richardson et al., 2007).

It is also important to consider what aspects of a social partner's behaviour are particularly informative. The direction of a partner's gaze and the kinematic cues provided by their limb movements convey information about their future movements (Sartori et al., 2011; Ansuini et al., 2014; Khoramshahi et al., 2016), particularly where the end of point of an action is otherwise unclear (Fulceri et al., 2018). Such cues are likely to facilitate both spontaneous and intentional IS. Mutual gaze between partners may also support the emergence of IS. Although there is no direct evidence that mutual gaze plays a role in IS, mutual gaze is considered to facilitate social engagement, and is known to promote spontaneous partner imitation (Wang and Hamilton, 2012; Prinsen et al., 2017). The shared element of

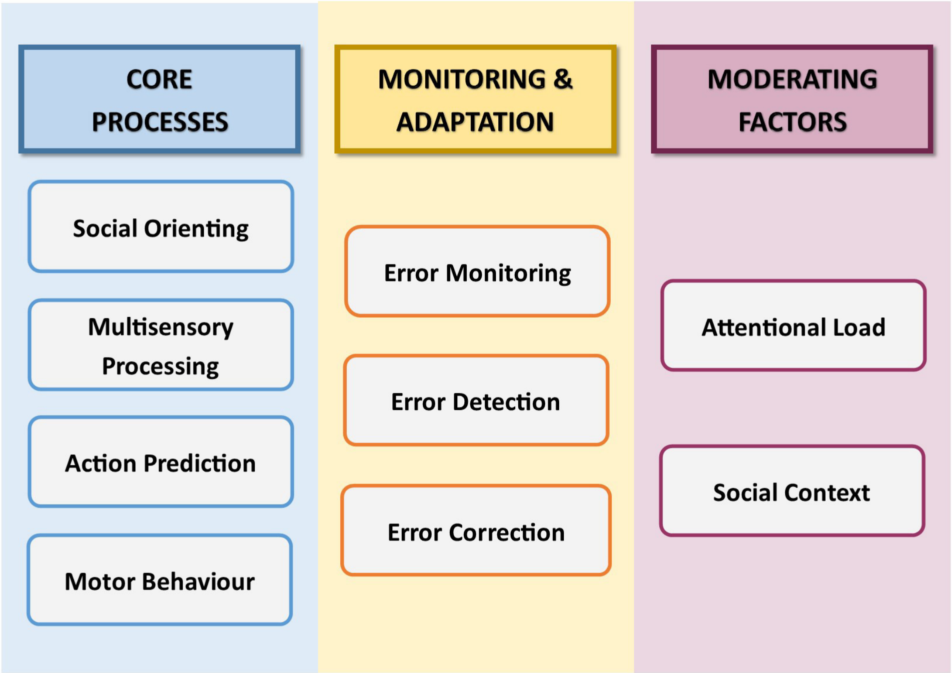


FIGURE 1 | Schematic illustration of the components of non-verbal interpersonal synchrony, as discussed in this review.

behavioural matching in IS and in imitation suggests that mutual gaze may have a similar role in promoting IS.

Flexibility in attending to social cues is another key consideration for effective IS. Naturalistic interactions almost always contain multiple social cues, occurring either simultaneously or in quick succession. To capitalise on their predictive value, it seems likely that attention must be flexibly directed and redirected toward relevant cues. A further consideration is that social cues often consist of limb or whole-body movements, which are inherently dynamic. For example, in order for a handshake to be successfully synchronised, each partner must match their eye movements to the movement of the other's hand, such that the motion of the other partner is efficiently tracked *via* smooth pursuit (Johnson et al., 2016). In summary, social orienting is a core process necessary for IS to take place. Dynamically attending to a partner's behavioural cues, particularly kinetic and eye gaze cues, is likely to be the specific means by which attending to a social partner supports IS.

Multisensory Processing

Interacting with a social partner is a multisensory experience, involving the integration of social information delivered *via* multiple modalities. Sensory inputs from two or more modalities that occur with sufficient temporal proximity are perceptually combined *via* multisensory integration (Alais et al., 2010; Murray et al., 2016). There is evidence that efficient multisensory integration supports IS. For example, synchronisation of both fine and gross motor activity was achieved with significantly greater accuracy when multisensory stimuli were available, as compared with when stimuli were unisensory only (Elliott et al., 2010; Su, 2014; Wright et al., 2014; Khan et al., 2020). Further, increased multisensory acuity was significantly associated with increased spontaneous IS (Noel et al., 2018). Multisensory integration of social information is likely to support IS in a number of ways. First, it likely enhances the perceptual cohesion and salience of the behaviour (Bahrick and Todd, 2012) with which synchrony is to be achieved. Second, the combination of information from multiple streams enables the individual to make enhanced statistical predictions about the stimulus (Elliott et al., 2010). Third, multisensory integration may promote perception action coupling, whereby audio-visual cues are likely to engage the observer's own sensorimotor system more than single modality cues, which then simulates the observed movement as a means of predicting its ongoing trajectory (Su, 2014). Fourth, in addition to supporting the processing of incoming social stimuli, multisensory integration is likely to support the effective planning, monitoring and adjustment of one's *own* actions, which requires the effective integration of visual and proprioceptive information about one's own motor behaviour (Foster, 2019). Overall, therefore, efficient multisensory processing is a core process that likely supports effective IS *via* multiple pathways.

Action Prediction

A defining feature of IS is that partners move together simultaneously. In order to achieve this, a partner's movements must be anticipated, both spatially and temporally (Sebanz and Knoblich, 2009; Meyer and Hunnius, 2020;

Gvirts Probolovski and Dahan, 2021), rather than merely reacted to. The tendency to make spontaneous, online predictions about the actions of others is present throughout the lifespan (Flanagan and Johansson, 2003; Reddy et al., 2013; Sebanz and Knoblich, 2021), and is likely influenced by low level kinematic information (Ansuini et al., 2014) and gaze cues (Khoramshahi et al., 2016), as well as higher-level cognitive processes, such as assessing a partner's likely intentions in light of other contextual information (Falck-Ytter, 2012). A number of studies suggest a role for action prediction in intentional IS. For example, participants who displayed more accurate temporal prediction in a solo task showed significantly greater accuracy and stability of synchronisation during a joint finger tapping task (Pecenka and Keller, 2011). Similarly, toddlers who made more temporally accurate predictions in an observation task moved with greater temporal stability during a joint action task with a partner (Meyer et al., 2015). Thus, the evidence suggests that being able to form accurate predictions about a partner's movements is a core process facilitating intentional IS. There is no direct evidence as to how action prediction abilities inform spontaneous IS. However, some form of prediction must necessarily occur in order for behaviour to become temporally aligned (Gvirts Probolovski and Dahan, 2021). Further research is required to establish exactly how predictions are made and integrated with other component processes in the context of spontaneous IS.

Motor Behaviour

In addition to anticipating the movement of a partner, it is necessary to plan and execute one's own complimentary movement sequence, both temporally and spatially. The ability to synchronise simple motor movements with an external stimulus is limited in young children, partly by a relatively limited capacity to adapt motor behaviour to the tempo of an external stimulus, but develops during childhood (Drake et al., 2000; McAuley et al., 2006; Monier and Droit-Volet, 2018) as a function of developing motor skills (Monier and Droit-Volet, 2019). Immature motor skills are therefore believed to limit young children's levels of IS (Trainor and Cirelli, 2015), although there is limited direct evidence as to the role of motor skills in IS in typical populations. In support of a role for motor planning in intentional IS, reduced performance on a motor planning task was significantly associated with reduced intentional IS in a hand movement task (Granner-Shuman et al., 2021). However, the contribution to IS of other aspects of motor abilities are yet unknown.

Just as motor skills are likely important to the achievement of IS, so too is the form of movement people tend to produce. People tend to display an "individual motor signature," which is a distinct and stable pattern of movement that is personal to the individual in terms of direction, range and velocity of movement (Richardson and Johnston, 2005; Hart et al., 2014; Słowiński et al., 2016). Some individual motor signatures convey more predictive information than others (Koul et al., 2016), which is likely to make them easier to synchronise with. Further, evidence suggests that partners with similar individual motor signatures are better at predicting the timing of each other's movements (Colling et al., 2014) and achieve a higher degree of co-ordination when moving together (Słowiński et al., 2016), relative to partners whose motor signatures are relatively dissimilar. As well as displaying individualised patterns of movement, people also tend to exhibit a preferred *pace* of movement, or spontaneous motor tempo (Delevoye-Turrell et al., 2014). Just as the (dis)similarity of partners' motor signatures influences the degree of co-ordination they achieve when interacting, it seems likely that those with relatively similar spontaneous motor tempos would achieve higher levels of IS than those with relatively dissimilar motor tempos. In sum, interacting partners' motor abilities, their natural movement patterns as well as, potentially, their relative pace of movement, all contribute to IS.

Monitoring and Adaptation

The component processes of IS have so far been considered independently. However, during dynamic real-world interactions involving both intentional and spontaneous IS, these processes are believed to be embedded together in a continuous feedback loop (Shamay-Tsoory et al., 2019; Gvirts Probolovski and Dahan, 2021). Specifically, it is proposed that predictions about a partner's movement and one's own plan to align with it are integrated into a forward model of the shared movement between partners. As motor commands are executed, "error monitoring" occurs, whereby both one's own and one's partner's actual motor output is compared to the forward model. "Error detection" occurs when either partner's actual movement does not match the generated prediction. Error detection precipitates "error correction," where the predictive model and movement plan are updated (Shamay-Tsoory et al., 2019).

The component processes of IS, described above, may contribute to the effectiveness of the feedback loop in a number of ways. For example, effective error monitoring is likely underpinned by continued social orienting and dynamic attendance to behavioural cues. Further, error monitoring, in the context of IS, consists of detecting asynchrony between the actions of partners. Thus, perceptual sensitivity to the temporal alignment of events likely contributes to the achievement of IS. Error correction is likely to draw on action prediction and motor abilities. The effectiveness of the updating process also depends on how quickly it occurs (Vishne et al., 2021). The faster the internal model and movement plan can be updated and implemented, the more closely aligned partners' behaviour will be over time. Efficient updating is likely to be critical in real-world social interactions, in which the form and speed of partners' movements change over time, placing persistent demands on adaptive mechanisms.

The process of continuous mutual adaptation during IS, described by the above account, is supported by behavioural evidence. For example, dyads required to synchronise their finger tapping adjusted the time between their taps in opposite directions to one another, on a tap-by-tap basis (Konvalinka et al., 2010), suggesting that each partner continuously accounted for the pace of the other and modified the pace of their own tapping accordingly. During more complex interactions, there is evidence that partners spontaneously adapt both the spatial (Sartori et al., 2009; Candidi et al., 2015) and temporal (Vesper et al., 2011) qualities of their movements, so as to make them more predictable to their partner. The result of partners' mutually adaptive behaviour is that they coalesce into a third movement pattern distinct from either of their individual motor signatures (Hart et al., 2014). Overall, this line of research emphasises that the perceptual and motor abilities of each partner operate within a dynamic context of bidirectional adjustment and adaptation.

Attentional Load

During a social interaction, attentional resources are subject to demands from multiple sources. For example, while partners process visual information about their partner they will also be processing the content of their conversation and making inferences about the other person's mental state (Westra and Nagel, 2021). There may also be input from distractors in the environment (e.g., an interesting visual display or an overheard conversation), sensory input (e.g., feeling too hot or too cold), or other unrelated thoughts. The distribution of attention across multiple stimuli can influence the extent of intentional IS. Participants asked to synchronise arm movements were more accurate in their synchrony when attending to the task, compared to attending to a simultaneous reaction time task, or sharing attention across tasks (Temprado and Laurent, 2004). The extent to which distractors are present during real life interactions might similarly moderate levels of IS. However, it is notable that participants in this research were explicitly instructed to direct their attention away from IS-relevant stimuli. In real-world interactions, social cues are preferentially attended to (see the section "Social Orienting") and processed preferentially even when not task-relevant (Lavie et al., 2003). Preferential processing of social stimuli may mean that intentional IS is relatively unaffected when distractors are present, or affected only when distractors are particularly salient.

From a different perspective, it has been proposed that spontaneous IS might itself arise as a means of minimising overall attentional load (Koban et al., 2019). When IS arises (either spontaneously or intentionally), a social partner's actions are relatively similar in time and form to one's own, such that they are easier to predict and require less effortful processing. This, in turn, is likely to mean that greater attentional resources are available for processing other stimuli. However, given that intentional IS is an effortful process whereas spontaneous IS is not, spontaneous and intentional IS may relate differently to attentional load (Aubin et al., 2021). Further research is required to determine how attentional load and IS influence each other each other during everyday interactions.

Social Context

A number of studies have considered the role of social context in synchronisation, with some finding evidence of greater synchronisation when participants synchronise with a social stimulus (such as another person) as compared with a non-social stimulus (such as a mechanical arm) (Kirschner and Tomasello, 2009; Honisch et al., 2021; Howard et al., 2021). One interpretation is that the very existence of a social context motivates individuals to synchronise (Kirschner and Tomasello, 2009; Yu and Myowa, 2021). An alternative interpretation is that the perceived engagement of a partner, rather than their mere presence, provides increased motivation to synchronise. This explanation is supported by evidence that participants synchronised more accurately with a social partner than with a non-social stimulus, even when a social partner was present in both conditions (Howard et al., 2021). Another possibility, however, is that social stimuli provide greater congruency with the action to be produced by the participant, relative to non-social stimuli (Honisch et al., 2021; Howard et al., 2021). Studies that employed identical social and non-social stimuli (e.g., computer based, auditory signals), apart from being described to participants as originating either from a human partner or a computer, found that rates of synchronisation were comparable between conditions (Koehne et al., 2016; Mills et al., 2019). This suggests that the fact that participants thought they were interacting with a social partner, as opposed to a non-social stimulus, provided *no* intrinsic motivational effect. However, where a partner's actions are represented only by computer-based signals, the "presence" of the social partner is much less salient than in a more naturalistic interaction. This limited salience is a possible alternative explanation for lack of difference between conditions. Thus, the relative contributions of social context and physical congruency on IS remain unclear.

The studies described above considered whether social presence motivated synchronisation at the group level. An alternative approach is to examine how individual differences in trait levels of social motivation influence levels of IS. Using a self-reported measure of social motivation, participants classified as "pro-social" were found to spontaneously synchronise with a partner to a significantly greater extent than participants classified as "pro-self" (Lumsden et al., 2012), suggesting that higher social motivation at an individual level precipitates higher levels of IS.

The quality of the social relationships between partners may also affect IS. Partners with pre-existing affiliative relationships have been found to synchronise more than unfamiliar partners (Latif et al., 2014). Further, whether participants are positively or negatively disposed toward previously unfamiliar partners influences levels of IS. For example, participants spontaneously synchronised significantly more with partners they believed to be punctual (Miles et al., 2010), honest (Brambilla et al., 2016), and attractive (Zhao et al., 2015), relative to partners they believed to be tardy, dishonest, and unattractive, respectively. Collectively, the effects of social moderators on IS has led researchers to conclude that IS is influenced by the need or desire to make social connections with others (Lumsden et al., 2014; Hoehl et al., 2021) and serves as a means of co-constructing a social space (Cornejo et al., 2017).

Although there is relatively strong evidence that social context modulates IS, less is known about the mechanism by which it does so. One possibility is that the motivation to seek connection with a partner influences social orienting, such that the increased desire to connect with a partner increases attentional allocation to them (Lumsden et al., 2012; Gvirts and Perlmutter, 2020). Increased attention has been proposed to have cascading effects on other component processes, such as improved action prediction, leading to more accurate motor planning. Critically, the resulting behavioural alignment is thought to be experienced as rewarding, thus promoting continued mutual social attention and maintaining the integrity of the feedback loop described above (Kokal et al., 2011; Shamay-Tsoory et al., 2019; Gvirts and Perlmutter, 2020). Feelings of reward may arise because behavioural alignment leads to reduced processing demands (Koban et al., 2019; Shamay-Tsoory et al., 2019) but are also likely to be influenced by the social significance of the interaction to each partner. The extent to which IS is experienced as rewarding by each partner is therefore a likely further source of variation in levels of IS (Gvirts and Perlmutter, 2020; Gvirts Probolovski and Dahan, 2021).

PART 2 – INTERPERSONAL SYNCHRONY IN AUTISM

Converging evidence indicates that, on average, IS is reduced in autism (McNaughton and Redcay, 2020). Studies employing structured experimental tasks involving pendulum swinging (Fitzpatrick et al., 2016), chair rocking (Marsh et al., 2013), movement improvisation (Brezis et al., 2017) and gaze following (Liu et al., 2021) all found lower levels of IS when one of the interacting partners was autistic, relative to when both partners were non-autistic. A similar pattern of results has emerged from the analysis of naturalistic interactions. IS was reduced during a clinical diagnostic interview for adults who were subsequently given an autism diagnosis, compared to those who were not (Koehler et al., 2021). Similarly, conversations between dyads in which at least one partner was autistic were characterised by reduced IS, relative to conversations between typical dyads (Georgescu et al., 2020). In typical adults, higher levels of autistic traits within dyads were significantly associated with reduced spontaneous motor synchrony when partners walked and talked together (Cheng et al., 2017). Thus, there is a range of evidence suggesting reduced IS in autism. However, a substantial majority of studies that support this conclusion compared IS in mixed dyads (consisting of one autistic and one non-autistic partner) with IS in typical dyads. Relatively little is known about levels of IS in interactions between autistic people.

Additionally, a feature of many of the studies described above is that they involved a relatively sophisticated level of social interaction. By contrast, autistic and non-autistic participants achieved comparable levels of synchrony in an interaction in which the social, perceptual, and motoric content of the interaction was substantially reduced, in that it involved only the exchange of signals with an unseen partner *via* a computer button press (Koehne et al., 2016). Together with evidence that synchrony is reduced but still present at above chance levels in more complex interactions (Georgescu et al., 2020; Koehler et al., 2021), this finding suggests that a basic tendency to synchronise may be intact in autism. However, it is unclear which particular processes may account for the differences in IS during naturalistic social interactions. This section explores the potential points of divergence in IS between autistic and non-autistic individuals.

Social Orienting

Atypical social orienting has been proposed as one possible mechanism precipitating reduced IS in autism (Fitzpatrick et al., 2016; Brezis et al., 2017; McNaughton and Redcay, 2020). Recent meta-analyses have found that, on average, autistic individuals display reduced visual attention toward social stimuli relative to non-autistic individuals (Frazier et al., 2017; Hedger et al., 2020). However, there is also heterogeneity between studies, with a substantial number finding no differences in the tendency to visually attend to social stimuli (Frazier et al., 2017; Hedger et al., 2020). Further, a majority of studies examining social orienting involve passive viewing of stimuli, rather than social orienting during live interactions, which may prompt different patterns of gaze behaviour (von dem Hagen and Bright, 2017). There is less evidence of social orienting in autism during active engagement in an interaction. While some research reports an increased tendency amongst autistic individuals to visually attend to background information rather than to a social partner (Zhao et al., 2021), others have failed to observe such an effect (Canigueral and Hamilton, 2019). There is more substantial evidence that autistic people attend atypically to specific cues, with evidence, for example that mutual gaze (Jones and Klin, 2013; Vabalas and Freeth, 2016; Nyström et al., 2017; Hessels et al., 2018; McParland et al., 2021) and gaze following (Vivanti et al., 2011; Riby et al., 2013) are, on average, reduced in autism.

There is also evidence of heterogeneity within the *patterns* of social attending displayed by autistic people. For instance, while non-autistic participants viewing video footage of social interactions had highly predictable looking patterns, the gaze patterns of autistic participants were highly variable, and became less similar to typical gaze patterns with higher levels of autistic traits (Avni et al., 2020). Additionally, differences in social orienting may be moderated by gender, with evidence that autistic females and neurotypical individuals display comparable patterns of visual attention toward faces (Harrop et al., 2018; Harrop et al., 2019). However, autistic females are more likely than autistic males to employ strategies to display neurotypical social behaviour and/or compensate for social difficulties (Cook et al., 2021b). Thus, the moderating effect of gender in these studies may have resulted from female participants employing a learned strategy through which they consciously attend to faces (Harrop et al., 2019). Recent studies have also considered how social orienting proceeds over time during the course of an interaction. Autistic and non-autistic participants both displayed a high probability of initial visual attending to social stimuli, followed by a decline after several seconds. However, non-autistic participants were significantly more likely than autistic participants to return their visual attention to the social stimuli shortly afterward (Del Bianco et al., 2021; Hedger and Chakrabarti, 2021). Further, autistic children shifted their gaze in response to the gaze of a social partner significantly more slowly that non-autistic children (Liu et al., 2021). Longer latencies in gaze following are likely to reduce the extent to which relevant behavioural cues can be perceived and acted upon.

Overall, the evidence indicates that that some, but not all, autistic individuals are likely to demonstrate atypical social orienting, with patterns of visual attention to social stimuli unfolding differently over time. Given that social orienting facilitates IS in typical populations, it is likely that differences in social orienting over the course of an interaction play a role in reduced IS for some autistic individuals. Although there is currently no direct evidence to this effect, research in the related field of imitation provides some indicative support. For instance, reduced visual attention to a demonstrator by autistic children was significantly associated with reduced spontaneous imitation of the acts performed by the demonstrator (Gonsiorowski et al., 2015). Further, when explicitly instructed to pay attention to the features of an action, autistic and non-autistic participants imitated the action with an equivalent degree of accuracy (Gowen et al., 2020). Thus, atypical visual attention to a partner influences imitation in autism, and is likely to play an equivalent role in the context of IS.

Multisensory Processing

As discussed in Part 1, efficient multisensory integration of social stimuli is likely to support the emergence of IS. The balance of evidence suggests that autistic individuals demonstrate reduced multisensory acuity, reflected in an increased tendency to report relatively asynchronous visual and auditory stimuli as originating from the same source (Zhou et al., 2018; Wallace et al., 2020). This can potentially lead to inappropriate perceptual binding of incoming sensory stimuli and thus a less coherent picture of the immediate environment (Casassus et al., 2019), including IS-relevant social cues. However, a substantial minority of studies observe no difference in multisensory processing abilities between autistic and non-autistic participants (Feldman et al., 2018; Meilleur et al., 2020; Wallace et al., 2020). Possible explanations for this divergence include variation in the age of participants, with some researchers suggesting that maturation of multisensory integration is delayed in autism (Beker et al., 2018; Feldman et al., 2018). Additionally, there is more consistent evidence of differential multisensory processing when complex, speech-based stimuli are used, relative to simplified, non-social stimuli such as flashes and tones (Stevenson et al., 2014; Meilleur et al., 2020).

To our knowledge, only one study to date has examined the relation between multisensory processing and IS in autism, with autistic children displaying both reduced audio-visual multisensory acuity and reduced non-verbal synchrony, relative to typically developing children (Noel et al., 2018). However, the multisensory acuity of autistic children was not significantly associated with the amount of IS they displayed, potentially indicating that autistic participants did not make use of available audio-visual information to inform other component processes of IS (Noel et al., 2018).

As noted in relation to typical populations above, multisensory integration of proprioceptive and visual information may also be important in supporting IS, because it enables the effective monitoring of one's own motor behaviour. There is evidence that overreliance on proprioceptive information leads to less efficient multisensory integration of proprioceptive and visual information in autism (Greenfield et al., 2015), and that this precipitates reduced accuracy in motor behaviour (Glazebrook et al., 2009; Haswell et al., 2009). However, there is no evidence, to date, of an association with reduced IS specifically.

Overall, there is evidence of atypical multisensory processing in autism, which is likely to contribute to reduced IS. However, further research is required to establish the extent of its contribution and its relationship with other component processes.

Action Prediction

Evidence from typical populations, as outlined in Part 1, indicates that IS is facilitated by accurately anticipating the spatial and temporal aspects of a social partner's movement. In autism, several researchers have proposed that a generalised impairment in prediction underpins a variety of autistic traits (Sinha et al., 2014; Van de Cruys et al., 2014; Cannon et al., 2021), including reduced action co-ordination (e.g., Cerullo et al., 2021). There is evidence of atypical action prediction in autism, which potentially influences IS. For example, when observing the repeated actions of a cartoon character, autistic children generated fewer and less accurate spontaneous action predictions than non-autistic children (Schuwerk et al., 2016). In more naturalistic contexts, autistic individuals have displayed a reduced tendency to make spontaneous action predictions about others' behaviour from both gaze (Pierno et al., 2006) and kinematic (Hudson et al., 2021) cues. Accuracy of action prediction may depend not only on the ability of the person making the prediction, but also on who is being observed (Cook, 2016). There is evidence that, when predicting the actions of another based on observing their movement kinematics, autistic people are better at predicting the actions of other autistic people than they are at predicting the actions of non-autistic people, and vice versa (Montobbio et al., 2022). This suggests that, when autistic and non-autistic partners interact, they may experience bidirectional difficulties with action prediction, potentially leading to difficulties in establishing and maintaining IS.

While action prediction difficulties are a plausible cause of reduced IS in autism, there is no direct evidence of this relation. However, there is evidence that atypical action prediction in autism contributes to reduced co-ordination in joint action, where two people co-ordinate their actions to achieve a shared goal. Autistic and non-autistic children coordinated equally well with an experimenter when predictive demands were minimised because the end point of an action was unambiguous. In contrast, when the experimenter's movement had to be inferred from kinematic cues alone, autistic children were significantly impaired in their co-ordination (Fulceri et al., 2018). In the context of IS, many movements are likely to be non-transitive and thus lack a clear end point. Less frequent action prediction in autism, and less accurate action prediction, both by autistic people and their non-autistic social partners, may therefore contribute to reduced IS in autism.

Motor Behaviour

A key component of IS, described in Part 1, is the planning and execution of accurate and timely motor activity. Impairments in motor behaviour frequently co-occur with autism (for reviews see Fournier et al., 2010; Hocking and Caeyenberghs, 2017; Hudry et al., 2020; Zampella et al., 2021), and are therefore a plausible contributor to reduced IS in autistic populations. However, the evidence to support such a contribution is relatively limited.

For example, a number of studies have assessed basic motor synchrony, typically by requiring participants to tap a finger in synchrony with a simple, repetitive stimulus. The results indicate that the ability to synchronise simple motor output with basic and non-social stimuli in autism is broadly intact (Koehne et al., 2016; Tryfon et al., 2017; Morimoto et al., 2018; Honisch et al., 2021; Vishne et al., 2021), or even enhanced (Edey et al., 2019).

Further, while there is some evidence of a positive association between motor abilities, assessed in an individual context, and IS (Brezis et al., 2017), the existence of such an association has also been found to depend on the particular tasks used (Fitzpatrick et al., 2017a). Other studies have failed to find a significant association between motor abilities assessed in a solo context and IS in autism (Kaur et al., 2018; Koehler et al., 2021). However, these studies used generalised measures of motor ability, rather than specific component processes of motor functioning (Gowen and Hamilton, 2013), such as motor planning, motor timing, and motor control. The heterogeneity in study outcomes may therefore reflect differential loading of motor tasks onto particular aspects of motor functioning. There is some evidence that motor planning may mediate the association between increased autistic traits and reduced IS (Granner-Shuman et al., 2021), but the role of specific component motor processes in IS in autism is otherwise yet to be explored. In sum, the evidence for an association between motor abilities and reduced IS in autism is mixed. Further research is needed to establish whether specific motor processes might be the key contributors.

An additional consideration in relation to motor behaviour and IS is the motor signatures of interacting partners. As outlined in Part 1, the degree of similarity between partners' motor signatures affects the extent of IS (Hart et al., 2014; Słowiński et al., 2016). The movement patterns of autistic individuals tend to differ from those produced by non-autistic individuals (Bloch et al., 2019), including across gestures (Anzulewicz et al., 2016), head movements (Martin et al., 2018), and gait (Cho et al., 2022). The relative dissimilarity of individual motor signatures between partners may therefore lead to lower levels of IS in interactions between autistic and non-autistic people (Georgescu et al., 2020).

Monitoring and Adaptation

The IS-relevant processes described in this conceptual analysis are believed to be embedded in a feedback loop, which includes bidirectional monitoring and error correction (Konvalinka et al., 2010; Marieke Van der Steen and Keller, 2013; Shamay-Tsoory et al., 2019; Feniger-Schaal and Warzager, 2020), and facilitates adaptation of motor behaviour to bring about IS. Effective error monitoring likely depends partly on the extent to which interacting partners visually attend to their own and their partner's behaviour. Atypical social orienting and attention in autism, as discussed above, may compromise this process. Additionally, because IS depends on the relative timing of partners' behaviour, effective error monitoring also depends on effective detection of co-occurring behaviours (Shamay-Tsoory et al., 2019). The evidence regarding differences in perceptual sensitivity to the relative timing of stimuli in autism is inconclusive (Casassus et al., 2019; Meilleur et al., 2020), and there has been no exploration of whether autistic and non-autistic people differ in their perceptual sensitivity to IS during interactions. Thus the relevance of relative timing abilities for error monitoring in IS in autism is not yet well understood.

Following error monitoring, internal predictive models and motor plans must be efficiently updated (Shamay-Tsoory et al., 2019). Difficulties for autistic people with action prediction (Montobbio et al., 2022) as well motor planning (Granner-Shuman et al., 2021), both discussed above, are likely to be implicated here. As outlined in Part 1, the speed with which updating occurs is also an important consideration. There is evidence that the updating process may occur more slowly in autistic than in typical populations, resulting in lower levels of synchronous behaviour. For example, autistic participants took significantly longer to adjust their finger tapping when the pace of an isochronous stimulus changed, despite demonstrating comparable levels of synchronisation when the pace of the stimulus was constant (Vishne et al., 2021). This indicates that performance differences were attributable either to slower updating of the internal model, or slower incorporation of the model into updated movement plans (Vishne et al., 2021). Slow updating and error correction within the dynamic context of a social exchange are likely to result in difficulty both in establishing and in maintaining IS, and are therefore plausible contributors to reduced IS in autism.

Attentional Load

Substantial evidence suggests that attentional resources are atypically distributed in autism. Despite evidence of enhanced perceptual capacity overall (e.g., Remington et al., 2009; Bayliss and Kritikos, 2011), autistic people have been found to display reduced sensitivity to incoming information, narrowly focussed attention to particular stimuli, and difficulties disengaging from stimuli (for reviews see Allen and Courchesnel, 2001, and Keehn et al., 2013). Atypical patterns of attending are likely to lead to a reduction in the resources available for attending to social information relevant to IS, and lead to reduced IS as a result.

As discussed in Part 1, there is evidence that intentional IS is reduced with greater attentional load, through a diminished attentional capacity for IS-relevant stimuli (Temprado and Laurent, 2004). For autistic people, social interactions may involve multiple attention-demanding phenomena, which are not as exacting for non-autistic people. For example, the online process of understanding the mental states of a conversational partner is a relatively intuitive process for non-autistic people, but autistic people may need to engage in effortful compensatory strategies (Livingston et al., 2020). Autistic people may also engage in effortful behaviours aimed at masking autistic social tendencies, which could include suppressing repetitive behaviours, carefully monitoring their eye contact, or using behavioural rules or conversational scripts to guide social behaviour (Livingston et al., 2019b; Cook et al., 2021a,b). Attending to such strategies during conversation may deplete the attentional resources available to autistic people for attending to IS-relevant cues, leading to reduced IS during the interaction.

The influence of environmental distractors on IS may also be different for autistic people. Autistic people experience a range of unusual sensory experiences, including both hyper- and hypo- sensitivity to sensory features of the environment, as well as sensory seeking of preferred sensory experiences (American Psychiatric Association, 2013). These unusual sensory experiences can cause distress or distraction for autistic people (Robertson and Simmons, 2015) which may reduce their capacity for processing and attending to IS. Further, while neurotypical individuals preferentially process social stimuli even when attentional load from non-social stimuli is relatively high (Lavie et al., 2003), the same effect is not observed in autistic people (Remington et al., 2012). This evidence suggests that environmental distractors are more likely to result in reduced IS in autistic than in typical populations.

Further, it has been proposed that the emergence of spontaneous IS precipitates a *reduction* in attentional load, thus enhancing processing capacity for other stimuli (Koban et al., 2019). Given that lower levels of spontaneous IS are generally observed in autism, fewer attentional resources may be available for processing other social cues arising during the course of an interaction.

Overall, therefore, differences in attentional load and the way it is processed may influence IS in autism in a number of ways. Attentional resources for processing IS-relevant stimuli may be depleted by atypical distribution of attentional resources, the socio-cognitive demands of an interaction, the use of compensatory and camouflaging behaviour, or by atypical processing of the sensory environment. Further, autistic people may be relatively susceptible to distractor stimuli, leading to reduced intentional IS in autism as a result. Finally, lower levels of spontaneous IS in autism are likely to place relatively high attentional demands on autistic people, who are likely to experience a reduced capacity for processing other stimuli, including other social information, as a result.

Social Context

Autism is a condition characterised by atypical social communication, thus the social context is potentially highly relevant to differences in IS in autism. Differences in social orienting and in socially relevant attentional load during interactions, as well as their potential impact on IS, have already been considered above. A broader question concerns whether atypical social communication in autism leads to reduced IS; whether reduced IS may itself give rise to disrupted social communication; or whether there is a complex and bidirectional relationship between the two constructs.

In support of the former proposition, some researchers have proposed that social context is less relevant to autistic people because of a reduced desire to engage in social exchange and to forge social bonds (Chevallier et al., 2012). If social motivation were reduced in autism, this would be likely to have downstream effects on social communication including IS. Just as the desire for social connection is thought to drive IS in typical populations (Miles et al., 2010; Lumsden et al., 2012; Zhao et al., 2015; Brambilla et al., 2016), a reduced desire for social connections could drive reduced levels of IS in autism (Brezis et al., 2017). However, this account of reduced social motivation in autism has been challenged (Jaswal and Akhtar, 2019; Livingston et al., 2019a). There is evidence of substantial variation in levels of social motivation in autism (Garman et al., 2016; Sedgewick et al., 2016), as well as evidence to suggest that levels of social motivation are dissociated from social interaction outcomes (Morrison et al., 2020a). Further, it may be that social motivation is present but atypically expressed in autism (Jaswal and Akhtar, 2019; Livingston et al., 2019a). However, conventional markers of social motivation, such as eye contact (Akhtar and Jaswal, 2020), are themselves likely to facilitate IS (see above). It may be that the absence of such markers, rather than the absence of social motivation *per se*, plays a role in reduced IS for some autistic individuals.

The social context may also influence IS in autism because of elevated rates of social anxiety disorder among autistic people (Spain et al., 2018; Hollocks et al., 2019). People with social anxiety disorder display reduced IS relative to people without the disorder (Hessels et al., 2018). Several features of social anxiety disorder have been suggested as potential explanations for this reduction in IS, including reduced eye contact (Hessels et al., 2018), reduced non-verbal behaviour such as nodding or gesturing, and increased internally focussed attention (Asher et al., 2020). Elevated levels of anxiety may also precipitate increased levels of emotional arousal (O'Haire et al., 2015), which has been associated with faster motor output during an intentional synchronisation task (Monier and Droit-Volet, 2018). Thus, for autistic people, the contextual effect of increased anxiety or arousal during social situations may exacerbate any underlying differences in the component processes that contribute to IS.

It is also important to consider whether reduced IS may itself give rise to disrupted social communication over time. From a developmental perspective, it is possible that early difficulties with the perceptual and motor elements of IS initiate a developmental trajectory whereby early social interactions are experienced as less rewarding by autistic children, such that they are less likely to engage in social interaction. Less engagement would result in fewer opportunities to pick up on social cues and develop typical social skills. In turn, peers and caregivers may find asynchronous interactions with autistic children relatively less rewarding, leading to reduced social engagement from others, further reducing the opportunities for the development of typical social communication abilities (Delafield-Butt et al., 2019; Zampella et al., 2020). Thus, when considering the social context, the relationship between social communication difficulties and reduced IS in autism may in fact be developmental and bidirectional in nature. Longitudinal research is required to understand how these factors influence each other and relate to social communication difficulties over time.

DISCUSSION

IS is a significant and complex social process which contributes to positive social outcomes and to building social relationships throughout the lifespan. However, its underlying mechanisms and how they relate to one another are still not well understood. In Part 1 of this conceptual analysis, we synthesised a wide range of evidence outlining the contributions of social orienting, multisensory processing, action prediction, and motor planning and execution. We described how these mechanisms are believed to be embedded together in a feedback loop of error monitoring and correction, and reflected on the moderating effects of attentional load and social context. As well as discussing each component mechanism in turn, we also sought to draw out the interdependence between these constructs. A key feature of this interdependence is that one process is likely to have cascading effects on others, such that successful IS depends on each process providing appropriate input for the next. For example, successful action prediction depends in part on an individual's predictive ability, but cannot take place unless relevant social information has first been gathered *via* efficient social orienting. In Part 2, we applied this understanding of the component processes of IS to consider the factors that might precipitate reduced IS in autism. We outlined evidence of atypical functioning in autism across a number of component mechanisms, and highlighted the variation in the extent to which the evidence supported a link between such divergence in functioning and reduced IS. Overall, however, it is likely that differences across multiple processes contribute to reduced IS in autism, with atypicality in any given process having potential downstream effects on other relevant mechanisms.

Although our analysis described the key component processes that contribute to IS in typical and autistic populations, it is not intended as an exhaustive account of every potential influence on IS. For example, our focus was on non-verbal IS and further consideration should be given to verbal and vocal IS, as well as the complex interplay between non-verbal and verbal IS. Additionally, age-related changes in component processes are likely to influence the extent to which IS occurs. As yet, however, little is known about how developmental timing differences in the emergence of these component skills, and how they vary within autistic and non-autistic populations, influence IS during development.

Similarly, consideration of the neural substrates of IS is likely to shed light on the mechanisms involved in IS. For example, motor cortex activity may play a role in facilitating synchronisation, with evidence of a positive association between levels of sensorimotor activity when observing a partner's actions and more accurate synchronisation with the partner in finger tapping (Naeem et al., 2012), button pressing (Meyer et al., 2011) and drumming games (Endedijk et al., 2017). Further, increased activation of the motor cortex was observed where participants synchronised with a partner, but not where they synchronised with a non-social stimulus (Novembre et al., 2012), suggesting the motor cortex plays a specific role in facilitating synchronisation with a biological stimulus, rather than synchronisation with external stimuli more generally. Consideration of the neural underpinnings of each component mechanism of IS likely to enrich our understanding as to how IS arises – or is compromised – during social interaction.

Relatedly, synchrony between interacting partners is known to arise at a neural as well as behavioural level (Nam et al., 2020). While there is evidence that neural and behavioural synchrony tend to co-occur and are thought to be closely related (Dumas et al., 2010; Liu et al., 2018; Koban et al., 2019), the mechanisms and directions of influence are not yet fully understood. For example, inducing neural entrainment between partners *via* simultaneous transcranial stimulation of their motor cortices enhanced partners' finger tapping synchrony (Novembre et al., 2017) and levels of IS in a naturalistic interaction (Pan et al., 2021), suggesting that behavioural synchrony may be preceded and induced by synchrony at a neural level. By contrast, partners who observed themselves acting with IS became neurally synchronised with each other (Levy et al., 2017). Thus, neural entrainment might also arise as a consequence of behavioural entrainment (Wass et al., 2020). Further research is required to understand the potentially bidirectional nature of the relations between these different aspects of synchrony. The relations between neural and behavioural synchrony may also inform our understanding of reduced IS in autism, with some evidence of a dissociation between the two in autistic children (Kruppa et al., 2021).

In the context of autism in particular, the influence of relational factors is also likely to be an important avenue for future research. Our analysis highlights emerging evidence that differences between interacting partners, such as divergent motor signatures and mutual difficulties in action prediction between autistic and non-autistic partners, can impact IS. Further relational considerations are also likely to be relevant. For example, non-autistic people may lack understanding of autistic social behaviour, which may precipitate reduced IS when autistic and non-autistic people interact. Non-autistic people may hold stereotypical assumptions about autistic people and their behaviour, for example, characterising autistic people as unfriendly or odd, or assuming that averted social gaze connotes a lack of social interest (Turnock et al., 2022). Such preconceptions may reduce their inclination to forge a social connection with autistic social partners (Sasson et al., 2017), leading to reduced IS within mixed dyads. Similarly, low levels of acceptance of autistic social behaviour by non-autistic people may impact the extent to which autistic people feel the need to engage in resource-intensive camouflaging behaviours (Livingston et al., 2019b), which may negatively impact on the attentional resources available for processing IS-relevant cues. Few studies to date have examined IS within autistic dyads (Georgescu et al., 2020, being a notable exception). Further investigation of how autistic people synchronise with other autistic people may advance understanding of relational factors and how they might impede IS in mixed interactions.

Future research should also consider the extent to which IS influences social bonding in autism, and its importance to building social relationships relative to other aspects of social behaviour. Existing evidence indicates that autistic people experience equivalent levels of IS (Georgescu et al., 2020) but increased rapport (Crompton et al., 2020a,b; Morrison et al., 2020b) when interacting with other autistic people, relative to when interacting with non-autistic people. Taken together, these findings suggest that, for autistic people, IS may be somewhat dissociated from social bonding. It may be that different aspects of the interaction, such as the extent to which information is efficiently exchanged (Crompton et al., 2020a), or feelings of shared experience (Crompton et al., 2020b), are more important in establishing rapport and bonding for many autistic individuals (Heasman and Gillespie, 2019;

Crompton et al., 2020b; Morrison et al., 2020b). Indeed, if different aspects of an interaction are socially salient, then increased allocation of attention to such factors, potentially at the expense of attending to IS-relevant information, might logically be expected. Further investigation is needed to understand the relevance of IS within the broader context of social interaction in autism.

In conclusion, our conceptual analysis has highlighted the importance of understanding both the component processes of IS and the interrelationships between them. Drawing on research that has examined the role of individual component mechanisms, we have provided a framework for understanding how these mechanisms contribute and interact to bring about IS. Our framework has provided a conceptual basis for understanding how non-verbal IS operates in autism and how it relates to autistic experiences of social communication more generally. Finally, there is emerging evidence that IS is reduced in other conditions, including ADHD (Problovski et al., 2021) and schizophrenia (Dean et al., 2021), although the underlying reasons for reduced IS are likely to be different across different disorders. By mapping out the component mechanisms of IS and how they interact, our conceptual analysis may provide a useful starting point for identifying which component mechanisms are uniquely implicated within these different conditions.

AUTHOR CONTRIBUTIONS

CB-M, CJ, and EvdH conceived the idea for the article. CB-M, CJ, EvdH, and SG wrote the article. All authors contributed to the article and approved the submitted version.

REFERENCES

Allen, G., and Courchesne, E. (2001). Attention function and dysfunction in autism. *Front. Biosci*. 6, D105-D119. doi: 10.2741/allen

Allman, M. J., and Falter, C. M. (2015). "Abnormal timing and time perception in autism spectrum disorder? A review of the evidence," in *Time Distortions in Mind: Temporal Processing in Clinical Populations*, eds A. Vatakis and M. J. Allan (Leiden: Brill), 37–56.

Akhtar, N., and Jaswal, V. K. (2020). Stretching the social: broadening the behavioral indicators of sociality. *Child Dev. Perspect.* 14, 28–33.

Alais, D., Newell, F., and Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* 23, 3–38. doi: 10.1163/187847510X488603

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Arlington, VA: American Psychiatric Association.

Ansuini, C., Cavallo, A., Bertone, C., and Becchio, C. (2014). The visible face of intention: why kinematics matters. *Front. Psychol.* 5:815. doi: 10.3389/fpsyg.2014.00815

Anzulewicz, A., Sobota, K., and Delafield-Butt, J. T. (2016). Toward the autism motor signature: gesture patterns during smart tablet gameplay identify children with autism. *Sci. Rep.* 6:31107. doi: 10.1038/srep31107

Asher, M., Kauffmann, A., and Aderka, I. M. (2020). Out of sync: nonverbal synchrony in social anxiety disorder. *Clin. Psychol. Sci.* 8, 280–294.

Aubin, L., Mostafaoui, G., Schmidt, R., Serré, H., and Marin, L. (2021). Effects of unintentional coordination on attentional load. *Hum. Mov. Sci.* 80:102880. doi: 10.1016/j.humov.2021.102880

Avni, I., Meiri, G., Bar-Sinai, A., Reboh, D., Manelis, L., Flusser, H., et al. (2020). Children with autism observe social interactions in an idiosyncratic manner. *Autism Res.* 13, 935–946. doi: 10.1002/aur.2234

Bahrick, L. E., and Todd, J. T. (2012). "Multisensory processing in autism spectrum disorders: intersensory processing disturbance as atypical development," in *The New Handbook of Multisensory Processes*, ed. B. E. Stein (Cambridge, MA: MIT Press), 657–674.

Bayliss, A. P., and Kritikos, A. (2011). Brief report: perceptual load and the autism spectrum in typically developed individuals. *J. Autism Dev. Disord.* 41, 1573–1578. doi: 10.1007/s10803-010-1159-8

Beker, S., Foxe, J. J., and Molholm, S. (2018). Ripe for solution: delayed development of multisensory processing in autism and its remediation. *Neurosci. Biobehav. Rev.* 84, 182–192. doi: 10.1016/j.neubiorev.2017.11.008

Bernieri, F., Reznick, S., and Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. *J. Pers. Soc. Psychol.* 54:243. doi: 10.1037/0022-3514.54.2.243

Bloch, C., Vogeley, K., Georgescu, A. L., and Falter-Wagner, C. M. (2019). INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. *Front. Robot. AI* 6:73. doi: 10.3389/frobt.2019.00073

Brambilla, M., Sacchi, S., Menegatti, M., and Moscatelli, S. (2016). Honesty and dishonesty don't move together: trait content information influences behavioral synchrony. *J. Nonverbal Behav.* 40, 171–186. doi: 10.1007/s10919-016-0229-9

Brezis, R.-S., Noy, L., Alony, T., Gotlieb, R., Cohen, R., Golland, Y., et al. (2017). Patterns of joint improvisation in adults with autism spectrum disorder. *Front. Psychol.* 8:1790. doi: 10.3389/fpsyg.2017.01790

Candidi, M., Curioni, A., Donnarumma, F., Sacheli, L. M., and Pezzulo, G. (2015). Interactional leader-follower sensorimotor communication strategies during repetitive joint actions. *J. R. Soc. Interface* 12:0644. doi: 10.1098/rsif.2015.0644

Canigueral, R., and Hamilton, A. (2019). The role of eye gaze during natural social interactions in typical and autistic people. *Front. Psychol.* 10:560. doi: 10.3389/fpsyg.2019.00560

Cannon, J., O'Brien, A. M., Bungert, L., and Sinha, P. (2021). Prediction in autism spectrum disorder: a systematic review of empirical evidence. *Autism Res.* 14, 604–630. doi: 10.1002/aur.2482

Casassus, M., Poliakoff, E., Gowen, E., Poole, D., and Jones, L. A. (2019). Time perception and autistic spectrum condition: a systematic review. *Autism Res.* 12, 1440–1462. doi: 10.1002/aur.2170

Catmur, C., and Heyes, C. (2013). Is it what you do, or when you do it? The roles of contingency and similarity in pro-social effects of imitation. *Cogn. Sci.* 37, 1541–1552. doi: 10.1111/cogs.12071

Cerullo, S., Fulceri, F., Muratori, F., and Contaldo, A. (2021). Acting with shared intentions: a systematic review on joint action coordination in Autism Spectrum Disorder. *Brain Cogn.* 149:105693. doi: 10.1016/j.bandc.2021.105693

Cheng, M., Kato, M., and Tseng, C.-H. (2017). Gender and autistic traits modulate implicit motor synchrony. *PLoS One* 12:e0184083. doi: 10.1371/journal.pone.0184083

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn. Sci.* 16, 231–239. doi: 10.1016/j.tics.2012.02.007

Cho, A. B., Otte, K., Baskow, I., Ehlen, F., Maslahati, T., Mansow-Model, S., et al. (2022). Motor signature of autism spectrum disorder in adults without intellectual impairment. *Sci. Rep*. 12:7670.

Colling, L. J., Thompson, W. F., and Sutton, J. (2014). The effect of movement kinematics on predicting the timing of observed actions. *Exp. Brain Res.* 232, 1193–1206. doi: 10.1007/s00221-014-3836-x

Condon, W. S., and Sander, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science (New York, N.Y.)* 183:99. doi: 10.1126/science.183.4120.99

Cook, J. (2016). From movement kinematics to social cognition: the case of autism. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150372. doi: 10.1098/rstb.2015.0372

Cook, J., Hull, L., Crane, L., and Mandy, W. (2021b). Camouflaging in autism: a systematic review. *Clin. Psychol. Rev.* 89:102080.

Cook, J., Crane, L., Hull, L., Bourne, L., and Mandy, W. (2021a). Self-reported camouflaging behaviours used by autistic adults during everyday social interactions. *Autism* 26, 406–421. doi: 10.1177/13623613211026754

Cornejo, C., Cuadros, Z., Morales, R., and Paredes, J. (2017). Interpersonal coordination: methods, achievements, and challenges. *Front. Psychol.* 8:1685. doi: 10.3389/fpsyg.2017.01685

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., and Fletcher-Watson, S. (2020a). Autistic peer-to-peer information transfer is highly effective. *Autism* 24, 1704–1712. doi: 10.1177/1362361320919286

Crompton, C. J., Sharp, M., Axbey, H., Fletcher-Watson, S., Flynn, E. G., and Ropar, D. (2020b). Neurotype-Matching, but not being autistic, influences self and observer ratings of interpersonal rapport. *Front. Psychol.* 11:586171. doi: 10.3389/fpsyg.2020.586171

Cross, L., Turgeon, M., and Atherton, G. (2019). How moving together binds us together: the social consequences of interpersonal entrainment and group processes. *Open Psychol.* 1, 273–302.

Dean, D. J., Scott, J., and Park, S. (2021). Interpersonal coordination in schizophrenia: a scoping review of the literature. *Schizophr. Bull.* 47, 1544–1556. doi: 10.1093/schbul/sbab072

Del Bianco, T., Mason, L., Charman, T., Tillman, J., Loth, E., Hayward, H., et al. (2021). Temporal profiles of social attention are different across development in autistic and neurotypical people. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 6, 813–824. doi: 10.1016/j.bpsc.2020.09.004

Delafield-Butt, J., Trevarthen, C., Rowe, P., and Gillberg, C. (2019). Being misunderstood in autism: the role of motor disruption in expressive communication, implications for satisfying social relations. *Behav. Brain Sci.* 42:E86. doi: 10.1017/S0140525X1800242X

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. (2012). Interpersonal synchrony: a survey of evaluation methods across disciplines. *IEEE Trans. Affect. Comput.* 3, 349–365. doi: 10.1109/T-AFFC.2012.12

Delevoye-Turrell, Y., Dione, M., and Agneray, G. (2014). Spontaneous motor tempo is the easiest pace to act upon for both the emergent and the predictive timing modes. *Procedia Soc. Behav. Sci.* 126, 121–122.

Dominguez, S., Devouche, E., Apter, G., and Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant Child Dev.* 25, 240–255. doi: 10.1002/icd.1976

Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 251–288. doi: 10.1016/S0010-0277(00)00106-2

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One* 5:e12166. doi: 10.1371/journal.pone.0012166

Edey, R., Brewer, R., Bird, G., and Press, C. (2019). Brief report: typical auditory-motor and enhanced visual-motor temporal synchronization in adults with autism spectrum disorder. *J. Autism Dev. Disord.* 49, 788–793. doi: 10.1007/s10803-018-3725-4

Elliott, M. T., Wing, A. M., and Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *Eur. J. Neurosci.* 31, 1828–1835. doi: 10.1111/j.1460-9568.2010.07205.x

Endedijk, H. M., Meyer, M., Bekkering, H., Cillessen, A. H. N., and Hunnius, S. (2017). Neural mirroring and social interaction: motor system involvement during action observation relates to early peer cooperation. *Dev. Cogn. Neurosci.* 24, 33–41. doi: 10.1016/j.dcn.2017.01.001

Evans, C. A., and Porter, C. L. (2009). The emergence of mother–infant co-regulation during the first year: links to infants' developmental status and attachment. *Infant Behav. Dev.* 32, 147–158. doi: 10.1016/j.infbeh.2008.12.005

Falck-Ytter, T. (2012). Predicting other people's action goals with low-level motor information. *J. Neurophysiol.* 107, 2923–2925. doi: 10.1152/jn.00783.2011

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., and Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 95, 220–234. doi: 10.1016/j.neubiorev.2018.09.020

Feldman, R. (2007). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *J. Child Psychol. Psychiatry* 48, 329–354. doi: 10.1111/j.1469-7610.2006.01701.x

Feniger-Schaal, R., and Warzager, D. (2020). Getting synchronised in the mirror game: an exploratory study. *Body Mov. Dance Psychother.* 15, 20–37. doi: 10.1080/17432979.2019.1694071

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

Fitzpatrick, P., Romero, V., Amaral, J. L., Duncan, A., Barnard, H., Richardson, M. J., et al. (2017b). Social motor synchronization: insights for understanding social behavior in autism. *J. Autism Dev. Disord.* 47, 2092–2107. doi: 10.1007/s10803-017-3124-2

Fitzpatrick, P., Romero, V., Amaral, J. L., Duncan, A., Barnard, H., Richardson, M. J., et al. (2017a). Evaluating the importance of social motor synchronization and motor skill for understanding autism. *Autism Res.* 10, 1687–1699. doi: 10.1002/aur.1808

Flanagan, J. R., and Johansson, R. S. (2003). Action plans used in action observation. *Nature* 424, 769–771. doi: 10.1038/nature01861

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., and Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception (London)* 37, 571–583. doi: 10.1068/p5705

Foster, N. C. (2019). *Sensorimotor Learning and Control in Autism Spectrum Disorders: The Role of Sensorimotor Integration*. Doctoral thesis. Liverpool: Liverpool John Moores University.

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., and Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *J. Autism Dev. Disord.* 40, 1227–1240. doi: 10.1007/s10803-010-0981-3

Frazier, T. W. P., Strauss, M. P., Klingemier, E. W. B. A., Zetzer, E. E. M. A., Hardan, A. Y. M. D., Eng, C. M. D. P., et al. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *J. Am. Acad. Child Adolesc. Psychiatry* 56, 546–555. doi: 10.1016/j.jaac.2017.05.005

Fulceri, F., Tonacci, A., Lucaferro, A., Apicella, F., Narzisi, A., Vincenti, G., et al. (2018). Interpersonal motor coordination during joint actions in children with and without autism spectrum disorder: the role of motor information. *Res. Dev. Disabil.* 80, 13–23. doi: 10.1016/j.ridd.2018.05.018

Garman, H. D., Spaulding, C. J., Webb, S. J., Mikami, A. Y., Morris, J. P., and Lerner, M. D. (2016). Wanting it too much: an inverse relation between social motivation and facial emotion recognition in autism spectrum disorder. *Child Psychiatry Hum. Dev.* 47, 890–902. doi: 10.1007/s10578-015-0620-5

Gaziv, G., Noy, L., Liron, Y., and Alon, U. (2017). A reduced-dimensionality approach to uncovering dyadic modes of body motion in conversations. *PLoS One* 12:e0170786. doi: 10.1371/journal.pone.0170786

Georgescu, A. L., Koeroglu, S., Hamilton, A. F. d. C., Vogeley, K., Falter-Wagner, C. M., and Tschacher, W. (2020). Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. *Mol. Autism* 11:11. doi: 10.1186/s13229-019-0305-1

Glazebrook, C., Gonzalez, D., Hansen, S., and Elliott, D. (2009). The role of vision for online control of manual aiming movements in persons with autism spectrum disorders. *Autism* 13, 411–433. doi: 10.1177/1362361309105659

Gluckman, M., and Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Front. Psychol.* 4:527. doi: 10.3389/fpsyg.2013.00527

Gonsiorowski, A., Williamson, R. A., and Robins, D. L. (2015). Brief report: imitation of object-directed acts in young children with autism spectrum disorders. *J. Autism Dev. Disord.* 46, 691–697. doi: 10.1007/s10803-015-2596-1

Gowen, E., and Hamilton, A. (2013). Motor abilities in autism: a review using a computational context. *J. Autism Dev. Disord.* 43, 323–344. doi: 10.1007/s10803-012-1574-0

Gowen, E., Vabalas, A., Casson, A. J., and Poliakoff, E. (2020). Instructions to attend to an observed action increase imitation in autistic adults. *Autism Int. J. Res. Pract.* 24, 730–743. doi: 10.1177/1362361319882810

Granner-Shuman, M., Dahan, A., Yozevitch, R., and Gvirts Problovski, H. Z. (2021). The association among autistic traits, interactional synchrony and typical pattern of motor planning and execution in neurotypical individuals. *Symmetry (Basel)* 13:1034. doi: 10.3390/sym13061034

Greenfield, K., Ropar, D., Smith, A. D., Carey, M., and Newport, R. (2015). Visuo-tactile integration in autism: atypical temporal binding may underlie greater reliance on proprioceptive information. *Mol. Autism* 6, 1–10. doi: 10.1186/s13229-015-0045-9

Gvirts Probolovski, H. Z., and Dahan, A. (2021). The potential role of dopamine in mediating motor function and interpersonal synchrony. *Biomedicines* 9:382. doi: 10.3390/biomedicines9040382

Gvirts, H. Z., and Perlmutter, R. (2020). What guides us to neurally and behaviorally align with anyone specific? A neurobiological model based on fNIRS hyperscanning studies. *Neuroscientist (Baltimore, Md.)* 26, 108–116. doi: 10.1177/1073858419861912

Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1984). The timing of shifts of head postures during conservation. *Hum. Mov. Sci.* 3, 237–245. doi: 10.1016/0167-9457(84)90018-6

Harrist, A. W., and Waugh, R. M. (2002). Dyadic synchrony: its structure and function in children's development. *Dev. Rev.* 22, 555–592. doi: 10.1016/S0273-2297(02)00500-2

Harrop, C., Jones, D., Zheng, S., Nowell, S. W., Boyd, B. A., and Sasson, N. (2018). Sex differences in social attention in autism spectrum disorder. *Autism Res.* 11, 1264–1275. doi: 10.1002/aur.1997

Harrop, C., Jones, D., Zheng, S., Nowell, S., Schultz, R., and Parish-Morris, J. (2019). Visual attention to faces in children with autism spectrum disorder: are there sex differences? *Mol. Autism* 10:28. doi: 10.1186/s13229-019-0276-2

Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E., and Alon, U. (2014). Individuality and togetherness in joint improvised motion. *PLoS One* 9:e87213. doi: 10.1371/journal.pone.0087213

Haswell, C. C., Izawa, J., Dowell, L. R., Mostofsky, S. H., and Shadmehr, R. (2009). Representation of internal models of action in the autistic brain. *Nat. Neurosci.* 12, 970–972. doi: 10.1038/nn.2356

Heasman, B., and Gillespie, A. (2019). Neurodivergent intersubjectivity: distinctive features of how autistic people create shared understanding. *Autism* 23, 910–921. doi: 10.1177/1362361318785172

Hedger, N., and Chakrabarti, B. (2021). Autistic differences in the temporal dynamics of social attention. *Autism* 25, 1615–1626. doi: 10.1177/1362361321998573

Hedger, N., Dubey, I., and Chakrabarti, B. (2020). Social orienting and social seeking behaviors in ASD. A meta analytic investigation. *Neurosci. Biobehav. Rev.* 119, 376–395. doi: 10.1016/j.neubiorev.2020.10.003

Hessels, R. S., Holleman, G. A., Cornelissen, T. H., Hooge, I. T., and Kemner, C. (2018). Eye contact takes two–autistic and social anxiety traits predict gaze behavior in dyadic interaction. *J. Exp. Psychopathol.* 9:jep.062917.

Hilbrink, E. E., Gattis, M., and Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother-infant interaction. *Front. Psychol.* 6:1492. doi: 10.3389/fpsyg.2015.01492

Hocking, D. R., and Caeyenberghs, K. (2017). What is the nature of motor impairments in autism, are they diagnostically useful, and what are the implications for intervention? *Curr. Dev. Disord. Rep.* 4, 19–27. doi: 10.1007/s40474-017-0109-y

Hoehl, S., Fairhurst, M., and Schirmer, A. (2021). Interactional synchrony: signals, mechanisms and benefits. *Soc. Cogn. Affect. Neurosci.* 16, 5–18. doi: 10.1093/scan/nsaa024

Hollocks, M. J., Lerh, J. W., Magiati, I., Meiser-Stedman, R., and Brugha, T. S. (2019). Anxiety and depression in adults with autism spectrum disorder: a systematic review and meta-analysis. *Psychol. Med.* 49, 559–572. doi: 10.1017/S0033291718002283

Honisch, J. J., Mane, P., Golan, O., and Chakrabarti, B. (2021). Keeping in time with social and non-social stimuli: synchronisation with auditory, visual, and audio-visual cues. *Sci. Rep.* 11:8805. doi: 10.1038/s41598-021-88112-y

Hove, M., and Risen, J. (2009). It's all in the timing: interpersonal synchrony increases affiliation. *Soc. Cogn.* 27, 949–960. doi: 10.1521/soco.2009.27.6.949

Howard, E. M., Ropar, D., Newport, R., and Tunçgenç, B. (2021). Social context facilitates visuomotor synchrony and bonding in children and adults. *Sci. Rep.* 11:22869. doi: 10.1038/s41598-021-02372-2

Hudry, K., Chetcuti, L., and Hocking, D. R. (2020). Motor functioning in developmental psychopathology: a review of autism as an example context. *Res. Dev. Disabil.* 105, 103739–103739. doi: 10.1016/j.ridd.2020.103739

Hudson, M., Nicholson, T., Kharko, A., McKenzie, R., and Bach, P. (2021). Predictive action perception from explicit intention information in autism. *Psychon. Bull. Rev.* 28, 1556–1566. doi: 10.3758/s13423-021-01941-w

Jaswal, V. K., and Akhtar, N. (2019). Being versus appearing socially uninterested: challenging assumptions about social motivation in autism. *Behav. Brain Sci.* 42:e82. doi: 10.1017/S0140525X18001826

Johnson, B. P., Lum, J. A. G., Rinehart, N. J., and Fielding, J. (2016). Ocular motor disturbances in autism spectrum disorders: systematic review and comprehensive meta-analysis. *Neurosci. Biobehav. Rev.* 69, 260–279. doi: 10.1016/j.neubiorev.2016.08.007

Jones, W., and Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504, 427–431.

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

Keehn, B., Müller, R.-A., and Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neurosci. Biobehav. Rev.* 37, 164–183. doi: 10.1016/j.neubiorev.2012.11.014

Khan, O., Ahmed, I., Cottingham, J., Rahhal, M., Arvanitis, T. N., and Elliott, M. T. (2020). Timing and correction of stepping movements with a virtual reality avatar. *PLoS One* 15:e0229641. doi: 10.1371/journal.pone.0229641

Khoramshahi, M., Shukla, A., Raffard, S., Bardy, B. G., and Billard, A. (2016). Role of gaze cues in interpersonal motor coordination: towards higher affiliation in human-robot interaction. *PLoS One* 11:e0156874. doi: 10.1371/journal.pone.0156874

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Koban, L., Ramamoorthy, A., and Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the brain's optimization principle. *Soc. Neurosci.* 14, 1–9. doi: 10.1080/17470919.2017.1400463

Koehler, J. C., Georgescu, A. L., Weiske, J., Spangemacher, M., Burghof, L., Falkai, P., et al. (2021). Brief report: specificity of interpersonal synchrony deficits to autism spectrum disorder and its potential for digitally assisted diagnostics. *J. Autism Dev. Disord.* 1–9 doi: 10.1007/s10803-021-05194-3 [Epub ahead of print].

Koehne, S., Hatri, A., Cacioppo, J. T., and Dziobek, I. (2016). Perceived interpersonal synchrony increases empathy: insights from autism spectrum disorder. *Cognition* 146, 8–15. doi: 10.1016/j.cognition.2015.09.007

Kokal, I., Engel, A., Kirschner, S., and Keysers, C. (2011). Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment–if the rhythm comes easily. *PLoS One* 6:e27272. doi: 10.1371/journal.pone.0027272

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol. (2006)* 63, 2220–2230. doi: 10.1080/17470218.2010.497843

Koul, A., Cavallo, A., Ansuini, C., and Becchio, C. (2016). Doing it your way: how individual movement styles affect action prediction. *PLoS One* 11:e0165297. doi: 10.1371/journal.pone.0165297

Kruppa, J. A., Reindl, V., Gerloff, C., Oberwelland Weiss, E., Prinz, J., Herpertz-Dahlmann, B., et al. (2021). Brain and motor synchrony in children and adolescents with ASD—a fNIRS hyperscanning study. *Soc. Cogn. Affect. Neurosci.* 16, 103–116. doi: 10.1093/scan/nsaa092

Latif, N., Barbosa, A. V., Vatiokiotis-Bateson, E., Castelhano, M. S., and Munhall, K. G. (2014). Movement coordination during conversation. *PLoS One* 9:e105036. doi: 10.1371/journal.pone.0105036

Lavie, N., Ro, T., and Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychol. Sci.* 14, 510–515.

Leppanen, J. M. (2016). Using eye tracking to understand infants' attentional bias for faces. *Child Dev. Perspect.* 10, 161–165. doi: 10.1111/cdep.12180

Levy, J., Goldstein, A., and Feldman, R. (2017). Perception of social synchrony induces mother–child gamma coupling in the social brain. *Soc. Cogn. Affect. Neurosci.* 12, 1036–1046. doi: 10.1093/scan/nsx032

Liu, D., Liu, S., Liu, X., Zhang, C., Li, A., Jin, C., et al. (2018). Interactive brain activity: review and progress on EEG-based hyperscanning in social interactions. *Front. Psychol.* 9:1862. doi: 10.3389/fpsyg.2018.01862

Liu, Q., Wang, Q., Li, X., Gong, X., Luo, X., Yin, T., et al. (2021). Social synchronization during joint attention in children with autism spectrum disorder. *Autism Res.* 14, 2120–2130. doi: 10.1002/aur.2553

Livingston, L. A., Shah, P., and Happé, F. (2019a). Compensation in autism is not consistent with social motivation theory. *Behav. Brain Sci.* 42:E99. doi: 10.1017/S0140525X18002388

Livingston, L. A., Shah, P., and Happé, F. (2019b). Compensatory strategies below the behavioural surface in autism: a qualitative study. *Lancet Psychiatry* 6, 766–777. doi: 10.1016/S2215-0366(19)30224-X

Livingston, L. A., Shah, P., Milner, V., and Happé, F. (2020). Quantifying compensatory strategies in adults with and without diagnosed autism. *Mol. Autism* 11, 1–10. doi: 10.1186/s13229-019-0308-y

Louwerse, M. M., Dale, R., Bard, E. G., and Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cogn. Sci.* 36, 1404–1426. doi: 10.1111/j.1551-6709.2012.01269.x

Lumsden, J., Miles, L. K., and Macrae, C. N. (2014). Sync or sink? Interpersonal synchrony impacts self-esteem. *Front. Psychol.* 5:1064. doi: 10.3389/fpsyg.2014.01064

Lumsden, J., Miles, L. K., Richardson, M. J., Smith, C. A., and Macrae, C. N. (2012). Who syncs? Social motives and interpersonal coordination. *J. Exp. Soc. Psychol.* 48, 746–751. doi: 10.1016/j.jesp.2011.12.007

Marieke Van der Steen, M. C., and Keller, P. E. (2013). The adaptation and anticipation model (ADAM) of sensorimotor synchronization. *Front. Hum. Neurosci.* 7:253. doi: 10.3389/fnhum.2013.00253

Marsh, K. L., Isenhower, R. W., Richardson, M. J., Helt, M., Verbalis, A. D., Schmidt, R. C., et al. (2013). Autism and social disconnection in interpersonal rocking. *Front. Integr. Neurosci.* 7:4. doi: 10.3389/fnint.2013.00004

Martin, K. B., Hammal, Z., Ren, G., Cohn, J. F., Cassell, J., Ogihara, M., et al. (2018). Objective measurement of head movement differences in children with and without autism spectrum disorder. *Mol. Autism* 9:14. doi: 10.1186/s13229-018-0198-4

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *J. Exp. Psychol. Gen.* 135, 348–367. doi: 10.1037/0096-3445.135.3.348

McNaughton, K. A., and Redcay, E. (2020). Interpersonal synchrony in autism. *Curr. Psychiatry Rep.* 22:12. doi: 10.1007/s11920-020-1135-8

McParland, A., Gallagher, S., and Keenan, M. (2021). Investigating gaze behaviour of children diagnosed with autism spectrum disorders in a classroom setting. *J. Autism Dev. Disord.* 51, 4663–4678. doi: 10.1007/s10803-021-04906-z

Meilleur, A., Foster, N. E. V., Coll, S.-M., Brambati, S. M., and Hyde, K. L. (2020). Unisensory and multisensory temporal processing in autism and dyslexia: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 116, 44–63. doi: 10.1016/j.neubiorev.2020.06.013

Meyer, M., and Hunnius, S. (2020). Becoming better together: the early development of interpersonal coordination. *Prog. Brain Res.* 254, 187–204. doi: 10.1016/bs.pbr.2020.06.012

Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., and Hunnius, S. (2015). The role of action prediction and inhibitory control for joint action coordination in toddlers. *J. Exp. Child Psychol.* 139, 203–220. doi: 10.1016/j.jecp.2015.06.005

Meyer, M., Hunnius, S., Elk, M. v., Ede, F. L. v., and Bekkering, H. (2011). Joint action modulates motor system involvement during action observation in 3-year-olds. *Exp. Brain Res.* 211, 581–592. doi: 10.1007/s00221-011-2658-3

Miles, L. K., Griffiths, J. L., Richardson, M. J., and Macrae, C. N. (2010). Too late to coordinate: contextual influences on behavioral synchrony. *Eur. J. Soc. Psychol.* 40, 52–60. doi: 10.1002/ejsp.721

Mills, P. F., Harry, B., Stevens, C. J., Knoblich, G., and Keller, P. E. (2019). Intentionality of a co-actor influences sensorimotor synchronisation with a virtual partner. *Q. J. Exp. Psychol. (2006)* 72, 1478–1492. doi: 10.1177/1747021818796183

Miyata, K., Varlet, M., Miura, A., Kudo, K., and Keller, P. E. (2017). Modulation of individual auditory-motor coordination dynamics through interpersonal visual coupling. *Sci. Rep.* 7:16220. doi: 10.1038/s41598-017-16151-5

Mogan, R., Fischer, R., and Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J. Exp. Soc. Psychol.* 72, 13–20. doi: 10.1016/j.jesp.2017.03.009

Monier, F., and Droit-Volet, S. (2018). Synchrony and emotion in children and adults. *Int. J. Psychol.* 53, 184–193. doi: 10.1002/ijop.12363

Monier, F., and Droit-Volet, S. (2019). Development of sensorimotor synchronization abilities: motor and cognitive components. *Child Neuropsychol.* 25, 1043–1062. doi: 10.1080/09297049.2019.1569607

Montobbio, N., Cavallo, A., Albergo, D., Ansuini, C., Battaglia, F., Podda, J., et al. (2022). Intersecting kinematic encoding and readout of intention in autism. *Proc. Natl. Acad. Sci. U.S.A.* 119:e2114648119. doi: 10.1073/pnas.2114648119

Morimoto, C., Hida, E., Shima, K., and Okamura, H. (2018). Temporal processing instability with millisecond accuracy is a cardinal feature of sensorimotor impairments in autism spectrum disorder: analysis using the synchronized finger-tapping task. *J. Autism Dev. Disord.* 48, 351–360. doi: 10.1007/s10803-017-3334-7

Morrison, K. E., DeBrabander, K. M., Jones, D. R., Ackerman, R. A., and Sasson, N. J. (2020a). Social cognition, social skill, and social motivation minimally predict social interaction outcomes for autistic and non-autistic adults. *Front. Psychol.* 11:591100. doi: 10.3389/fpsyg.2020.591100

Morrison, K. E., DeBrabander, K. M., Jones, D. R., Faso, D. J., Ackerman, R. A., and Sasson, N. J. (2020b). Outcomes of real-world social interaction for autistic adults paired with autistic compared to typically developing partners. *Autism* 24, 1067–1080. doi: 10.1177/1362361319892701

Murray, M. M., Lewkowicz, D. J., Amedi, A., and Wallace, M. T. (2016). Multisensory processes: a balancing act across the lifespan. *Trends Neurosci. (Regular ed.)* 39, 567–579. doi: 10.1016/j.tins.2016.05.003

Naeem, M., Prasad, G., Watson, D. R., and Kelso, J. A. S. (2012). Electrophysiological signatures of intentional social coordination in the 10–12Hz range. *NeuroImage (Orlando, Fla.)* 59, 1795–1803. doi: 10.1016/j.neuroimage.2011.08.010

Nam, C. S., Choo, S., Huang, J., and Park, J. (2020). Brain-to-brain neural synchrony during social interactions: a systematic review on hyperscanning studies. *Appl. Sci.* 10:6669.

Noel, J.-P., De Niear, M. A., Lazzara, N. S., and Wallace, M. T. (2018). Uncoupling between multisensory temporal function and nonverbal turn-taking in autism spectrum disorder. *IEEE Trans. Cogn. Dev. Syst.* 10, 973–982. doi: 10.1109/TCDS.2017.2778141

Novembre, G., Knoblich, G., Dunne, L., and Keller, P. E. (2017). Interpersonal synchrony enhanced through 20 Hz phase-coupled dual brain stimulation. *Soc. Cogn. Affect. Neurosci.* 12, 662–670. doi: 10.1093/scan/nsw172

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., and Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cereb. Cortex (New York, N.Y. 1991)* 22, 2894–2903. doi: 10.1093/cercor/bhr364

Nyström, P., Bölte, S., and Falck-Ytter, T. (2017). Responding to other people's direct gaze: alterations in gaze behavior in infants at risk for autism occur on very short timescales. *J. Autism Dev. Disord.* 47, 3498–3509. doi: 10.1007/s10803-017-3253-7

Oh Kruzic, C., Kruzic, D., Herrera, F., and Bailenson, J. (2020). Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Sci. Rep.* 10:20626. doi: 10.1038/s41598-020-76672-4

O'Haire, M. E., McKenzie, S. J., Beck, A. M., and Slaughter, V. (2015). Animals may act as social buffers: skin conductance arousal in children with autism spectrum disorder in a social context. *Dev. Psychobiol.* 57, 584–595. doi: 10.1002/dev.21310

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., and Scott Kelso, J. A. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192. doi: 10.1080/17470910701563392

Pan, Y., Novembre, G., Song, B., Zhu, Y., and Hu, Y. (2021). Dual brain stimulation enhances interpersonal learning through spontaneous movement synchrony. *Soc. Cogn. Affect. Neurosci.* 16, 210–221. doi: 10.1093/scan/nsaa080

Pecenka, N., and Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Exp. Brain Res.* 211, 505–515. doi: 10.1007/s00221-011-2616-0

Pierno, A. C., Mari, M., Glover, S., Georgiou, I., and Castiello, U. (2006). Failure to read motor intentions from gaze in children with autism. *Neuropsychologia* 44, 1483–1488. doi: 10.1016/j.neuropsychologia.2005.11.013

Prinsen, J., Bernaerts, S., Wang, Y., de Beukelaar, T. T., Cuypers, K., Swinnen, S. P., et al. (2017). Direct eye contact enhances mirroring of others' movements: a transcranial magnetic stimulation study. *Neuropsychologia* 95, 111–118. doi: 10.1016/j.neuropsychologia.2016.12.011

Problovski, H. Z. G., Lavi, D., Yozevitch, R., Sherman, M., Hagay, Y., and Dahan, A. (2021). Impairments of interpersonal synchrony evident in attention deficit hyperactivity disorder (ADHD). *Acta Psychol.* 212:103210. doi: 10.1016/j.actpsy.2020.103210

Rabinowitch, T.-C., and Meltzoff, A. N. (2017). Synchronized movement experience enhances peer cooperation in preschool children. *J. Exp. Child Psychol.* 160, 21–32. doi: 10.1016/j.jecp.2017.03.001

Reddish, P., Fischer, R., and Bulbulia, J. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PLoS One* 8:e71182. doi: 10.1371/journal.pone.0071182

Reddy, V., Markova, G., and Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PLoS One* 8:e65289. doi: 10.1371/journal.pone.0065289

Remington, A., Swettenham, J., Campbell, R., and Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychol. Sci.* 20, 1388–1393.

Remington, A., Campbell, R., and Swettenham, J. (2012). Attentional status of faces for people with autism spectrum disorder. *Autism* 16, 59–73. doi: 10.1177/1362361311409257

Rennung, M., and Goritz, A. (2016). Prosocial consequences of interpersonal synchrony a meta-analysis. *Z. Psychol. J. Psychol.* 224, 168–189. doi: 10.1027/2151-2604/a000252

Riby, D. M., Hancock, P. J., Jones, N., and Hanley, M. (2013). Spontaneous and cued gaze-following in autism and Williams syndrome. *J. Neurodev. Disord.* 5, 1–12. doi: 10.1186/1866-1955-5-13

Richardson, M. J., and Johnston, L. (2005). Person recognition from dynamic events: the kinematic specification of individual identity in walking style. *J. Nonverbal Behav.* 29, 25–44. doi: 10.1007/s10919-004-0888-9

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., and Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002

Robertson, A. E., and Simmons, D. R. (2015). The sensory experiences of adults with autism spectrum disorder: a qualitative analysis. *Perception* 44, 569–586. doi: 10.1068/p7833

Rösler, L., End, A., and Gamer, M. (2017). Orienting towards social features in naturalistic scenes is reflexive. *PLoS One* 12:e0182037. doi: 10.1371/journal.pone.0182037

Sartori, L., Becchio, C., and Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition* 119, 242–252. doi: 10.1016/j.cognition.2011.01.014

Sartori, L., Becchio, C., Bara, B. G., and Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Conscious. Cogn.* 18, 766–772. doi: 10.1016/j.concog.2009.06.004

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., and Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Sci. Rep.* 7, 1–10. doi: 10.1038/srep40700

Schuwerk, T., Sodian, B., and Paulus, M. (2016). Cognitive mechanisms underlying action prediction in children and adults with autism spectrum condition. *J. Autism Dev. Disord.* 46, 3623–3639. doi: 10.1007/s10803-016-2899-x

Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367. doi: 10.1111/j.1756-8765.2009.01024.x

Sebanz, N., and Knoblich, G. (2021). Progress in joint-action research. *Curr. Dir. Psychol. Sci.* 30, 138–143.

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sedgewick, F., Hill, V., Yates, R., Pickering, L., and Pellicano, E. (2016). Gender differences in the social motivation and friendship experiences of autistic and non-autistic adolescents. *J. Autism Dev. Disord.* 46, 1297–1306. doi: 10.1007/s10803-015-2669-1

Shamay-Tsoory, S. G., Saporta, N., Marton-Alper, I. Z., and Gvirts, H. Z. (2019). Herding brains: a core neural mechanism for social alignment. *Trends Cogn. Sci.* 23, 174–186. doi: 10.1016/j.tics.2019.01.002

Shockley, K., Santana, M.-V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29:326. doi: 10.1037/0096-1523.29.2.326

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., et al. (2014). Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15220–15225. doi: 10.1073/pnas.1416797111

Słowiński, P., Zhai, C., Alderisio, F., Salesse, R., Gueugnon, M., Marin, L., et al. (2016). Dynamic similarity promotes interpersonal coordination in joint action. *J. R. Soc. Interface* 13:20151093. doi: 10.1098/rsif.2015.1093

Spain, D., Sin, J., Linder, K. B., McMahon, J., and Happé, F. (2018). Social anxiety in autism spectrum disorder: a systematic review. *Res. Autism Spectr. Disord.* 52, 51–68.

Stevenson, R. A., Siemann, J. K., Schneider, B., Eberly, H., Woynaroski, T., Camarata, S. M., et al. (2014). Multisensory temporal integration in autism spectrum disorders. *J. Neurosci.* 34, 691–697. doi: 10.1523/JNEUROSCI.3615-13.2014

Su, Y.-H. (2014). Visual enhancement of auditory beat perception across auditory interference levels. *Brain Cogn.* 90, 19–31. doi: 10.1016/j.bandc.2014.05.003

Tarr, B., Launay, J., Cohen, E., and Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biol. Lett.* 11:20150767. doi: 10.1098/rsbl.2015.0767

Temprado, J.-J., and Laurent, M. (2004). Attentional load associated with performing and stabilizing a between-persons coordination of rhythmic limb movements. *Acta Psychol.* 115, 1–16. doi: 10.1016/j.actpsy.2003.09.002

Trainor, L. J., and Cirelli, L. (2015). Rhythm and interpersonal synchrony in early social development. *Ann. N. Y. Acad. Sci.* 1337:45. doi: 10.1111/nyas.12649

Tryfon, A., Foster, N. E., Ouimet, T., Doyle-Thomas, K., Anagnostou, E., Sharda, M., et al. (2017). Auditory-motor rhythm synchronization in children with autism spectrum disorder. *Res. Autism Spectr. Disord.* 35, 51–61. doi: 10.1016/j.rasd.2016.12.004

Tunçgenç, B., and Cohen, E. (2016). Movement synchrony forges social bonds across group divides. *Front. Psychol.* 7:782. doi: 10.3389/fpsyg.2016.00782

Tunçgenç, B., and Cohen, E. (2018). Interpersonal movement synchrony facilitates pro-social behavior in children's peer-play. *Dev. Sci.* 21:e12505. doi: 10.1111/desc.12505

Tunçgenç, B., Cohen, E., and Fawcett, C. (2015). Rock with me: the role of movement synchrony in infants' social and nonsocial choices. *Child Dev.* 86, 976–984. doi: 10.1111/cdev.12354

Turnock, A., Langley, K., and Jones, C. R. (2022). Understanding stigma in autism: a narrative review and theoretical model. *Autism Adulthood* 4, 76–91.

Vabalas, A., and Freeth, M. (2016). Brief report: patterns of eye movements in face to face conversation are associated with autistic traits: evidence from a student sample. *J. Autism Dev. Disord.* 46, 305–314. doi: 10.1007/s10803-015-2546-y

Vacharkulksemsuk, T., and Fredrickson, B. L. (2011). Strangers in sync: achieving embodied rapport through shared movements. *J. Exp. Soc. Psychol.* 48, 399–402. doi: 10.1016/j.jesp.2011.07.015

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., et al. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* 121, 649–675. doi: 10.1037/a0037665

Vesper, C., Wel, R. P. R. D. v. d., Knoblich, G. K., and Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Exp. Brain Res.* 211, 517–530. doi: 10.1007/s00221-011-2706-z

Vicaria, I. M., and Dickens, L. (2016). Meta-analyses of the intra- and interpersonal outcomes of interpersonal coordination. *J. Nonverbal Behav.* 40, 335–361. doi: 10.1007/s10919-016-0238-8

Vishne, G., Jacoby, N., Malinovitch, T., Epstein, T., Frenkel, O., and Ahissar, M. (2021). Slow update of internal representations impedes synchronization in autism. *Nat. Commun.* 12:5439. doi: 10.1038/s41467-021-25740-y

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., et al. (2011). Intact and impaired mechanisms of action understanding in autism. *Dev. Psychol.* 47:841. doi: 10.1037/a0023105

von dem Hagen, E. A. H., and Bright, N. (2017). High autistic trait individuals do not modulate gaze behaviour in response to social presence but look away more when actively engaged in an interaction. *Autism Res.* 10, 359–368. doi: 10.1002/aur.1666

Wallace, M. T., Woynaroski, T. G., and Stevenson, R. A. (2020). Multisensory integration as a window into orderly and disrupted cognition and communication. *Annu. Rev. Psychol.* 71, 193–219. doi: 10.1146/annurev-psych-010419-051112

Wang, Y., and Hamilton, A.F.d.C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Front. Hum. Neurosci.* 6:153. doi: 10.3389/fnhum.2012.00153

Wass, S. V., Whitehorn, M., Haresign, I. M., Phillips, E., and Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends Cogn. Sci.* 24, 329–342. doi: 10.1016/j.tics.2020.01.006

Westra, E., and Nagel, J. (2021). Mindreading in conversation. *Cognition* 210:104618. doi: 10.1016/j.cognition.2021.104618

Wright, R. L., Spurgeon, L. C., and Elliott, M. T. (2014). Stepping to phase-perturbed metronome cues: multisensory advantage in movement synchrony but not correction. *Front. Hum. Neurosci.* 8:724. doi: 10.3389/fnhum.2014.00724

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: a study of 18-to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/infa.12403

Zampella, C. J., Wang, L. A. L., Haley, M., Hutchinson, A. G., and de Marchena, A. (2021). Motor skill differences in autism spectrum disorder: a clinically focused review. *Curr. Psychiatry Rep*. 23:64. doi: 10.1007/s11920-021-01280-6

Zampella, C. J., Csumitta, K. D., Simon, E., and Bennetto, L. (2020). Interactional synchrony and its association with social and communication ability in children with and without autism spectrum disorder. *J. Autism Dev. Disord.* 50, 3195–3206. doi: 10.1007/s10803-020-04412-8

Zhao, Z., Salesse, R. N., Gueugnon, M., Schmidt, R. C., Marin, L., and Bardy, B. (2015). Moving attractive virtual agent improves interpersonal coordination stability. *Hum. Mov. Sci.* 41, 240–254. doi: 10.1016/j.humov.2015.03.012

Zhao, Z., Tang, H., Zhang, X., Zhu, Z., Xing, J., Li, W., et al. (2021). Characteristics of visual fixation in Chinese children with autism during face-to-face conversations. *J. Autism Dev. Disord.* 1–13 doi: 10.1007/s10803-021-04985-y [Epub ahead of print].

Zhou, H.-y., Cai, X.-l., Weigl, M., Bang, P., Cheung, E. F., and Chan, R. C. (2018). Multisensory temporal binding window in autism spectrum disorders and schizophrenia spectrum disorders: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 86, 66–76. doi: 10.1016/j.neubiorev.2017.12.013

Zivotofsky, A. Z., and Hausdorff, J. M. (2007). The sensory feedback mechanisms enabling couples to walk synchronously: an initial investigation. *J. Neuroeng. Rehabil.* 4:28. doi: 10.1186/1743-0003-4-28

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fpsyg.2022.897230

Assessing Sensorimotor Synchronisation in Toddlers Using the Lookit Online Experiment Platform and Automated Movement Extraction

Sinead Rocha[1,2]* and *Caspar Addyman*[3]

[1] *School of Psychology and Sport Science, Anglia Ruskin University, Cambridge, United Kingdom,* [2] *Department of Psychology, University of Cambridge, Cambridge, United Kingdom,* [3] *Department of Psychology, Goldsmiths, University of London, London, United Kingdom*

Adapting gross motor movement to match the tempo of auditory rhythmic stimulation (sensorimotor synchronisation; SMS) is a complex skill with a long developmental trajectory. Drumming tasks have previously been employed with infants and young children to measure the emergence of rhythmic entrainment, and may provide a tool for identification of those with atypical rhythm perception and production. Here we describe a new protocol for measuring infant rhythmic movement that can be employed at scale. In the current study, 50 two-year-olds drummed along with the audiovisual presentation of four steady rhythms, using videos of isochronous drumming at 400, 500, 600, and 700 ms IOI, and provided their spontaneous motor tempo (SMT) by drumming in silence. Toddlers' drumming is observed from video recordings made in participants' own homes, obtained *via* the Lookit platform for online infant studies. We use OpenPose deep-learning model to generate wireframe estimates of hand and body location for each video. The vertical displacement of the hand was extracted, and the power and frequency of infants' rhythmic entrainment quantified using Fast Fourier Transforms. We find evidence for age-appropriate tempo-flexibility in our sample. Our results demonstrate the feasibility of a fully digital approach to measuring rhythmic entrainment from within the participant's home, from early in development.

Keywords: sensorimotor synchronisation, infancy, development, machine learning, OpenPose, Lookit, automated movement analysis

Edited by:
Stefanie Hoehl,
University of Vienna, Austria

Reviewed by:
Joëlle Provasi,
Université de Sciences Lettres de Paris, France
Daniel Charles Comstock,
University of California, Davis, United States

***Correspondence:**
Sinead Rocha
sineadrocha@gmail.com

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *15 March 2022*
Accepted: *06 June 2022*
Published: *30 June 2022*

Citation:
Rocha S and Addyman C (2022) Assessing Sensorimotor Synchronisation in Toddlers Using the Lookit Online Experiment Platform and Automated Movement Extraction. Front. Psychol. 13:897230. doi: 10.3389/fpsyg.2022.897230

INTRODUCTION

Rhythmic timing underlies a broad set of human behaviours, including music and dance. Critically, the emerging ability to produce an internally generated rhythm (spontaneous motor tempo; SMT), and adapt one's movement to match an external stimulus (sensorimotor synchronisation; SMS), is related to success in language acquisition. Poor SMS is related to language difficulties in typically developing pre-schoolers (Carr et al., 2014; Politimou et al., 2019; Rios-Lopez et al., 2019), and across language disorders, including dyslexia (Thomson and Goswami, 2008; Lee et al., 2015; Persici et al., 2019), developmental language disorder (DLD; Corriveau and Goswami, 2009;

Cumming et al., 2015), and speech impediments (Olander et al., 2010; Falk et al., 2015). Sensorimotor synchronisation is suggested as a useful tool for identifying those who may struggle with language, which is suitable from infancy and through the lifespan (Ladányi et al., 2020).

Since Fraisse (1982), laboratory measures of SMT in adulthood are often measured *via* tapping paradigms, where discreet intervals are produced by the vertical displacement of the index finger onto a keypad or equivalent surface. Whilst in adulthood the self-produced rate of tapping is stable within-subjects over short periods of time (Vanneste et al., 2001), across the lifespan SMT is known to change, with children's tapping significantly faster than adult's (McAuley et al., 2006). Adult SMT is in the range of 630 ms inter-onset-interval (IOI; McAuley et al., 2006). In very early childhood, SMT measured *via* tapping has been demonstrated as fast as 400–450 ms (IOI; Bobin-Bègue and Provasi, 2008). Whilst tapping dominates the SMT literature, and adult SMS studies, across early childhood, this difficult fine-motor task is not always the most appropriate. Particularly in determining infant ability to synchronize with external stimuli, different research groups have used a variety of tasks to facilitate synchrony within populations with poor fine motor skills.

Seminal studies into infant movement to music have simply allowed infants to move freely to auditory stimuli (Zentner and Eerola, 2010; Fujii et al., 2014). However, such scenarios do not provide auditory/haptic feedback equivalent to the tapping measures used in adulthood. Other infant paradigms used small hand-held instruments such as bells (Rocha and Mareschal, 2017). The closest experimental paradigm to tapping involves whole-hand drumming. From 5 months-of-age, infants can produce their own SMT *via* drumming (Rocha et al., 2021b) with the tempo and regularity of their drumming increasing over the first 2 years of life. Whilst infants cannot reliably synchronise their movements to music, a longitudinal investigation of infant drumming to nursery rhymes of different tempi suggests that by 11-months-of-age infants are beginning to shift away from their SMT to better match the rate of the song (Rocha et al., 2021a). Studies of toddlers evidence good tempo adaptation in older infants, when drumming along with a human and non-human partner (Kirschner and Tomasello, 2009; Yu and Myowa, 2021). In contrast to whole-body free movement analysis, constraints imposed in a drumming task allow more direct comparison of SMS over age, with a common effector and motion as is commonly used in adult tapping studies. As drumming can be used across contexts, with minimal apparatus and instruction, and from 5-months-of-age with no upper limit, we suggest that this could be a candidate marker of SMS that could be used at scale to detect early individual differences. Whilst there is a strong movement toward identification of risk of language disorders using neural markers, e.g., (Attaheri et al., 2022) an accessible behavioural assessment of rhythmic skill would have multiple practical advantages in identifying children at risk.

In the current study we test the feasibility of measuring SMS in the child's own home, using asynchronous data collection methods, and largely automated data processing. If viable, such a technique will allow for large scale data collection. Our approach is focussed on creating an open source tool to evaluate rhythm in developing populations using a task that is low cost, easy to administer, and easy to adapt for research and clinical needs. Here we describe the implementation of our paradigm on a group of 2-year-olds, a notoriously difficult age to test, and document the successes and failures of our approach. We first ask whether we can detect the rate of drumming from home-video footage. We then characterise toddler Spontaneous Motor Tempo, and ask whether toddlers of this age show signs of successful SMS at a group level.

MATERIALS AND METHODS

Participants

The initial sample included 68 infants who completed the online drumming task. These include 39 female, 28 male and one gender not specified. Their mean age was 816.2 ± 94.1 days. A further 24 participants (12 female) were excluded because they did not complete the task (22) or withdrew (2). Participants were recruited through the Lookit website and *via* the experimenters' research networks. Ethical approval for the study was obtained from Psychology Ethics Committee at Goldsmiths, University of London.

Design

The study used a mixed design with all participants completing the same set of six video recorded trials, with counterbalanced order of target inter-stimulus intervals. The first and sixth trials were to designed measure spontaneous motor tempo. The middle four trials each demonstrated drumming at a different interonset interval (IOI) ranging from 400 to 700 ms in steps of 100 ms, chosen to capture the possible range of SMT over childhood. In order one these were presented in the sequence (400, 600, 500, and 700), and in order two (700, 500, 600, and 400; See **Figure 1**).

Materials

The demographic and video data were collected on the Lookit online child lab website (Scott and Schulz, 2017). The Lookit website[1] managed the sign up of participants and collection of demographic details (date of birth, sex, race, geographic location, number of children in the family, languages, parent education level, household income, number of children's books at home). It presented informed consent and data-sharing agreements for caregivers. During the data collection phase Lookit presented the stimulus and reward videos, created by the experimenters. For both Spontaneous Motor Tempo trials, the same 20-s-long silent video was used. It displayed written prompts "Can you drum for me?," "What sound does it make when you drum?," for the caregiver to read aloud to the infant. In the experimental trials, 20 s videos showed a woman's hand tapping out a steady beat on a flat surface at an interval of 400, 500, 600, or 700 ms. All the materials are available online at https://github.com/InfantLab/little-drummers.

---

[1] https://lookit.mit.edu

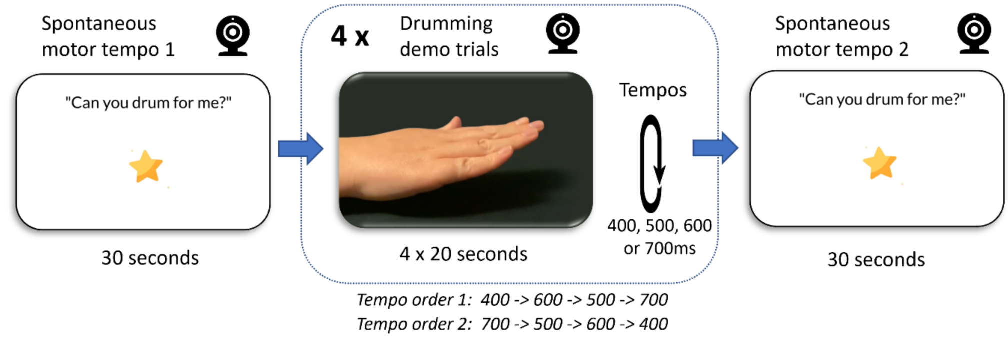


FIGURE 1 | Schematic of the data collection trials. Using the Lookit platform, children saw six short videos and their responses were video recorded. First and last trial lasted 20 s and captured infants own spontaneous drumming. Trials 2–5 each demonstrated drumming at four different tempos (an inter-stimulus intervals of 400, 500, 600, or 700 ms per beat).

Procedure

Data collection took place in participants own homes with caregivers following online instructions to run the experiment using their own personal computers. Prior to the study, caregivers created an account on the Lookit website and provided basic demographic details. At a time of their choosing they recorded verbal consent to their participation and followed instructions to position their child in view of their webcam. They were asked not to have the child on their lap and make sure that child's hands were visible in the shot, though compliance with these instructions could not be ensured. The camera view was shown on screen to help with positioning.

When child and caregiver were in position the six trials began. The trials progressed automatically but caregivers could pause the study by pressing the spacebar. If a trial was paused, it could be restarted or the caregiver could choose to end the study early if the child became too fussy. To capture children's spontaneous motor tempo, the first and last trial provided no tempo information but showed onscreen prompts for the caregiver to encourage drumming. Parents were instructed not to demonstrate drumming themselves. The experiment software randomly assigned participants to one of two order conditions which determined the sequence of Trials 2–5. In each of these trials a 20 s long video of a woman's hand drumming on table was presented. The videos were accompanied by an onscreen caption "Adults, in this video please say 'Can you drum along'?". Each trial was followed by a 5 s "reward" video. Following data collection, caregivers were asked to confirm their child's date of birth and specify a level of data sharing (Public, Scientific, and Private). Finally, a debrief explained the experiment and thanked them for their participation. A video walkthrough of the experiment can be found online at https://github.com/InfantLab/little-drummers#experiment-walkthrough.

Data Analysis

Our novel approach to coding infant rhythmic behaviour uses the OpenPose software for markerless motion tracking (Cao et al., 2021). OpenPose is a deep learning model that has been trained to identify multiple human figures in images and video that is widely used in research settings (e.g., Fujiwara and Yokomitsu, 2021; Kim et al., 2021; Zeng and Chen, 2021). When presented with a video it analyses each frame independently, labelling all people present. For each identified person, it can tag up to 25 key points on the body (depending on visibility) and has an optional hand-model that identifies up to 21 key points per hand. Each identified marker is given as x and y coordinates within the frame (see **Figure 2**). OpenPose also has the ability to label face markers but this was not used in the current project. OpenPose is an open source project that is free to use in non-commercial applications (For further information, see https://github.com/CMU-Perceptual-Computing-Lab/openpose).

Using the keypoint data generated by OpenPose, the vertical displacement of hand can then be extracted and the power and frequency of infants rhythmic entrainment can be measured using Fast Fourier Transforms (see **Figure 3**). However, to make use of the raw data generated by OpenPose, a substantial amount of additional data processing is required. In this section, we briefly walk-through the steps involved in data transformation, cleaning and analysis.

All analysis was performed in Python using the Jupyter notebooks interface which creates an annotated analysis script, allowing for direct reproduction of all analysis steps. A general toolkit for performing these steps and a short tutorial are freely available (open source) at https://github.com/InfantLab/VASC. The specific versions of libraries, scripts and their output for this dataset are found at https://github.com/InfantLab/little-drummers.

Step 1: Motion Capture Video Conversion

Lookit provides video for each individual trial as a separate file with a unique identifier per child and condition. We downloaded all videos for all children and all conditions. The Step 1 script then passed each video to OpenPose. It processes videos frame-by-frame outputting a single structured data file per frame (JSON format), containing all key point information (screen X- and Y-coordinates and a percentage confidence score per key point). OpenPose processed all 402 videos in our dataset, producing

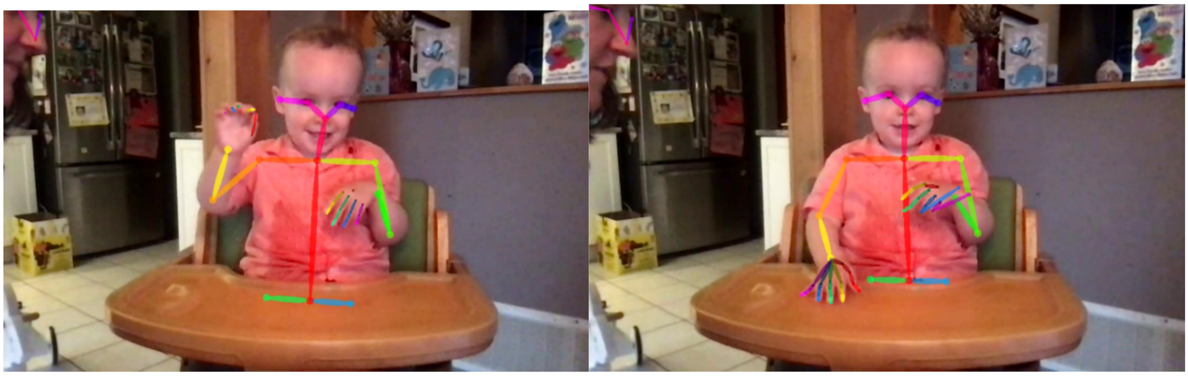


FIGURE 2 | A toddler filmed drumming at home using the Lookit platform. Parents supervise the study following instructions on laptop screen. Wireframe overlay created using OpenPose (Cao et al., 2021).

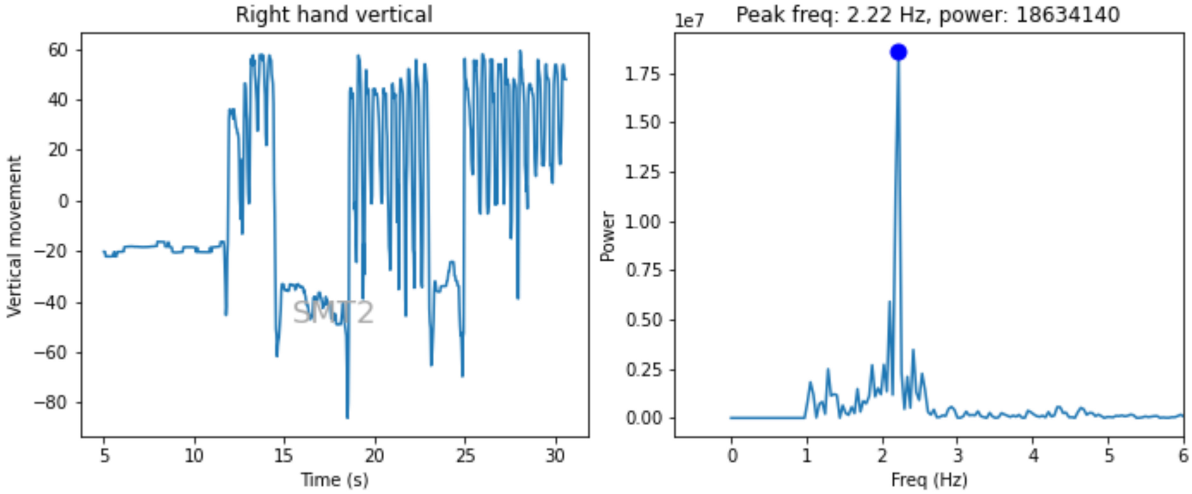


FIGURE 3 | (A) The graph shows the vertical movement of the toddler's right hand using averaged OpenPose data over their second spontaneous motor tempo trial, which contained three bursts of drumming. **(B)** The graph shows the power spectrum derived using fast fourier transform of the data, revealing a spontaneous motor frequency of 2.22 Hz.

approximately 250,000 JSON files. Next the script parsed the outputted JSON to extract and combine all data into a single multi-dimensional NumPy array. We save this in a compressed format to pass to Step 2.

Step 2: Data Cleaning and Collation

The OpenPose software has limitations so a considerable amount of data cleaning is required. For each video the experimenters had to manual check the data generated by OpenPose and make sure it correctly identified the drumming infant. The biggest problem is that OpenPose operates on a per frame basis and so can have inconsistent labelling between frames. For example, it may label infant and caregiver as person 0 and person 1 in one frame but as person 1 and person 0 in the next. Additionally, there may be additional people who temporarily enter the field of view or OpenPose can include false negatives (failure to label person) and false positives (labelling background scenery as a "ghost" person). The step 2 script provides visual inspection tools for manual corrections. It allows the experimenter to see plots of the average locations of the figures in the video across the whole time-series. Mislabelling shows up as large jumps and cross-overs in the plots. The user can select the affected frame and relabel the data and remove erroneous false positives. To speed up this process a set of simple automatic algorithms to relabel the figures consistently were created. These operate by comparing each frame to the one previous and matching the figure labels by location or by figure size. More details can be found in the online tutorial. The output of this process is a multidimensional time-series of consistently labelled body and hand points for each infant in each trial. We saved these as multi-index Pandas dataframes to pass to Step 3b.

Step 3a: Manual Tagging of Drumming Trials

To identify videos without drumming, experimenters watched all videos and manually coded infant behaviour. For each trial we recorded whether the infants hands were visible, whether they drummed with either left hand, right hand or both, and whether there was any interference from the caregiver. Trials were retained for analysis if infants took at least four consecutive strikes (<2,000 ms between hits) of the surface in front of them, and discarded if an adult moved the infants' hand themselves. 18 infants demonstrated no drumming, from the remaining 50 participant drumming was seen in a total of 208 trials (Mean = 4.16 per infant). These were labelled in a spreadsheet that was read by Step 3b.

Step 3b: Extracting Rhythmic Hand Data With Fast Fourier Transforms

For each trial we first linearly interpolated any missing data in the time series of marker points. Across the 208 trails selected in Step 3a, this affected less than 1% of the data. Then we found the average location of left and right hands by creating weighted sums of the hand and wrist X- and Y-coordinates. We use this derived y-coordinate as a measure of the vertical movement of the drumming hand over time. Next, we subtracted the mean vertical displacement in a given trial from each of the time-series to give a measure of movement. This data was then transformed into a power spectrum using the discrete Fourier transform routines in SciPy (Virtanen et al., 2020). To filter out larger, non-rhythmical movements we cut off the power spectrum below 1 Hz. The maximum power was found and the corresponding frequency saved as the tempo for that trial for each hand. The infants "best" hand (i.e., with an FFT with the highest power) is used in further analyses. See **Figures 3A,B** for example of vertical movement and corresponding power spectrum.

TABLE 1 | Inter-onset-interval (IOI) of infant drumming derived from FFT.

Target IOI (ms)	N	Mean	SD	Median	SE
700	38	712.208	204.097	695.797	33.109
600	38	649.802	182.196	596.068	29.556
500	41	599.066	190.883	523.657	29.811
400	35	599.418	192.16	549.679	32.481
SMT1	30	640.453	189.621	607.89	34.62
SMT2	26	531.843	152.001	491.127	29.81

N reflects number of infants that drummed in each trial.

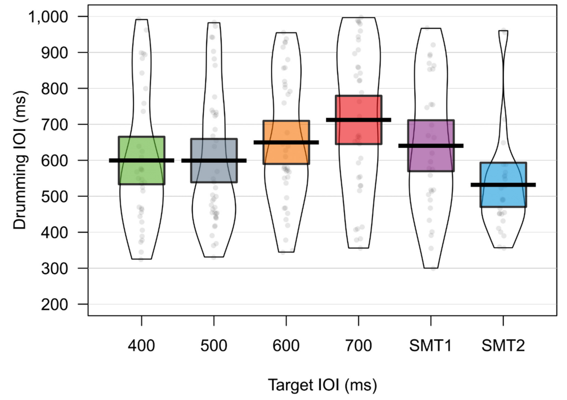


FIGURE 4 | Figure shows violin plots and jittered raw data of the rate of infant drumming in each experimental condition. The Mean is signified by a bold line, and the 95% CI is shaded. Whilst infant drumming does not directly correspond to the rate of presentation, infants appear to be adapting their rate of drumming to be slower in the slow trials (600 and 700 ms IOI).

RESULTS

A total of 68 infants completed the experiment. Of these, 18 demonstrated no visible drumming and are excluded from further analysis. This included 14 who were partially off-camera, and a further four who presented no drumming in any trial. The remaining 50 infants provided drumming data in 208 different trials averaging 1.12 SMT trials each and 3.04 drumming trials across the different conditions. For each trial, drumming frequency was recorded for subsequent analysis.

Rate of Drumming

The median IOI of infant drumming in silence (SMT) trial 1 was 608 ms, decreasing to 491 ms at trial 6. Infant drumming during stimulation appears to show some tempo flexibility (i.e., slower drumming to longer IOIs), see **Table 1** and **Figure 4**.

In order to test whether infants were indeed drumming at different rates across the different IOI trial types, a linear mixed effects model with a random slope on participant was conducted in RStudio Team (2020.09.01), RStudio (2021.09.01) (RStudio Team, 2020), with the specification "rate of drumming ∼ trial type + (1| participant)". The slowest, 700 ms condition was taken as the basecase. An ANOVA using Satterthwaite's method reveals a highly significant main effect of trial IOI ($F = 3.573$, $p = 0.004$). Full results are shown in **Table 2**. *Post-hoc* tests show that all tempi except 600 ms and SMT1 elicited significantly faster drumming than in the 700 ms trial.

Tempo Mismatch

In order to quantify how accurately infants were tempo-matching during the different tempo trials, we calculated a tempo mismatch score as the rate of infant drumming minus the target IOI. Positive mismatch values therefore reflect infants drumming slower than the target IOI, and negative values reflect faster than target drumming. Tempo mismatch is plotted in **Figure 5A**.

For further analysis, the absolute (i.e., non-signed) tempo mismatch is taken as the dependent variable, see **Figure 5B**. Descriptive statistics are shown in **Table 3**. The mismatch between infant drumming and the target IOI was approximately 150–200 ms, across the four target tempi.

If infants are not tempo-matching, we would expect to see higher "accuracy," or lower tempo-mismatch, in the trials with a target tempo closer to their SMT. To test for this pattern a further linear mixed model with a random slope on participant was conducted, with the specification "tempo mismatch ∼ trial IOI + (1| participant)". Data provided in the SMT conditions were not included, as there was no target for infants to match. We do not find a main effect of trial IOI ($F = 1.605$, $p = 0.192$) nor any *post hoc* differences, suggesting

TABLE 2 | Table of coefficients for linear mixed effect models.

	Rate of drumming			Tempo mismatch		
Predictors	***Estimates***	***CI***	***p***	***Estimates***	***CI***	***p***
(Intercept)	713.88	654.31 to 773.45	**<0.001**	175.84	132.17 to 219.50	**<0.001**
600 ms	−63.87	−141.97 to 14.23	0.108	−24.98	−84.19 to 34.23	0.406
500 ms	−114.42	−190.52 to −38.31	**0.003**	−19.26	−77.13 to 38.61	0.512
400 ms	−110.05	−190.17 to −29.94	**0.007**	35.73	−24.94 to 96.40	0.246
SMT 1	−72.96	−156.93 to 11.00	0.088			
SMT 2	−168.01	−255.25 to −80.78	**<0.001**			
Random effects						
σ^2	29,036.09			16,769.25		
τ_{00}	6,176.53 $_{\text{Participant}}$			1,850.26 $_{\text{Participant}}$		
ICC	0.18			0.10		
N	50 $_{\text{Participant}}$			49 $_{\text{Participant}}$		
Observations	208			152		
Marginal R^2/Conditional R^2	0.068/0.232			0.028/0.125		

Bold values indicate significant values at p < 0.05.

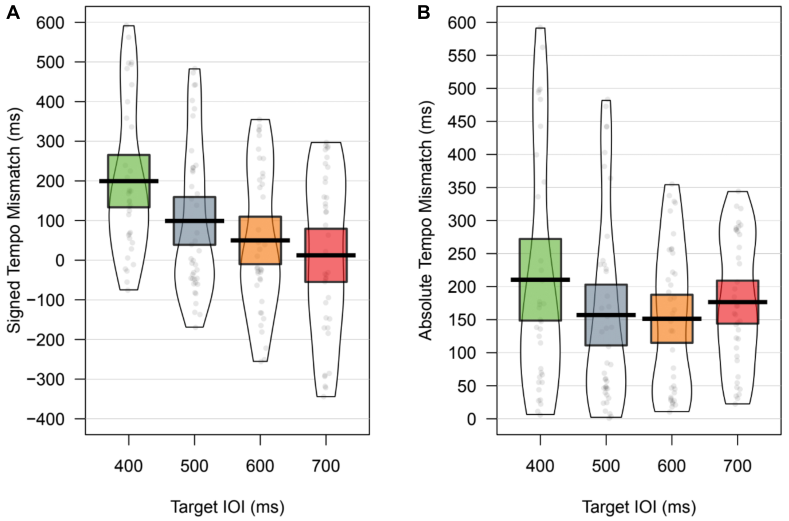


FIGURE 5 | Panel **(A)** depicts the tempo mismatch of infant drumming (IOI of drumming – Target IOI). Positive numbers reflect drumming slower than target, and negative numbers reflect drumming faster than target. Values close to zero reflect accurate drumming. Panel **(B)** depicts the absolute (non-signed) tempo mismatch. The Mean is signified by a bold line, and the 95% CI is shaded.

that infants performed similarly across conditions (all p n.s., see **Table 2**).

Finally, as infant SMT in trial SMT1 was slower than predicted for this age group, manual video coding was used to determine if SMT was related to the number of hits performed in each trial. Descriptive statistics for all trials are presented in **Table 4**.

Infants were indeed seemingly less engaged in the SMT trials (where there was no drumming video to follow), reflected in both a lower N of infants participating in these trials, and a lower number of hits by those who did participate in SMT1. Notably, after the presentation of drumming videos, in SMT2, infants who did participate were drumming to a similar extent as during the test trials. It is therefore possible that the slower than expected SMT for this age group recorded in SMT1 is the product of infants not producing a reliable estimate due to insufficient data. If this were the case, we might expect that infants who drummed more in this trial would have a faster SMT. However, SMT1 is not significantly correlated with the number of hits produced, with evidence for the null hypothesis of no relationship between number of hits and rate of drumming [$r(28) = 0.024$, $p = 0.899$, $BF_{10} = 0.229$].

DISCUSSION

Here we demonstrate the feasibility of using online measurement of infant drumming as an index of infant rhythmic skill. Infants in our sample showed age-appropriate tempo-flexibility, drumming faster to faster tempi and slower to slower tempi. Infants showed evidence of adjusting their rate of drumming away from their intrinsic rate of movement, or Spontaneous Motor Tempo (SMT). However, infants were not close to adult

TABLE 3 | Tempo mismatch of infant drumming at each target IOI.

Target IOI (ms)	N	Mean	SD	Median	SE
700	38	176.558	98.963	165.026	16.054
600	38	151.344	110.538	153.469	17.932
500	41	157.076	145.624	109.394	22.743
400	35	210.369	179.738	149.679	30.381

TABLE 4 | Number of drum hits made by infants in each trial type.

Target IOI (ms)	N	Mean	SD	Median	Minimum	Maximum
400	38	23.171	12.215	22.000	4	45
500	38	25.525	13.263	24.500	6	53
600	41	22.816	13.096	24.000	4	51
700	35	18.395	10.709	16.000	4	49
SMT1	30	16.800	7.284	17.000	4	35
SMT2	26	23.500	13.064	20.000	6	51

levels of tempo-matching (in the range of tens of milliseconds), showing an average mismatch of greater than 100 ms. This level of tempo-matching is in line with prior observations of 18-month-olds (Rocha and Mareschal, 2017). Previously, 24-month-olds have been shown to synchronise their drumming, but only when interacting with a live, social partner (see Kirschner and Tomasello, 2009; Yu and Myowa, 2021). Whilst our stimuli involved a video recording of a human hand drumming, it was not an overtly "social" signal. Nonetheless, our results show that it is possible to gain a behavioural index of infant sensorimotor synchronisation using a low-cost and accessible open-source platform.

Methodologically this work has multiple strengths. Firstly, by leveraging the Lookit platform, we were able to collect high resolution behavioural data with good ecological validity from a notoriously challenging age group, with minimal experimenter oversight. Because data were collected at home, infants were in a highly familiar setting and caregivers could run the study at time of their choosing. Parents could even abandon an attempt and try at a later time. Secondly, despite the variability of testing circumstances, we were able to get good compliance with instructions and engagement with the task. For example, in several cases older siblings were present but data could be screened for interference and distraction. Finally, the data processing pipeline provided objective measures of movement and rhythmicity with a relatively small amount of manual coding.

Our study demonstrates that markerless motion capture data with infants can be collected in a home with no specialist equipment. The data quality was sufficient for us to extract measures of infant motor tempo with automated Fourier transforms. This is a promising proof of concept, particularly given that the OpenPose model was trained primarily with adult data (Cao et al., 2021). The best infants participants produced data comparable to adult pilot participants (see "**Supplementary Material**"). It is important to observe that for infants with little or no drumming the Fourier method will not automatically extract a drumming frequency due to lower frequency noise. If additional manual coding was used to tag periods of drumming then more accurate tempo scores could be extracted. Future work will develop this functionality. Further, the trial lengths were purposefully short (20-s), in order to minimise attrition from the study in this unique testing scenario where the experimenter is not present, but this may not have allowed enough time for all infants to provide data. The toddlers mostly tolerated the length very well, and increasing the trial length to 1-min may allow more time for the infants to "warm up" to the drumming and produce enough data for analysis.

One general limitation of this approach to motion capture is that data is only two-dimensional, in the plane of the camera (X and Y coordinates). OpenPose does have the capability to combine data from multiple cameras to reconstruct three-dimensional poses and movement (Nakano et al., 2020). However, this only works in highly optimised conditions with a single participant in view, making it unsuitable for our web-based, parent supervised paradigm. However, the design of data processing workflow allows us to substitute alternative, improved pose estimation algorithms. See Wang et al. (2021) for a review of recent advances.

Future iterations of this paradigm can benefit from clearer instructions to both the caregiver and infant. Of the 68 infants recruited for the study, 18 did not provide drumming data in any of the experimental conditions. Further, the SMT exhibited in the first trial (SMT1) was slower than we would have predicted, and indeed slower than the SMT produced in SMT2, which was recorded at the end of the experiment. The caregivers were asked to prompt the infant "Can you drum for me?", "Can you show me how you drum?", but it is possible this was not sufficient for some infants to understand the task. Some parents reported that their child was unfamiliar with the word "drum," while other infants simply responded "No!". The original design did not include a video example, in an effort to not bias the infants' SMT. However, a possible solution would be to include a video montage where two or more infants are drumming side-by-side, giving a clear demonstration of the action expected (repeated whole hand hits), whilst not giving a strong timing signal, adding clarity and motivation to participate. Further, if infants do not produce drumming during SMT1, it could be possible for the parent to replay the demonstration video, and the child attempt the trial again. This would better enable the collection of a representative SMT.

The largest source of missing data was poor camera angles that meant the infant hand could not be tracked ($N = 14$). Lookit is optimised for desktop/laptop computers (i.e., not tablets or mobile phones), and anecdotally, the angle of webcams is normally optimised for centring the adult face. The infant hand is considerably lower in the camera's field. Prior to commencing the study, the caregiver was shown a preview of their camera angle and asked to check that their infant was in shot. However, as the infant was not yet drumming at this point, it may not have been obvious as to whether the area that would be drummed upon was in view. Future iterations can preview the view of the webcam between each experimental trial, such that the caregiver can adapt their angle as needed. However, care must be taken not to make the task instruction too complex or demanding for caregivers to follow.

A further limitation of remote asynchronous testing that we did not foresee was that not all children were sat at a table that provided good auditory and haptic feedback from the children's drumming. Subtle differences in the surface infants were drumming on could potentially affect their ability to adapt their behaviour to the auditory stimuli we asked them to synchronise with. Future iterations could therefore also ask the parent and child to "sound check" their drumming to ensure they are getting auditory feedback when they hit the table. Further, if infants are recruited for a study outside of Lookit (e.g., where they are already participating in a lab based session), the paradigm could also be used either within the lab, or at home with a standardised surface (e.g., drum), provided by the research team, to ameliorate these differences.

The study provides promising basis for further exploration of other domains. Firstly, rhythmic movements of the whole body could be examined, allowing investigation of dancing and entrainment to music. OpenPose has been used in this context in laboratory studies with adults (Zeng and Chen, 2021). Our work demonstrates that this could be feasibly done with relatively large samples of young children. Even more promising would be to investigate synchrony between individuals, especially in the context of bonding and responsive caregiving where current human coding measures are labour intensive and lack standardization and predictive validity (Lotzin et al., 2015). Automated solutions have been a goal of social signal processing for a long time (Chetouani et al., 2017) and movement data has emerged as a promising signal (Egmose et al., 2017; López Pérez et al., 2017). However, progress has been slow and most methods are not suitable for field data (Chu et al., 2015). Therefore, we are currently adapting our methods for use in this context.

In summary, here we provide a successful proof-of-concept that we can extract the rate and accuracy of infant drumming from home video, using largely automated and fully open-source procedures. In an initial study of 2-year-old toddlers, we find evidence for tempo-flexibility, but not synchronisation, in response to an isochronous external beat presented at different tempi. The overarching goal of the current work was to develop a tool that is suitable to assess rhythmic movement in very young children, which can be employed at scale, and potentially even identify children at risk of neurodevelopmental disorders, including speech and language difficulties. Such longitudinal assessments, that are appropriate over developmental time, are key to understanding the mechanistic profiles of rhythm impairments across a broad range of neurodevelopmental disorders (Lense et al., 2021). Now that the feasibility of this online approach has been demonstrated, future work can refine the procedure, and further develop this promising tool for deeper insights into infant behaviour.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Goldsmiths, University of London. Verbal informed consent to participate in this study was provided by the participants' legal guardian via a recorded video statement. Written informed consent was obtained from the minor(s)' legal guardian, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SR and CA co-designed the study, collected and analysed the data, wrote the manuscript, and approved the submitted version.

FUNDING

This project was supported by the Royal Academy of Engineering Global Challenges Research Fund Grant: Frontiers of Development - Tranche 2 - FoDSF\1920\2\100020, awarded to CA.

ACKNOWLEDGMENTS

We would like to thank all of the families who took part, and followed our instructions so well at home! We would also like to thank James Kellett Smith for his help in creating the drumming videos, and to Kim Scott and everyone working so hard to support the Lookit platform, which has been vital for infant research during the global pandemic.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.897230/full#supplementary-material

REFERENCES

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *Neuroimage* 247:118698. doi: 10.1016/j.neuroimage.2021.118698

Bobin-Bègue, A., and Provasi, J. (2008). Régulation rythmique avant 4 ans: effet d'un tempo auditif sur le tempo moteur. *Ann. Psychol.* 108, 631–658.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., and Sheikh, Y. (2021). OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. *Pattern Anal. Mach. Intell.* 43, 172–186. doi: 10.1109/TPAMI.2019.2929257

Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., and Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14559–14564. doi: 10.1073/pnas.1406219111

Chetouani, M., Delaherche, E., Dumas, G., and Cohen, D. (2017). "Chapte 15 – Interpersonal synchrony: from social perception to social interaction," in *Social Signal Processing*, eds J. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli (Cambridge: Cambridge University Press), 202–212. doi: 10.1017/9781316676202.015

Chu, W.-S., Zeng, J., De la Torre, F., Cohn, J. F., and Messinger, D. S. (2015). "Unsupervised synchrony discovery in human interaction," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 3146–3154. doi: 10.1109/ICCV.2015.360

Corriveau, K. H., and Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex* 45, 119–130. doi: 10.1016/j.cortex.2007.09.008

Cumming, R., Wilson, A., Leong, V., Colling, L. J., and Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments. *Front. Hum. Neurosci.* 9:672. doi: 10.3389/fnhum.2015.00672

Egmose, I., Varni, G., Cordes, K., Smith-Nielsen, J., Væver, M. S., Køppe, S., et al. (2017). Relations between automatically extracted motion features and the quality of mother-infant interactions at 4 and 13 months. *Front. Psychol.* 8:2178. doi: 10.3389/fpsyg.2017.02178

Falk, S., Müller, T., and Dalla Bella, S. (2015). Non-verbal sensorimotor timing deficits in children and adolescents who stutter. *Front. Psychol.* 6:847. doi: 10.3389/fpsyg.2015.00847

Fraisse, P. (1982). Rhythm and tempo. *Psychol. Music* 1, 149–180.

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., and Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS One* 9:e97680. doi: 10.1371/journal.pone.0097680

Fujiwara, K., and Yokomitsu, K. (2021). Video-based tracking approach for nonverbal synchrony: a comparison of motion energy analysis and OpenPose. *Behav. Res. Methods* 53, 2700–2711. doi: 10.3758/s13428-021-01612-7

Kim, W., Sung, J., Saakes, D., Huang, C., and Xiong, S. (2021). Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose). *Int. J. Ind. Ergon.* 84:103164. doi: 10.1016/j.ergon.2021.103164

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., and Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wires Cogn. Sci.* 11:e1528. doi: 10.1002/wcs.1528

Lee, H.-Y., Sie, Y.-S., Chen, S.-C., and Cheng, M.-C. (2015). The music perception performance of children with and without dyslexia in Taiwan. *Psychol. Rep.* 116, 13–22. doi: 10.2466/15.28.PR0.116k15w8

Lense, M. D., Ladányi, E., Rabinowitch, T.-C., Trainor, L., and Gordon, R. (2021). Rhythm and timing as vulnerabilities in neurodevelopmental disorders. *Philos. Trans. R. Soc. B Biol. Sci.* 376:20200327. doi: 10.1098/rstb.2020.0327

López Pérez, D., Leonardi, G., Niedźwiecka, A., Radkowska, A., Rączaszek-Leonardi, J., and Tomalski, P. (2017). Combining recurrence analysis and automatic movement extraction from video recordings to study behavioral coupling in face-to-face parent-child interactions. *Front. Psychol.* 8:2228. doi: 10.3389/fpsyg.2017.02228

Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., et al. (2015). Observational tools for measuring parent–infant interaction: a systematic review. *Clin. Child Fam. Psychol. Rev.* 18, 99–132. doi: 10.1007/s10567-015-0180-z

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *J. Exp. Psychol. Gen.* 135, 348–367. doi: 10.1037/0096-3445.135.3.348

Nakano, N., Sakura, T., Ueda, K., Omura, L., Kimura, A., Iino, Y., et al. (2020). Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. *Front. Sports Act. Living* 2:50. doi: 10.3389/fspor.2020.00050

Olander, L., Smith, A., and Zelaznik, H. N. (2010). Evidence that a motor timing deficit is a factor in the development of stuttering. *J. Speech Lang. Hear. Res.* 53, 876–886. doi: 10.1044/1092-4388(2009/09-0007)

Persici, V., Stucchi, N., and Arosio, F. (2019). "Predicting the future in rhythm and language: the anticipation abilities of a group of Italian-speaking children," in *Proceedings of GALA 2017: Language Acquisition and Development*, eds P. Guijarro-Fuentes and C. Suárez-Gómez (Newcastle upon Tyne: Cambridge Scholars Publishing), 451–467.

Politimou, N., Dalla Bella, S., Farrugia, N., and Franco, F. (2019). Born to speak and sing: musical predictors of language development in pre-schoolers. *Front. Psychol.* 10:948. doi: 10.3389/fpsyg.2019.00948

Rios-Lopez, P., Molinaro, N., and Lallier, M. (2019). Tapping to a beat in synchrony predicts brain print sensitivity in pre-readers. *Brain Lang.* 199:104693. doi: 10.1016/j.bandl.2019.104693

Rocha, S., and Mareschal, D. (2017). Getting into the groove: the development of tempo-flexibility between 10 and 18 months of age. *Infancy* 22, 540–551. doi: 10.1111/infa.12169

Rocha, S., Southgate, V., and Mareschal, D. (2021b). Infant spontaneous motor tempo. *Dev. Sci.* 24:e13032. doi: 10.1111/desc.13032

Rocha, S., Attaheri, A., Choisdealbha, Á. N., Brusini, P., Flanagan, S. A., Mead, N., et al. (2021a). Infant sensorimotor synchronisation to speech and non-speech rhythms: a longitudinal study. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/jbrga

RStudio Team (2020). *RStudio: Integrated Development for* R. Boston, MA: RStudio, PBC.

Scott, K. M., and Schulz, L. E. (2017). Lookit (part 1): a new online platform for developmental research. *Open Mind* 1, 4–14. doi: 10.1162/opmi_a_00002

Thomson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol.* 102, 120–129. doi: 10.1016/j.jphysparis.2008.03.007

Vanneste, V., Pouthas, J. H., and Wearden, S. (2001). Temporal control of rhythmic performance: a comparison between young and old adults. *Exp. Aging Res.* 27, 83–102. doi: 10.1080/03610730125798

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.

Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z., et al. (2021). Deep 3D human pose estimation: a review. *Comput. Vis. Image Underst.* 210:103225. doi: 10.1016/j.cviu.2021.103225

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: a study of 18-to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/infa.12403

Zeng, H., and Chen, W. (2021). "An evaluation approach of multi-person movement synchronization level using OpenPose," in *Proceedings of the 2021 40th Chinese Control Conference (CCC)*, Shanghai, 3900–3905. doi: 10.23919/CCC52363.2021.9549562

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773. doi: 10.1073/pnas.1000121107

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

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fnins.2022.901013

A neural marker of rapid discrimination of facial expression in 3.5- and 7-month-old infants

Fanny Poncet[1,2]*, Arnaud Leleu[1], Diane Rekow[1], Fabrice Damon[1], Milena P. Dzhelyova[3], Benoist Schaal[1], Karine Durand[1], Laurence Faivre[4], Bruno Rossion[3,5] and Jean-Yves Baudouin[6]*

[1]Development of Olfactory Communication and Cognition Laboratory, Centre des Sciences du Goût et de l'Alimentation, CNRS, Université Bourgogne Franche-Comté, INRAE, Institut Agro, Dijon, France, [2]Université Grenoble Alpes, Saint-Martin-d'Hères, France, [3]Université de Lorraine, CNRS, CRAN–UMR 7039, Nancy, France, [4]Inserm UMR 1231 GAD, Genetics of Developmental Disorders, and Centre de Référence Maladies Rares "Anomalies du Développement et Syndromes Malformatifs," FHU TRANSLAD, CHU Dijon and Université de Bourgogne-Franche Comté, Dijon, France, [5]Service de Neurologie, Université de Lorraine, CHRU-Nancy, Nancy, France, [6]Laboratoire "Développement, Individu, Processus, Handicap, Éducation", Département Psychologie du Développement, de l'Éducation et des Vulnérabilités, Institut de Psychologie, Université de Lyon, Université Lumière Lyon 2, Bron, France

OPEN ACCESS

EDITED BY
Chiahuei Tseng,
Tohoku University, Japan

REVIEWED BY
Stefania Conte,
University of South Carolina,
United States
Chiara Turati,
University of Milano-Bicocca, Italy

*CORRESPONDENCE
Fanny Poncet
fanny.poncet@univ-grenoble-alpes.fr
Jean-Yves Baudouin
j.baudouin@univ-lyon2.fr

SPECIALTY SECTION
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

RECEIVED 21 March 2022
ACCEPTED 29 July 2022
PUBLISHED 18 August 2022

CITATION
Poncet F, Leleu A, Rekow D, Damon F, Dzhelyova MP, Schaal B, Durand K, Faivre L, Rossion B and Baudouin J-Y (2022) A neural marker of rapid discrimination of facial expression in 3.5- and 7-month-old infants. *Front. Neurosci.* 16:901013. doi: 10.3389/fnins.2022.901013

Infants' ability to discriminate facial expressions has been widely explored, but little is known about the rapid and automatic ability to discriminate a given expression against many others in a single experiment. Here we investigated the development of facial expression discrimination in infancy with fast periodic visual stimulation coupled with scalp electroencephalography (EEG). EEG was recorded in eighteen 3.5- and eighteen 7-month-old infants presented with a female face expressing disgust, happiness, or a neutral emotion (in different stimulation sequences) at a base stimulation frequency of 6 Hz. Pictures of the same individual expressing other emotions (either anger, disgust, fear, happiness, sadness, or neutrality, randomly and excluding the expression presented at the base frequency) were introduced every six stimuli (at 1 Hz). Frequency-domain analysis revealed an objective (i.e., at the predefined 1-Hz frequency and harmonics) expression-change brain response in both 3.5- and 7-month-olds, indicating the visual discrimination of various expressions from disgust, happiness and neutrality from these early ages. At 3.5 months, the responses to the discrimination from disgust and happiness expressions were located mainly on medial occipital sites, whereas a more lateral topography was found for the response to the discrimination from neutrality, suggesting that expression discrimination from an emotionally neutral face relies on distinct visual cues than discrimination from a disgust or happy face. Finally, expression discrimination from happiness was associated with a reduced activity over posterior areas and an additional response over central frontal scalp regions at 7 months as compared to 3.5 months. This result suggests developmental changes in the processing of happiness expressions as compared to negative/neutral ones within this age range.

KEYWORDS

infant, visual perception, facial expression of emotions, fast periodic visual stimulation, EEG, development

Introduction

A large body of research suggests that the first year of life is critical in the development of facial emotion perception [e.g., Campos et al. (1983); for reviews, see Nelson (1987), Leppänen and Nelson (2009), and Maria et al. (2018)]. Some studies have reported early abilities for discriminating facial expressions (e.g., Field et al., 1982; Farroni et al., 2007; Addabbo et al., 2018), while other studies have challenged this view [Kaitz et al., 1988; Oostenbroek et al., 2016; see also Soussignan et al. (2018)]. In fact, the ability of infants to discriminate, recognize, and adapt their own behavior to the facial expressions of others seems to develop gradually over infancy and childhood. For most facial expressions, while discriminative abilities have been evidenced at an early age, the recognition of a specific emotion and its meaning has not been clearly established before the second half of the first year (Walker-Andrews, 1997; Leppänen and Nelson, 2009).

Many authors have delineated significant changes in the processing of emotional facial expression during the first year (Campos and Stenberg, 1981; Oster, 1981; Walker-Andrews, 1997, 2005; Leppänen and Nelson, 2009; Leppänen, 2011; Quinn et al., 2011). Infants seem to start differentiating facial expressions by progressively decoding specific cues and configurations, before being able to attribute emotional meaning to specific patterns of facial actions contingent with repeated social interactions. For instance, some studies reveal that newborns discriminate between dynamic facial expressions of happy and disgusted [but only after being habituated to a happy or disgusted one: Addabbo et al. (2018)]. From 3 to 4 months of age, infants visually discriminate some facial expressions, especially happiness from frowning, anger, sadness, or neutrality (LaBarbera et al., 1976; Young-Browne et al., 1977; Barrera and Maurer, 1981; Haviland and Lelwica, 1987). However, they do not reliably "categorize" facial expressions (i.e., generalize an emotion category across different identities or views) before 5–7 months of age (Caron et al., 1982; Serrano et al., 1992; Kotsoni et al., 2001; Bornstein and Arterberry, 2003; Bornstein et al., 2011). Infants react to positive emotions (smiling more to happy than to neutral and negative faces) from about 3 months of age (e.g., Millar, 1976; Legerstee, 1997; Soussignan et al., 2018), and progressively extend this reaction to negative emotions during the second part of the first year [e.g., Kaiser et al. (2017); for reviews, see Campos and Stenberg (1981) and Oster (1981)]. In particular, with the development of referential looking behaviors between 7 and 12 months of age (Rochat and Striano, 1999), infants progressively refer to the negative facial expressions of adults and adjust their behavior accordingly (Feinman and Lewis, 1983; Klinnert et al., 1986; Campos et al., 2003). For instance, fearful faces elicit an adult-like attentional orienting over neutral or happy faces in 7-month-olds (e.g., Nelson et al., 1979; Nelson and Dolgin, 1985; Kotsoni et al., 2001; Peltola et al., 2008, 2009, 2013). It is generally suggested that the ability to assign meaning to facial expressions emerges from this age of 7 months onward, after experience-expectant developmental processes (Leppänen and Nelson, 2009; Leppänen, 2011).

In line with the behavioral literature (Walker-Andrews, 1997; Quinn et al., 2011), studies on brain activity further indicate critical differences between the different expressions in the early development of facial expression processing from the middle of the first year. At 7 months of age, an "adult-like neural circuitry" is engaged to process some emotional faces [sad and happy faces: Rotem-Kohavi et al. (2017); fearful faces: Leppänen and Nelson (2009)]. EEG studies have revealed that, as in adults, looking facial expressions elicits sensorimotor activity at 7 months of age, but only for happy faces (Quadrelli et al., 2021). At this age, the brain response to happiness is different from negative expressions like angry faces for the temporal aspect (both right-lateralized), and the response to angry faces is associated with higher scores on a Negative Affect temperamental dimension (Quadrelli et al., 2019). In both these two studies, the stimulus dynamicity has been observed as more efficiently processed by functional brain networks at this age as compared to static stimuli. For fear, modulations were observed over medial occipital and occipito-temporal sites (i.e., related to the visual processing of faces) in studies measuring event-related potentials (ERPs) (e.g., de Haan and Nelson, 1999; Halit et al., 2004; Leppänen et al., 2007; Xie et al., 2019). Other modulations were reported over central frontal regions in response to fear, happiness or anger in ERP studies [Nelson and de Haan (1996), Leppänen et al. (2007), and Xie et al. (2019): with an effect emerging at 5 months of age before becoming well-established at 7 months of age] or studies recording functional near-infrared spectroscopy [fNIRS; Minagawa-Kawai et al. (2008), Fox et al. (2013), and Bayet et al. (2021): especially for happiness over temporo-parietal sites]. ERP studies relate attentional orienting toward salient stimuli (Nelson, 1994; de Haan, 2007) to cortical sources either in the prefrontal and anterior cingulate cortices (Reynolds and Richards, 2005), or in the posterior cingulate cortex/precuneus and temporal areas (Guy et al., 2016; Xie et al., 2019). In fNIRS studies, greater activity occurs in the medial prefrontal cortex (mPFC) for smiling over neutral faces [in 7-month-olds: Fox et al. (2013); in 9- to 13-month-olds: Minagawa-Kawai et al. (2008)], the mPFC being credited to play an important role in the early acquisition of socio-cognitive skills (Grossmann, 2013).

Overall, the studies reviewed above suggest that the nature and topography of the brain response to facial expressions differ between the facial expressions, at least in infants in the second half of the first year [that could start to emerge at 5 months: e.g., Xie et al. (2019)], with posterior responses, temporo-parietal and/or central frontal regions according to the emotional expressions. Some studies suggest that more central frontal responses emerge with age, posterior regions responding more at an earlier stage (e.g., Xie et al., 2019). However, studies using standard ERP and fNIRS approaches report quite variable results, with the difficulty of isolating clear brain responses to the discrimination of different facial expressions, as also noted in adult studies [for reviews on the ERP approach on this topic, see Vuilleumier and Pourtois (2007) and Calvo and Nummenmaa (2016)]. To overcome this limitation, recent studies in adults have used fast periodic visual stimulation (FPVS) coupled with scalp EEG. Robust and specific neural responses to brief expression changes were isolated, each emotional expression being directly contrasted to a neutral face (e.g., Dzhelyova et al., 2017; Leleu et al., 2018; Matt et al., 2021) or to all other expressions (Poncet et al., 2019). This FPVS-EEG approach relies on the property of the brain to synchronize with stimuli displayed periodically (Adrian and Matthews, 1934), eliciting EEG responses at the same frequency [Regan (1989) and Norcia et al. (2015), for reviews]. This allows isolating an objective response (i.e., measured at a predefined frequency of stimulation) to a specific visual content in a few minutes of recording. By presenting stimuli at a rapid rate (i.e., the base frequency) and introducing a specific type of target stimuli periodically at a slower rate, a variation of this approach isolates a brain response that directly reflects the difference between the target stimuli and the base stimuli (i.e., without *post hoc* subtraction) [for review see Rossion et al. (2020)]. While this approach has been used to isolate face categorization abilities in the infant brain [de Heering and Rossion, 2015; Peykarjou et al., 2017; Leleu et al., 2020; Rekow et al., 2020, 2021; see also Barry-Anwar et al. (2018)], to date, it has not been used to measure the discrimination of facial expressions of emotion in this population.

Here, we used FPVS-EEG to isolate neural responses to the discrimination of specific facial expressions in 3.5- and 7-month-old infants. In particular, our goal was to dissociate the response from one expression to several other expressions in a single stimulation sequence. In classical behavioral or ERP studies, constraints related to the limited attentional availability of infants, combined with the need to have a sufficient number of trials per experimental condition, usually makes it necessary to limit the number of contrasted expressions to avoid an exponential increase in stimulation time or number of participants. With FPVS, it is possible to present one expression at one frequency and all the others at another frequency, without weighting down the procedure. Above all, the dissociation then carried out isolates the *specific* response to the expression, i.e., what differentiates it from *all the others*. For example, in the study by Poncet et al. (2019) on adult participants, every expression was displayed at a base frequency of 6 Hz (i.e., six stimuli per second), and a target expression was interspersed every 6th stimulus (i.e., at a specific frequency of 6/6 = 1 Hz). As a result, the specific brain response to the target expression was dissociated from all the other categories at the 1-Hz frequency. Another procedure to capture the differential response between an information of interest and a baseline control information is to display the first information at the base frequency and the baseline information at the oddball frequency. For example, identity discrimination was evidenced by displaying one identity at the base frequency and multiple other identities at the oddball frequency (e.g., Rossion et al., 2020). Here, we adapted this procedure to infants: the target expression was repeatedly displayed at a base frequency of 6 Hz, and other emotion categories were interspersed every 6th stimulus (at 1 Hz). We opted to present the target expression at the base frequency to reduce the visual variability of the stimulation; within six images, infants were exposed to two distinct expressions (five times the target expression and one time another randomly selected expression). This procedure provides additional time to process the target expression and reduces backward and forward masking effects (**Figure 1**). Hence, given that the brain response recorded at 1 Hz reflects a generalized *differential* activity elicited by all expression changes within a sequence, it remains a clear marker that the infant brain discriminates the facial expressions inserted at 1 Hz from the target expression displayed at 6 Hz while making the rapid stimulation less challenging for infants. Three facial expressions - neutrality, happiness and disgust–were considered. According to previous studies (e.g., de Haan and Nelson, 1999; Halit et al., 2004; Leppänen et al., 2007; Xie et al., 2019), we first hypothesized that each expression elicits a specific brain response over posterior regions at both ages, reflecting the ability of the infant brain to detect the specific visual characteristics of an expression that differentiates it from other expressions. In addition, according to the progressive integration of affective and socio-cognitive processes in the perception of emotional expressions from the second half of the first year, we expected an evolution of the brain response to the expression that acquired significance between 3.5 and 7 months, i.e., happiness, with the emergence of central frontal responses (Nelson and de Haan, 1996; Leppänen et al., 2007; Minagawa-Kawai et al., 2008; Fox et al., 2013; Xie et al., 2019). By contrast, we used disgust as a "control" expression, since the age at which infants start to understand the meaning of disgust—or whether this ability appears in infancy—is not established during the first year [see Widen and Russell (2010, 2013) and Ruba et al. (2019); even if a discriminative ability is observed in newborns between dynamic faces of disgust and happiness: Addabbo et al. (2018)]. In the literature investigating dynamic presentation of expressions as compared to static ones, evidence have been revealed that although infants from 6-months showed clear diagnostic scanning of expressions (e.g., exploration of lower part of the face, nose and mouth), a developing sophistication in scanning for negative expressions of angry and fearful but also disgusted expressions was observed from 6 to 12 months: Prunty et al. (2021). Therefore, we did not expect an evolution of the response to an expression change from disgust between 3.5 and 7 months.

Materials and methods

Participants

Twenty-one 3.5-month-old and twenty-two 7-month-old infants participated in the study. They were recruited by mail through the local birth registry. Before testing, all parents gave written informed consent. Testing was conducted according to the Declaration of Helsinki and approved by a French ethics committee (Comité de Protection des Personnes Sud-Est III–2016-A02056-45). Data from three 3.5- and four 7-month-old infants were excluded from the analyses due to less than two valid sequences for one condition or unusable EEG data because of too noisy signals. The final sample thus consisted of eighteen 3.5-month-olds (four females, mean age ± *SD* = 107.8 ± 4.7 days, range: 101–116 days), and eighteen 7-month-olds (7 females, mean age ± *SD* = 216.8 ± 6.7 days, range: 204–229 days).

Visual stimuli

Face pictures of four females depicting basic emotional expressions (disgust, happiness, anger, fear, sadness) and neutrality were selected from the KDEF database (models coded 07F, 09F, 14F, and 20F) (Lundqvist et al., 1998). They were equalized in terms of luminance and put into a medallion-shaped window to discard information from the background and hairstyle (**Figure 1**). They were displayed on a mid-level gray background (i.e., 128/255 in grayscale) with face-size adjusted to 22.6 × 17 cm. Stimuli were presented at the center of a screen at an approximate viewing distance of 57 cm, subtending a large part of the visual field of infants (i.e., 21.6 × 16.6° of visual angle). Hence, the size of the face images was close to the size of faces encountered by infants during typical social interactions (de Heering and Rossion, 2015; Leleu et al., 2020).

Procedure

The procedure was largely borrowed from experiments that successfully isolated and quantified brain signatures of facial expression categorization in adults using FPVS (Dzhelyova et al., 2017; Leleu et al., 2018; Poncet et al., 2019). Stimuli were presented without inter-stimulus interval on a 24-inch LED screen (60 Hz refresh rate, resolution: 1920 × 1,080 pixels) on a mid-level gray background (i.e., 128/255 in grayscale) at a base rate of 6 Hz (i.e., six images per second). At this rate, each stimulus lasts about 167 ms (i.e., 1 s/6). To minimize low-level repetition effects, face size randomly varied between 85 and 115% at every stimulus-onset (minimum step for size change: 5%). Each stimulation sequence presented the face of only one individual (**Figure 1**). The four individual faces were used and presented to the infants, in different stimulation sequences (i.e., with a different expression at the base). The association between an expression and an identity has been counterbalanced between infants. One target expression (i.e., disgust, happiness, or neutrality attributed to separate sequences) was presented at the base rate and a change of expression with the remaining five basic emotional expressions (except surprise) was introduced every 6th stimulus (i.e., at a lower rate of 6/6 = 1 Hz). For example, when neutrality is the target expression presented at the base rate, faces displayed at 1 Hz displayed in random order either happiness, disgust, anger, fear, or sadness. With this design, we test the discrimination from neutrality by opposing a neutral/unexpressive face (i.e., without facial actions) to all potential emotional facial expressions displayed randomly; this could isolate a marker of the detection of emotional/expressive facial configurations. By testing the discrimination from happiness, we also explore the discrimination of positive vs. negative/neutral (i.e., non-positive) emotions, all emotions but happiness being negative (4 out of 6) or unexpressive (1 out of 6) in our design. Consequently, this contrast may also capture the processing of emotional valence. Finally, when testing the discrimination from disgust, this negative expression is contrasted with expressions which are also negative (anger, fear, and sadness), but also positive (happiness) or neutral. Thus, this contrast tests the potential acquisition of a discrete status of disgust. In general, the approach dissociates two brain responses within a single stimulation sequence: the 6-Hz base frequency tags a general visual response to the rapid train of one individual face varying in size, while the 1-Hz expression-change rate tags a

FIGURE 1
Fast periodic visual stimulation in electroencephalography (FPVS-EEG) to isolate a neural marker of rapid discrimination of facial expression. Among the five basic emotional facial expressions used (disgust, happiness, anger, fear, sadness) and neutrality, the discrimination of three target expressions was tested (disgust, happiness and neutrality). A 2-s sequence of fast periodic stimulation is depicted with images from one individual face. From top to bottom: happy, disgust, and neutral facial expressions are respectively used in dedicated sequences and periodically displayed at the 6-Hz base rate (i.e., six pictures per second; 1 cycle ≈ 167 ms) without inter-stimulus interval, and the five other expressions are randomly displayed at the 1-Hz expression-change frequency (every 6th cycle = 1 s between each expression-change). Images were presented through ±15% randomized size variation at each cycle. This design thus isolates two dissociated responses at two different frequencies: a general visual response (6 Hz) to the rapid train of one individual face varying in size and an expression-change response (1 Hz) reflecting rapid (i.e., single-glance) discrimination of all inserted expressions from the target expression presented at the base rate. Source: KDEF database (models coded 07F, 09F, 14F, and 20F) (Lundqvist et al., 1998).

discrimination response selectively reflecting the perception of a change of expression from the target expression. All contrasted expressions were equally presented throughout each stimulation sequence, avoiding consecutive repetition. In sum, due to the periodic nature of the tagged EEG response that captures brain activities common to all expression changes within a sequence, the 1-Hz expression-change frequency indexes whether the infant brain discriminates five facial expressions from the target expression.

After EEG-cap placement, the infants were installed in a baby car seat in front of the screen in a dedicated light- and sound-attenuated Baby-lab. A camera placed on top of the screen continuously monitored them to check their well-being and attention to the screen. Each 34.5-s sequence started with a pre-stimulation interval of 0.5 s of blank screen, followed by a fade-in of increasing contrast for 1.833 s. Employed in earlier FPVS studies on infants (e.g., Leleu et al., 2020; Rekow et al., 2020, 2021), this sequence duration was thus well-adapted to the attentional span of infants and to technical issues (good signal to noise ratio; SNR). Full-contrast stimulation then lasted 31.167 s before a 0.833-s fade-out of decreasing contrast, and a post-stimulation interval of 0.167 s of blank screen. For each stimulation sequence, the target expression was displayed as the base rate and one of the five other expressions was randomly inserted every 6th image, at the 1-Hz rate of expression-change. Infants were constantly monitored *via* a webcam and stimulation sequences were launched when the signal was artifact-free and the infant was quietly looking at the screen. Auditory tones were transiently used to reorient infants' attention toward the screen, without contaminating the frequency-tagged responses. Each infant was presented with the three target expressions (disgust, happiness, neutrality) in distinct sequences (**Figure 1**), their presentation order being counterbalanced across infants. Infants included in the final sample were exposed to 6–13 sequences, for a total testing duration ranging from 3.5 to 7 min.

Electroencephalography recording and preprocessing

Electroencephalography was continuously recorded from 32 Ag/AgCl channels inserted in a cap (Waveguard, ANT Neuro, Netherlands) according to the 10-10 classification system. Channel AFz was used as reference during acquisition. Electrode impedance was kept below 40 kΩ and EEG was digitalized using ASAlab 4.7 (ANT Neuro, Netherlands) at a sampling rate of 1,024 Hz. EEG analyses were carried out using Letswave 6[1] running on Matlab 2017 (MathWorks, United States). Left and right mastoid (M1 and M2) and prefrontal (Fp1, Fpz, Fp2) channels were removed before processing since they were noisy or artifact-ridden for most infants.

First, EEG data were bandpass filtered at 0.1–100 Hz using a 4th order Butterworth filter and resampled to 200 Hz. Data were then cropped for each sequence in segments lasting 36 s starting from the fade-in. To reduce high-amplitude artifacts, each segment was processed using the Artifact Blocking algorithm (Mourad et al., 2007; Fujioka et al., 2011) with a threshold of ±500 μV windowed on the overall segment (Leleu et al., 2020). For four 3.5-month-old infants a remaining noisy channel was rebuilt using linear interpolation from the nearest electrodes. As a result, only one channel was interpolated for four 3.5-month-olds (average: 0.22 ± 0.43 *SD*) and none for the 7-month-olds. Data were then re-referenced according to a common average reference and EEG data were further segmented in 32-s epochs from the start of the full-contrast sequence (i.e., 32 1-Hz cycles, removing the fade-in).

Two data-driven criteria were used for each infant to remove sequences when no general response was found to the fast train of the individual face changing in size, or when the 1-Hz expression-change response across the whole scalp presented with atypical noise-corrected amplitude compared to the other sequences [for a similar procedure, see Leleu et al. (2020) and Rekow et al. (2020, 2021)]. Fast Fourier transform (FFT) was first applied to every segment and amplitude spectra were extracted for all electrodes with a frequency resolution of 1/32 = 0.03125 Hz. The first criterion was based on Z-scores calculated for each channel and each frequency bin as the difference between the signal and the mean noise (estimation from the 20 surrounding bins, 10 on each side, excluding the two immediately adjacent and the two most extreme) divided by the standard deviation of the noise. According to previous FPVS-EEG studies showing a general response of the infant visual system to a 6-Hz stimulation sequence over medial occipital sites as a general marker of adequate looking at the stimulation screen (e.g., de Heering and Rossion, 2015; Peykarjou et al., 2017; Barry-Anwar et al., 2018; Leleu et al., 2020; Rekow et al., 2020, 2021), sequences were included in the analysis when at least two electrodes were associated with a Z-score above 1.64 ($p < 0.05$, one-tailed, signal > noise) or at least one electrode with a Z-score above 2.32 ($p < 0.01$, one-tailed) over medial occipital electrodes (Oz, POz, O1, O2) for the 6-Hz base frequency or the second harmonic (i.e., 12 Hz). For the second criterion, amplitude at each frequency bin was first corrected by subtracting the mean noise amplitude estimated from the six surrounding bins [for a similar procedure, see Leleu et al. (2020) and Rekow et al. (2020, 2021)]. Here, noise was estimated from fewer frequency bins since EEG amplitude is high in the low-frequency range with a non-linear decrease as frequency increases (Fransson et al., 2013). Hence, to consider too many bins would overestimate noise level (and therefore underestimate the expression-change response) because FFT amplitude spectrum is steeper for lower than for higher frequencies around the 1-Hz frequency. The global amplitude of the brain response over the scalp (i.e., square root of the sum of squared amplitudes of all channels) was then calculated at the 1-Hz expression-change frequency. A sequence was considered atypical when its global noise-corrected amplitude was above or below 2 *SDs* of the mean of all sequences (regardless of the expression) of the infant considered individually and retained after application of the first criterion. Once these two criteria were individually applied, between 6 and 14 sequences were kept per infant, with an average of 10.11 ± 1.78 *SD* for the 3.5-month-olds (for disgust condition: 3.44 ± 0.62 *SD*, range 2-4; for happiness: 3.22 ± 0.73 *SD*, range 2–4; for neutrality: 3.50 ± 0.71 *SD*, range 2–4), and an overall rejection of only 17 out of 199 sequences (8.54%). Similarly, in the 7-month-olds group, between 6 and 14 sequences were kept per infant, with an average of 10.56 ± 2.06 *SD* (for disgust condition: 3.56 ± 0.78 *SD*, range 2–5; for happiness: 3.56 ± 0.92 *SD*, range 2–6; for neutrality: 3.44 ± 0.86 *SD*, range 2–5), and an overall rejection of 24 out of 214 (11.2%). The resulting number of stimulation sequences was equivalent across conditions and age groups (i.e., 6–12 for 3.5-month-olds and 6–14 for 7-month-olds).

Frequency-domain analysis

For each infant, the 32-s sequences were sorted according to each target expression condition and then averaged in the time-domain into a single epoch per condition. FFT was then applied to extract amplitude spectra for each electrode. To determine significant responses for both base and expression-change frequencies and their harmonics (i.e., integer multiples), amplitude at each channel was first normalized by dividing by the square root of the sum of squared amplitudes of all channels (McCarthy and Wood, 1985). Normalization was used to identify the main electrodes presenting a significant response by scaling differences between electrodes on the global power of the response across the scalp to determine whether the different expressions elicit distinguishable topographical patterns when

---

[1] http://nocions.github.io/letswave6

scalp-wide amplitude is equalized across them. Then, data were grand-averaged across infants for each age group and Z-scores were calculated. For the general visual response at 6 Hz and harmonics, we considered electrodes located over the middle occipital cortex (Oz, POz, O1/2) as in previous FPVS-EEG infant studies (e.g., de Heering and Rossion, 2015; Peykarjou et al., 2017; Barry-Anwar et al., 2018; Leleu et al., 2020; Rekow et al., 2020, 2021). For the expression-change response, since this study is the first to investigate the response to facial expressions in infants with the FPVS-EEG approach, we first explored all electrodes over the scalp to identify those that showed a response for the different expressions at each age ($Z > 1.96$, $p < 0.05$, two-tailed). This bottom-up procedure allowed us to determine the electrodes that significantly responded to facial expressions [for a similar approach, see Dzhelyova et al. (2017) and Leleu et al. (2019)]. These electrodes were then included in the analyses if they were consistent with the electrodes that had been previously reported in other EEG studies of the brain response to facial expression in the first year (see below). Harmonics were considered for further analysis until Z-scores for two consecutive harmonics over one channel were no longer significant. For each response, individual normalized amplitudes were summed across significant harmonics (Retter et al., 2021) and corresponding Z-scores were calculated for these summed amplitudes for each infant and for grand-averaged data in each age group to estimate the significance of the overall responses at both group and individual levels. For illustration purpose, SNR of each response was computed on grand-averaged data as the amplitude (before normalization) divided by the mean amplitude of the noise (same estimation as for noise-corrected amplitudes, see above).

To analyze the differences between the three facial expression discrimination conditions and between age groups, each response was also quantified as a single value expressed in microvolts by summing noise-corrected amplitudes (before normalization) for significant harmonics. Individual summed noise-corrected amplitudes were extracted for each electrode with a significant response in at least one expression condition for at least one age group (as determined in the previous analysis; see Section "Results"). Based on the criteria exposed previously, for the expression-change response, we first explored electrodes located over occipito-temporal (Oz, O1/2, P7/8) and central regions (Cz, FC1/2, CP1/2) according to previous EEG studies on facial expression discrimination in infants (e.g., Nelson and de Haan, 1996; de Haan and Nelson, 1999; Halit et al., 2004; Leppänen et al., 2007; Xie et al., 2019). We selected electrodes O1/2, P7/8, T7/8, CP1/2, FC1/2 that are close to these locations and that showed a significant response in the previous analyses (see Section "Results"). For lateral electrodes, the homologous channel in the other hemisphere was also considered in the statistical analysis. A repeated-measures ANOVA was performed on the normalized noise-corrected amplitudes in both age groups using Expression (disgust, happiness, neutrality) and Electrode (POz, Oz, O1, O2 for the general visual response and O1/2, P7/8, T7/8, CP1/2, FC1/2 for the expression-change response) as within-subject factors. The factor Hemisphere (right, left) was also used as a within-subject factor for the expression-change response only. In addition, with the aim to directly explore the effect of age, we performed a repeated-measures ANOVA using Age (3.5- and 7-month-old) as a between-subject factor and Expression and Electrode as within-subject factors. For each analysis, Mauchly's test for sphericity violation was performed and Greenhouse-Geisser correction was applied whenever sphericity was violated. Comparisons for significant effects were conducted using T-tests.

Results

Expression-change response

In both age groups, exposition to rapid changes of expression (from either a disgust, happy, or neutral face) gave rise to identifiable brain responses with a high SNR (between 1.3 and 1.5; i.e., indicating 30–50% of signal increase compared with surrounding noise level), and with different scalp topographies (**Figure 2**; for topographies of non-normalized noise-corrected amplitudes, see **Supplementary Figure 1**). For 3.5-month-old infants (**Figure 2A**), the expression-change response for the target expression of disgust was significant for the 1st harmonic (i.e., 1 Hz) over the medial occipital channels O1 ($Z = 2.97$, $p = 0.003$) and O2 ($Z = 4.34$, $p < 0.0001$). For neutrality, the response was significant over the right occipito-temporal channel P8 until the 2nd harmonic (i.e., 2 Hz), with a significant response over T8 ($Z = 2.93$, $p = 0.0034$) and P8 ($Z = 4.09$, $p < 0.001$) when amplitude is summed across the two first harmonics. For happiness, only one electrode showed a significant response: Oz ($Z = 2.05$, $p = 0.0404$). At 7 months, every expression condition led to a significant expression-change response at the 1st harmonic. It was recorded over O2 ($Z = 2.86$, $p = 0.0042$) for disgust, over FC2 ($Z = 2.79$, $p = 0.0053$) for happiness and over CP2 ($Z = 2.60$, $p = 0.0093$) for neutrality (**Figure 2B**).

Altogether, these results reveal that, at 3.5 months of age, the brain response to a change of expression from disgust and happiness was concentrated over posterior regions, markedly over medial occipital sites. The response was more lateral for neutrality, with a right-hemispheric occipito-temporal distribution. In contrast, at 7 months of age, the expression-change response is still detected over the medial occipital region for disgust while recorded over central parietal sites for neutrality and over central frontal sites for happiness.

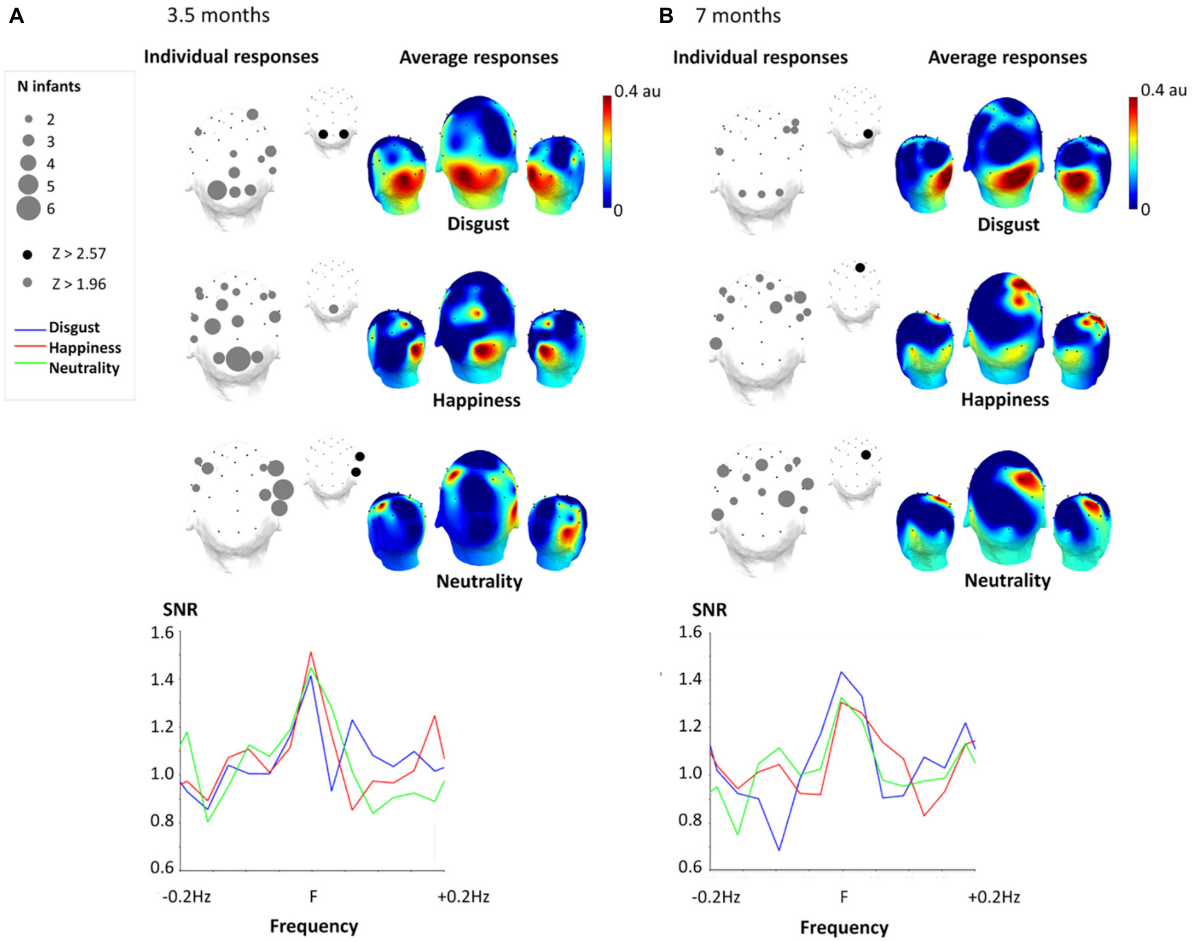


FIGURE 2
Expression-change response for the three facial expressions at 3.5 months **(A)** and 7 months **(B)**. Left part. Topographical representation (superior view) of significant individual responses. For each electrode, the number of infants with a significant response ($Z > 1.96$, $p < 0.05$) is represented by circle size when at least two individual responses were significant. The smaller topographical map also indicates significant electrodes at group level (gray: $Z > 1.96$, $p < 0.05$; black: $Z > 2.57$, $p < 0.01$). Right part. 3D topographical color maps (superior view) of the expression-change response (normalized noise-corrected amplitude in arbitrary units) for each expression condition. Bottom part. Signal-to-noise ratio (SNR) of the expression-change response and surrounding frequencies (±0.2 Hz, i.e., ±6 bins) averaged across significant electrodes for each expression condition.

Differences between expressions for each age group

To investigate the specific EEG response associated with the discrimination from the target expressions at each age, we first determined whether the distribution of the expression-change response over the scalp differentiates the different emotion categories for each age group. To do that, we analyzed normalized noise-corrected amplitudes (expressed in arbitrary units) over the different sites identified above and the corresponding site on the other hemisphere; namely, O1/2, P7/8, T7/8, CP1/2, and FC1/2. Medial electrodes were not included to investigate potential hemispheric differences.

In 3.5-month-olds, the analysis revealed a significant main effect of *Electrode* [$F_{(4,68)} = 4.281$, $\eta^2_p = 0.201$, $p = 0.0038$], with a larger expression-change response over O1/2 than over FC1/2, the other sites lying in between. More importantly for our purpose, the *Expression* × *Electrode* interaction was also significant [$F_{(8,136)} = 2.886$, $\eta^2_p = 0.145$, $p = 0.0053$; see **Figure 3**, left]. Complementary analyses using linear contrasts indicated that this interaction resulted from a different topography of the response to emotional expressions (i.e., disgust and happiness) compared to neutrality [*Expression* × *Electrode* interaction when disgust and happiness are pooled together: $F_{(4,68)} = 6.244$, $p = 0.0002$]. No significant difference emerged between disgust and happiness (*Expression* × *Electrode* interaction when neutrality is removed: $F < 1$). The difference between neutrality and the two emotional expressions resulted from a significantly lower contribution of O1/2 channels for the discrimination from neutrality [0.14 ± 0.02 (*SE*, standard error of the mean), arbitrary unit] than happiness (0.21 ± 0.02, $p = 0.039$), and disgust (0.22 ± 0.02, $p = 0.004$), together with a higher contribution of P7/8 electrodes for neutrality (0.20 ± 0.02) than disgust (0.13 ± 0.02, $p = 0.029$). No other main effect or interaction was significant for this age group.

In 7-month-olds, the only significant effect was the *Expression* × *Electrode* interaction [$F_{(8,136)} = 2.210$, $\eta^2{}_p = 0.115$, $p = 0.0303$; see **Figure 3**, right], and a trend for a larger response in the right hemisphere was noted [main effect of *Hemisphere*: $F_{(1,17)} = 3.792$, $p = 0.0682$]. Complementary analyses using linear contrasts indicated that by this age, and contrary to 3.5-month-old infants, the interaction was not explained by differences between emotional expressions and neutrality (*Expression* × *Electrode* interaction when disgust and happiness data are pooled: $F < 1$). Rather, the topography of the expression-change response was different between disgust and happiness [*Expression* × *Electrode* interaction when neutrality is removed: $F_{(4,68)} = 3.42$, $p = 0.0122$]. The interaction was mainly driven by a lower contribution of O1/2 to the response to happiness (0.11 ± 0.02) compared to both disgust (0.21 ± 0.03, $p = 0.011$) and neutrality (0.18 ± 0.02, $p = 0.028$). As for the 3.5-month-olds, P7/8 channels contributed more to the discrimination from neutrality (0.21 ± 0.02) than of disgust (0.15 ± 0.02, $p = 0.019$).

To directly assess the effect of age, we then performed a second ANOVA including *Age* (3.5- vs. 7-month-olds) as a between-subject factor. This analysis revealed a significant main effect of *Electrode* [$F_{(4,136)} = 5.41$, $\eta 2p = 0.134$, $p = 0.0004$] and a significant *Expression* × *Electrode* interaction [$F_{(8,272)} = 3.31$, $\eta^2{}_p = 0.089$, $p = 0.0013$], but only a trend for the *Age* × *Expression* × *Electrode* interaction [$F_{(8,272)} = 1.71$, $\eta^2{}_p = 0.048$, $p = 0.096$].

Individual expression-change responses

To assess the robustness of the expression-change EEG response at the individual level, we explored the responses across the whole scalp for each individual infant (see **Supplementary Table**).

Overall, individual responses confirm group-level observations with the discrimination of facial expression from disgust mainly eliciting medial occipital expression-change responses at both ages despite a broad distribution over the scalp (**Figure 2**). In contrast, while expression changes from happiness and neutrality lead to posterior brain responses for the 3.5-month-olds (i.e., over the medial occipital and right lateral channels respectively), they give rise to more anterior activity for happiness in 7-month-old infants. In sum, the expression-change response is reliably found at the individual level ensuring that the group response is not accounted for by a small subset of infants.

General visual response

As expected, the 6-Hz stimulation elicited a clear brain response at the same frequency and its harmonics (e.g., 12, 18 Hz) over the medial occipital cortex, reflecting the general visual processing of the rapidly presented individual faces changing in size (**Figure 4**). Exploration of the significant electrodes for each facial expression condition and each age group revealed a significant response at 6 Hz over the four occipital channels (Oz, POz, O1, O2) for every expression condition and both age groups (all $Zs > 3.06$, $ps < 0.0022$). Following harmonics were significant over at least one of these electrodes until the sixth harmonic (i.e., 36 Hz) for every expression and age group (all $Zs > 2.59$, $ps < 0.0096$). When summed across harmonics, the general visual response was still significant over the 4 medial occipital channels (all $Zs > 7.46$, $ps < 0.0001$). SNR was very high in 3.5-month-olds (SNR ≈ 6, i.e., signal six times larger than noise) and lower but still high in 7-month-olds (SNR ≈ 3, i.e., signal three times larger than noise) (**Figure 4**). Importantly for our purpose, the repeated-measures ANOVAs run separately at each age revealed that no effect involving the factor *Expression* was significant.

The second ANOVA revealed a trend for the effect of *Age* [$F_{(1,34)} = 3.69$, $\eta^2{}_p = 0.098$, $p = 0.063$], with a larger response in the 3.5-month-olds [5.22 ± 0.95 μV (SE)] than in the 7-month-olds (3.21 ± 0.44 μV). In addition, we found a significant main effect of *Electrode* [$F_{(1.93,65.64)} = 15.83$, $\eta^2{}_p = 0.64$, $p < 0.0001$] qualified by a significant *Electrode* × *Age* interaction [$F_{(1.93,65.64)} = 5.39$, $\eta^2{}_p = 0.64$, $p = 0.007$]. This latter interaction was characterized by a significant difference between age groups (3.5 vs. 7-month-olds) for electrodes O2 (6.32 ± 1.19 μV vs. 3.42 ± 0.43 μV, $p = 0.03$), Oz (6.91 ± 1.31 μV, vs. 3.83 ± 0.65 μV, $p = 0.04$), together with a trend for O1 (4.84 ± 0.94 μV, vs. 2.90 ± 0.46 μV, $p = 0.07$) and no difference for POz (2.84 ± 0.54 μV, vs. 2.69 ± 0.53 μV, $p = 0.85$).

Discussion

Using FPVS-EEG, the present study isolated direct brain markers of rapid facial expression discrimination in 3.5- and 7-month-old infants, by investigating the specific neural responses to brief changes of facial expression within rapid streams of neutral, disgust, or happy faces. A significant expression-change response was observed over posterior scalp regions for the discrimination from disgust and from happiness in 3.5-month-olds, with a similar response for the discrimination from disgust in 7-month-olds. For the discrimination from neutrality, the expression-change response was recorded over more anterior, occipito-temporal and parietal regions, in both age groups. Finally, a response to a change of expression from happiness emerged over central frontal scalp regions at 7 months of age. These results show that both 3.5- and 7-month-olds

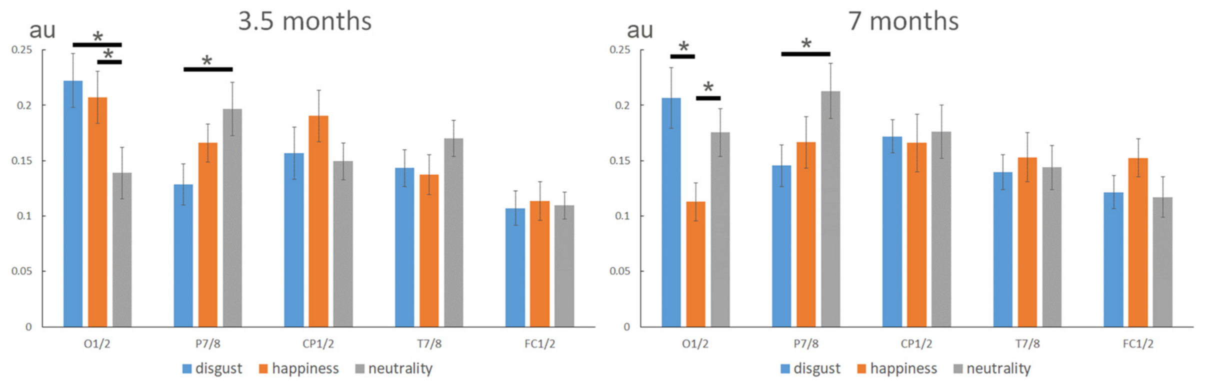


FIGURE 3
Expression-change response (normalized noise-corrected amplitudes, in arbitrary units) according to expressions over different sites (O1/2, P7/8, CP1/2, T7/8, and FC1/2) at 3.5 and 7 months. In 3.5-month-olds, the analysis revealed a similar topography for the two emotional expressions (i.e., disgust and happiness) that differed from neutrality with a lower contribution of O1/2 and a higher contribution of P7/8 for neutrality. In 7-month-olds, the topography of the expression-change response discriminated between disgust, neutrality, and happiness, with a lower contribution of O1/2 to the response to happiness compared to both disgust and neutrality. As for the 3.5-month-olds, P7/8 channels contributed more to the discrimination of neutrality than disgust. Note also the tendency for a higher contribution of FC1/2 to happiness, that was driven by a significant response over FC2 for happiness (see previous analyses). "*" indicated significant differences ($p < 0.05$).

discriminate several facial expressions from each target expression. They also suggest that distinct brain regions/neural networks could be involved in facial expression discrimination depending on the age of the infants and the expression.

For the discrimination from neutrality, by opposing a neutral/unexpressive face (i.e., without facial actions) to all potential emotional facial expressions displayed randomly, we isolated a brain response to the occurrence of facial actions within the face at both ages. Thus, this response could be a marker of the detection of expressive facial configurations. In infants, ERP studies revealed several cortical sources within occipital brain areas (notably the IOG), the latFG, the posterior temporal cortex (including the pSTS), the PCC/precuneus or the middle frontal cortex (including the medial prefrontal cortex: mPFC) (Guy et al., 2016; Xie et al., 2019), suggesting that the neural network delineated in adults is already partly functional during the first year of life (Leppänen and Nelson, 2009; Leppänen, 2011). Given that the expression-change response for neutrality was recorded over scalp regions analogous to those observed in adults [i.e., at right occipito-temporal sites; e.g., Poncet et al. (2019)], our findings might suggest that this network is already functional at 3.5 months of age, and subtends the discrimination of neutral vs. expressive faces. However, considering the lack of evidence for more finely tuned processes that distinguish the different expression categories at this age (Campos and Stenberg, 1981; Oster, 1981; Walker-Andrews, 1997; Leppänen and Nelson, 2009; Quinn et al., 2011), the functionality of this network could be limited to segregate any expressive face from a neutral face irrespective of the emotional content. It worth noting that these interpretations remain tentative as the analogy between topographies in infants and adults must be made with caution, and scalp EEG is limited by its coarse spatial resolution. In addition, we cannot ascertain that the response is specific to *emotional* configurations. Further research is needed to dismiss the possibility that opposing neutral faces to facial expressions without emotional content, such as tongue protrusion or speech-related facial movements (e.g., say "O"), would elicit a similar neural response.

An expression-change response from disgust and happiness was also isolated at both ages. This response was recorded over medial occipital scalp regions in 3.5-month-olds, and over medial occipital and fronto-central regions, for disgust and happiness respectively, in 7-month-olds, both regions responding more than to the expression-change from facial neutrality. This observation suggests that the infant brain discriminates several expressions from every target expression at both ages. Indeed, the expression-change response emerges only if the brain detects visual cues that both (1) occur reliably in the target expression, and (2) do not systematically occur in the other expressions. These cues can be either local properties of a single facial action (e.g., a smiling mouth or a wrinkled nose), or more complex and integrated patterns (i.e., the co-occurrence of facial actions over the whole face). In particular here, for the brain response to a brief change of expression from neutrality, the changes were associated with facial actions turning from unexpressive (i.e., neutral face) to expressive (i.e., anger, disgust, happiness, fear, or sadness) over the whole face. In contrast, the brain responses to disgust and happiness were not driven by all facial features at each change since facial actions can be shared across expressions (e.g., eyebrow lowering for anger, disgust, fear, and sadness). One may therefore suggest that discriminating from neutrality relies more on brain structures that integrate the configuration of facial actions over the whole face, whereas discriminating from both disgust and happiness

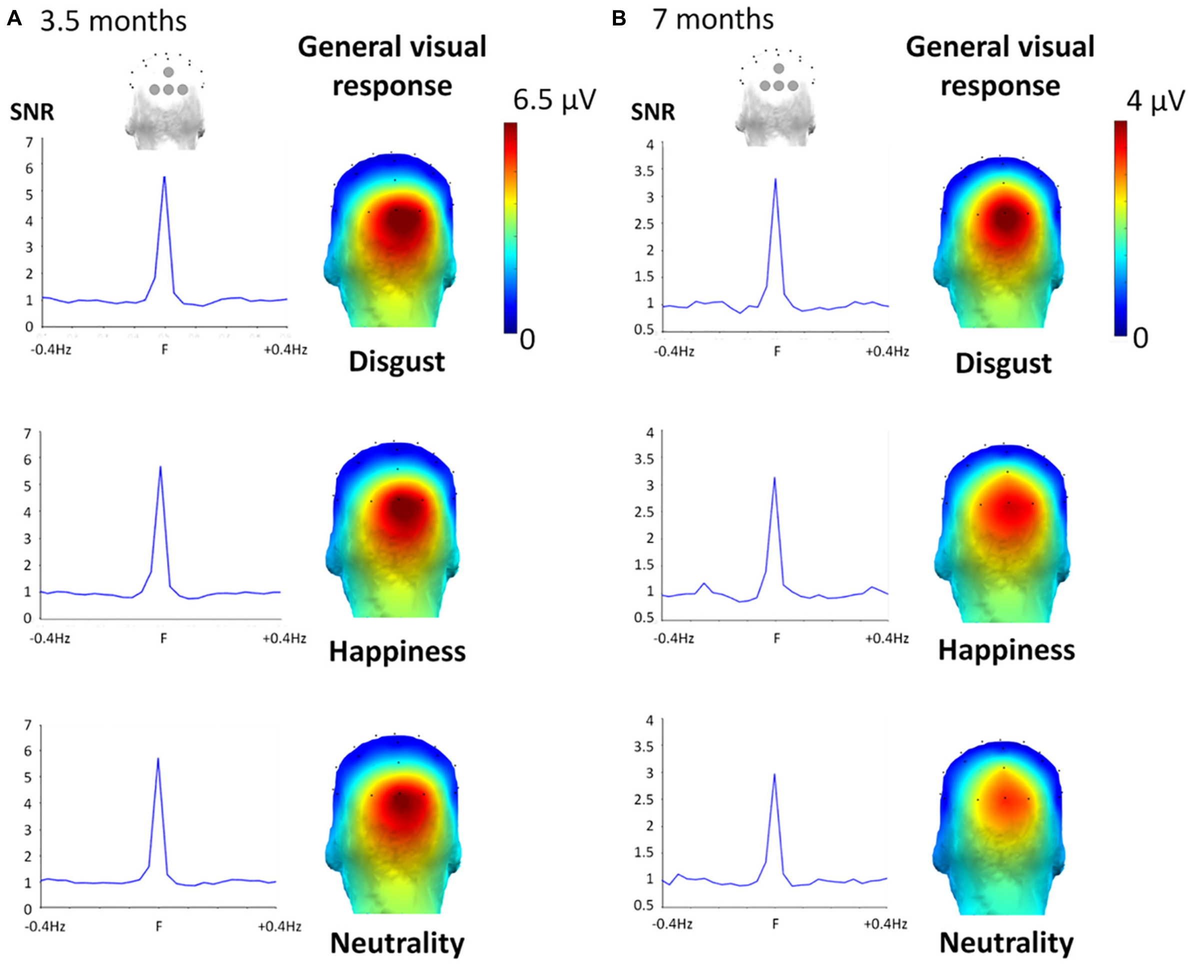


FIGURE 4
General visual response for the three target facial expressions at 3.5 months **(A)** and 7 months **(B)**. Left part. SNR was calculated on the summed harmonics values averaged between the four occipital electrodes, considering Z scores of at least one of the four occipital electrodes was significant (across harmonics *Zs* > 2.60, *p* < 0.006), showing a high SNR for each expression: around six for every expression-condition in the 3.5-month-olds, and around three in the 7-month-olds. Right part. Topographical 3D map (back view) shows a clear medial occipital cortex response of the noise-corrected amplitude (in µV) values summed until the 6th harmonic of the 6 Hz response for the three conditions.

is subtended by lower-level regions that process more local information. In sum, we propose that the medial occipital response to an expression change from disgust and happiness in 3.5-month-olds, and from disgust in 7-month-olds, could be related to the discrimination of local facial actions that differentiate several expressions from the target expression (e.g., the smiling mouth in happy faces, the wrinkled nose in disgust faces). By contrast, the more lateral response observed for an expression-change from neutrality could be elicited by the co-occurrence of several expressive features over the whole face, regardless of the (emotional) nature of the configuration of facial actions.

Contrary to the discrimination from disgust and neutrality, the brain response to a change of expression from happiness differed between 3.5 and 7 months. The response mainly appeared over occipital/posterior sites and was not different from the response to the discrimination from disgust in 3.5-month-olds, whereas the occipital response contributed less to the expression-change from happiness than other categories in 7-month-olds. Rather, a change of expression from happiness elicited a response over the central frontal region at this age. This topographic shift might reflect a specialization of the brain response, with the integration of affective and/or social meaning, as demonstrated by recent studies (e.g., Palama et al., 2018). The scalp topography observed here is consistent with those reported in ERP studies (notably with the topography of the "Negative Central": Nc component) in situations supposed to involve different levels of interest for infants [e.g., new vs. familiar objects, Reynolds and Richards (2005); faces vs. toys, Guy et al. (2016)], or for different expressions (e.g., Xie et al., 2019). These findings were thought of as reflecting stimulus salience (Nelson and de Haan, 1996) or attentional processes (e.g., Richards, 2003; Reynolds and Richards, 2005; Guy et al., 2016; Xie and Richards, 2016; Xie et al., 2019). For example, Xie et al. (2019) proposed that a central frontal response emerges for expressions that engage attention allocation and deeper processing. Similar central frontal activities were reported in response to happiness using fNIRS (Minagawa-Kawai et al., 2008; Fox et al., 2013). These findings were explained by the acquisition of socio-cognitive abilities (Grossmann, 2013).

An alternative (although non-exclusive) hypothesis is that the brain response to a rapid change of expression from happiness indexes the discrimination of positive vs. negative emotions at 7 months. In our design, all emotions but happiness were negative (4 out of 6) or unexpressive (1 out of 6). Thus, the brain response isolated by this contrast may also capture the processing of emotional valence, and not solely the processing of visually distinct expressions. However, irrespective of the nature of the response, it indicates that discriminating several expressions from happiness elicits a different brain response in 7-month-olds, possibly reflecting more attention or the recruitment of specific brain mechanisms implicated in affective, cognitive and/or social processing. Future studies should examine the hypothetical relationship between central frontal activities and emotional meaning attribution, for instance by testing the influence of contextual information, such as the emotional environment provided by the mother (de Haan et al., 2004; Jessen, 2020) or the multisensory context provided by auditory (Flom and Bahrick, 2007) or odor cues (Godard et al., 2016).

The brain response to an expression change from disgust was mainly found over occipital/posterior sites and did not evolve with age. It indicates that the brain has detected some visual cues that reliably occur in a disgust face and are absent in the other expressions (i.e., anger, happiness, fear, sadness, and neutrality). Considering the facial actions identified for these different expressions, the main candidates for disgust-specificity are nose wrinkling and lips parting together with upper lip rising; two actions more associated with disgust than with any other facial emotions (Ekman et al., 1978). The stability of the response between 3.5 and 7 months suggests that this expression is similarly processed during this period, contrary to happiness. This suggestion is in line with behavioral studies, which indicate that the emotional meaning of disgust faces is not integrated before 12 months (Moses et al., 2001; Hertenstein and Campos, 2004) or even later (Widen and Russell, 2010, 2013), as its understanding would imply a higher cognitive development (Rozin and Fallon, 1987; Widen and Russell, 2013). It may also be possible, however, that the contrast performed here did not allow to isolate the specific response to the emotional meaning of this facial expression. As mentioned earlier, the contrasted expressions were mainly negative (four expressions: disgust, anger, fear, and sadness), as opposed to only one positive expression (happiness) and neutrality. Thus, we cannot exclude that a face expressing disgust already acquired the status of a negative signal at 7 months, but may not be dissociated from other negative facial signals, except from visual characteristics (as suggested by the occipital expression-change response). This hypothesis should be further investigated by testing other negative emotions, such as fear or anger, already known to trigger infants' attention in relation to meaning attribution (Leppänen and Nelson, 2009; Xie et al., 2019).

Finally, as a limitation, it is worth noting that the different patterns of brain activity observed for each discrimination at each age were only partially supported by a trend for a 2-way interaction in the global analysis that included age as a factor. At least two main reasons can explain this finding. First, contrary to the response to the discrimination from happiness, the responses to the discrimination from disgust and neutrality appear similar at each age. The analysis including the three conditions was thus probably limited in its ability to evidence an effect of age that is entirely driven by only one condition. In addition, several confounding factors (e.g., brain maturation, skull thickness) may also lead to differences in the amplitude and topography of the brain responses between the two age groups, and partially hinder our ability to identify the effect of age on the response to a discrimination from happiness. Future studies should thus further investigate the brain responses to different facial expressions at different ages using complementary approaches.

Conclusion

Using FPVS-EEG, we characterized brain responses indicating that several basic facial expressions are discriminated from the expressions of disgust, happiness and neutrality at 3.5 and 7 months of age. The response to a change of expression from disgust was mainly located over medial occipital sites at both ages, likely reflecting visual discrimination based on local facial features. The distinct response noted at both ages for the expression-change from neutrality further suggests that the discrimination from this expression relies on more global cues (i.e., integration of facial actions over the whole face). Finally, for the discrimination from happiness, the expression-change response was recorded over the occipital region at 3.5 months, while we rather found a significant response over central frontal scalp regions at 7 months, potentially reflecting a critical developmental change in the processing of the emotional content of smiling faces.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Comité de Protection des Personnes Sud-Est III–2016-A02056-45. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

FP, AL, DR, and J-YB contributed to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript. FD contributed to the implementation of the research and to the writing of the manuscript. MD contributed to the design of the research and to the writing of the manuscript. BS, KD, and LF contributed to the writing of the manuscript. BR contributed to the design of the research, to the analysis of the results, and to the writing of the manuscript. DR contributed to the design and implementation of the research and to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was received support from the Conseil Régional de Bourgogne Franche-Comté, the European Fund for Regional Economic Development, the French "Investissements d'Avenir" program, project ISITE-BFC (contract ANR-15-IDEX-0003), and the French National Research Agency (contract ANR-19-CE28-0009). MD was supported by the Fonds National de la Recherche Scientifique (FNRS).

Acknowledgments

We thank Sylviane Martin for help in recruiting parents and their infants. We are also grateful to Vincent Gigot for assistance in EEG preprocessing.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.901013/full#supplementary-material

References

Addabbo, M., Longhi, E., Marchis, I., Tagliabue, P., and Turati, C. (2018). Dynamic facial expressions of emotions are discriminated at birth. *PLos One* 13:e0193868. doi: 10.1371/journal.pone.0193868

Adrian, E. D., and Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *J. Physiol.* 81, 440–471. doi: 10.1113/jphysiol.1934.sp003147

Barrera, M. E., and Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Dev.* 52, 203–206. doi: 10.2307/1129231

Barry-Anwar, R., Hadley, H., Conte, S., Keil, A., and Scott, L. S. (2018). The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia* 108, 25–31. doi: 10.1016/j.neuropsychologia.2017.11.019

Bayet, L., Perdue, K. L., Behrendt, H. F., Richards, J. E., Westerlund, A., Cataldo, J. K., et al. (2021). Neural responses to happy, fearful and angry faces of varying identities in 5-and 7-month-old infants. *Dev. Cogn. Neurosci.* 47:100882. doi: 10.1016/j.dcn.2020.100882

Bornstein, M. H., and Arterberry, M. E. (2003). Recognition, discrimination and categorization of smiling by 5-month-old infants. *Dev. Sci.* 6, 585–599. doi: 10.1111/1467-7687.00314

Bornstein, M. H., Arterberry, M. E., Mash, C., and Manian, N. (2011). Discrimination of facial expression by 5-month-old infants of nondepressed and clinically depressed mothers. *Infant Behav. Dev.* 34, 100–106. doi: 10.1016/j.infbeh.2010.10.002

Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., and Stenberg, C. (1983). "Socioemotional development," in *Handbook of Child Psychology, Volume 2: Infancy and Developmental Psychobiology*, eds M. M. Haith, J. W. Campos, and P. H. Mussen (New York, NY: Wiley), 783–915.

Campos, J. J., and Stenberg, C. (1981). "Perception, appraisal, and emotion: the onset of social referencing," in *Infant Social Cognition: Empirical and Theoretical Considerations*, eds M. E. Lamb and L. R. Sherrod (Hillsdale, NJ: Erlbaum), 273–314.

Campos, J. J., Thein, S., and Owen, D. (2003). A Darwinian legacy to understanding human infancy: emotional expressions as behavior regulators. *Ann. N.Y. Acad. Sci.* 1000, 110–134. doi: 10.1196/annals.1280.040

Caron, R. F., Caron, A. J., and Myers, R. S. (1982). Abstraction of invariant face expressions in infancy. *Child Dev.* 53, 1008–1015. doi: 10.2307/1129141

de Haan, M. (2007). "Visual attention and recognition memory in infancy," in *Infant EEG and Event-Related Potentials*, ed. M. de Haan (Hove: Psychology Press), 101–143.

de Haan, M., Belsky, J., Reid, V., Volein, A., and Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *J. Child Psychol. Psychiatry* 45, 1209–1218. doi: 10.1111/j.1469-7610.2004.00320.x

de Haan, M., and Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Dev. Psychol.* 35, 1113–1121. doi: 10.1037/0012-1649.35.4.1113

de Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife* 4:e06564. doi: 10.7554/eLife.06564

Dzhelyova, M., Jacques, C., and Rossion, B. (2017). At a single glance: fast periodic visual stimulation uncovers the spatio-temporal dynamics of brief facial expression changes in the human brain. *Cerebr. Cortex* 27, 4106–4123. doi: 10.1093/cercor/bhw223

Ekman, P., Friesen, W. V., and Hager, J. C. (1978). *Facial Action Coding System Investigator's Guide*. Salt Lake City, UT: Research Nexus.

Farroni, T., Menon, E., Rigato, S., and Johnson, M. H. (2007). The perception of facial expressions in newborns. *Eur. J. Dev. Psychol.* 4, 2–13. doi: 10.1080/17405620601046832

Feinman, S., and Lewis, M. (1983). Social referencing at ten months: a second-order effect on infants' responses to strangers. *Child Dev.* 54, 878–887. doi: 10.2307/1129892

Field, T. M., Woodson, R. W., Greenberg, R., and Cohen, C. (1982). Discrimination and imitation of facial expressions by neonates. *Science* 218, 179–181. doi: 10.1126/science.7123230

Flom, R., and Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: the role of intersensory redundancy. *Dev. Psychol.* 43, 238–252. doi: 10.1037/0012-1649.43.1.238

Fox, S. E., Wagner, J., Shrock, C. L., Flusberg, H. T., and Nelson, C. A. (2013). Neural processing of facial identity and emotion in infants at high-risk for autism spectrum disorders. *Front. Hum. Neurosci.* 7:89. doi: 10.3389/fnhum.2013.00089

Fransson, P., Metsäranta, M., Blennow, M., Åden, U., Lagercrantz, H., and Vanhatalo, S. (2013). Early development of spatial patterns of power-law frequency scaling in fMRI resting-state and EEG data in the newborn brain. *Cerebr. Cortex* 23, 638–646. doi: 10.1093/cercor/bhs047

Fujioka, T., Mourad, N., He, C., and Trainor, L. J. (2011). Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clin. Neurophysiol.* 122, 43–51. doi: 10.1016/j.clinph.2010.04.036

Godard, O., Baudouin, J.-Y., Schaal, B., and Durand, K. (2016). Affective matching of odors and facial expressions in infants: shifting patterns between 3 and 7 months. *Dev. Sci.* 19, 155–163. doi: 10.1111/desc.12292

Grossmann, T. (2013). The role of medial prefrontal cortex in early social cognition. *Front. Hum. Neurosci.* 7:340. doi: 10.3389/fnhum.2013.00340

Guy, M. W., Zieber, N., and Richards, J. E. (2016). The cortical development of specialized face processing in infancy. *Child Dev.* 87, 1581–1600. doi: 10.1111/cdev.12543

Halit, H., Csibra, G., Volein, Á., and Johnson, M. H. (2004). Face-sensitive cortical processing in early infancy. *J. Child Psychol. Psychiatry* 45, 1228–1234. doi: 10.1111/j.1469-7610.2004.00321.x

Haviland, J. M., and Lelwica, M. (1987). The induced affect response: 10-week-old infants' responses to three emotion expressions. *Dev. Psychol.* 23, 97–104. doi: 10.1037/0012-1649.23.1.97

Hertenstein, M. J., and Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Dev.* 75, 595–613. doi: 10.1111/j.1467-8624.2004.00695.x

Jessen, S. (2020). Maternal odor reduces the neural response to fearful faces in human infants. *Dev. Cogn. Neurosci.* 45:100858. doi: 10.1016/j.dcn.2020.100858

Kaiser, J., Crespo-Llado, M. M., Turati, C., and Geangu, E. (2017). The development of spontaneous facial responses to others' emotions in infancy: an EMG study. *Sci. Rep.* 7:17500. doi: 10.1038/s41598-017-17556-y

Kaitz, M., Meschulach-Sarfaty, O., Auerbach, J., and Eidelman, A. (1988). A reexamination of newborns' ability to imitate facial expressions. *Dev. Psychol.* 24, 3–7. doi: 10.1037/0012-1649.24.1.3

Klinnert, M. D., Emde, R. N., Butterfield, P., and Campos, J. J. (1986). Social referencing: the infant's use of emotional signals from a friendly adult with mother present. *Dev. Psychol.* 22, 427–432. doi: 10.1037/0012-1649.22.4.427

Kotsoni, E., de Haan, M., and Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception* 30, 1115–1125. doi: 10.1068/p3155

LaBarbera, J. D., Izard, C. E., Vietze, P., and Parisi, S. A. (1976). Four-and six-month-old infants' visual responses to joy, anger, and neutral expressions. *Child Dev.* 47, 535–538. doi: 10.2307/1128816

Legerstee, M. (1997). Contingency effects of people and objects on subsequent cognitive functioning in 3-month-old infants. *Soc. Dev.* 6, 307–321. doi: 10.1111/j.1467-9507.1997.tb00108.x

Leleu, A., Dzhelyova, M., Rossion, B., Brochard, R., Durand, K., Schaal, B., et al. (2018). Tuning functions for automatic detection of brief changes of facial expression in the human brain. *Neuroimage* 179, 235–251. doi: 10.1016/j.neuroimage.2018.06.048

Leleu, A., Favre, E., Yailian, A., Fumat, H., Klamm, J., Amado, I., et al. (2019). An implicit and reliable neural measure quantifying impaired visual coding of facial expression: Evidence from the 22q11.2 deletion syndrome. *Transl. Psychiatry* 9:67. doi: 10.1038/s41398-019-0411-z

Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., et al. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Dev. Sci.* 23:e12877. doi: 10.1111/desc.12877

Leppänen, J. M. (2011). Neural and developmental bases of the ability to recognize social signals of emotions. *Emot. Rev.* 3, 179–188. doi: 10.1177/1754073910387942

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., and Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Dev.* 78, 232–245. doi: 10.1111/j.1467-8624.2007.00994.x

Leppänen, J. M., and Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nat. Rev. Neurosci.* 10, 37–47. doi: 10.1038/nrn2554

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF, CD ROM from Department of Clinical Neuroscience, Psychology Section*. Solna: Karolinska Institutet.

Maria, A., Shekhar, S., Nissilä, I., Kotilahti, K., Huotilainen, M., Karlsson, L., et al. (2018). Emotional processing in the first 2 years of life: a review of near-infrared spectroscopy studies. *J. Neuroimaging* 28, 441–454. doi: 10.1111/jon.12529

Matt, S., Dzhelyova, M., Maillard, L., Lighezzolo-Alnot, J., Rossion, B., and Caharel, S. (2021). The rapid and automatic categorization of facial expression changes in highly variable natural images. *Cortex* 144, 168–184. doi: 10.1016/j.cortex.2021.08.005

McCarthy, G., and Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr. Clin. Neurophysiol.* 62, 203–208. doi: 10.1016/0168-5597(85)90015-2

Millar, W. S. (1976). Social reinforcement of a manipulative response in six- and nine-month-old infants. *J. Child Psychol. Psychiatry* 17, 205–212. doi: 10.1111/j.1469-7610.1976.tb00393.x

Minagawa-Kawai, Y., Matsuoka, S., Dan, I., Naoi, N., Nakamura, K., and Kojima, S. (2008). Prefrontal activation associated with social attachment: facial-emotion recognition in mothers and infants. *Cerebr. Cortex* 19, 284–292. doi: 10.1093/cercor/bhn081

Moses, L. J., Baldwin, D. A., Rosicky, J. G., and Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Dev.* 72, 718–735. doi: 10.1111/1467-8624.00311

Mourad, N., Reilly, J. P., Bruin, H. D., Hasey, G., and MacCrimmon, D. (2007). "A simple and fast algorithm for automatic suppression of high-amplitude artifacts in EEG data," in *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing*. (Piscataway, NJ: IEEE).

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: mechanisms of development. *Child Dev.* 58, 889–909. doi: 10.2307/1130530

Nelson, C. A. (1994). "Neural correlates of recognition memory in the first postnatal year of life," in *Human Behavior and the Developing Brain*, eds G. Dawson and K. Fischer (New York, NY: Guilford Press), 269–313.

Nelson, C. A., and de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotions. *Dev. Psychobiol.* 29, 577–595.

Nelson, C. A., and Dolgin, K. (1985). The generalized discrimination of facial expressions by 7-month-old infants. *Child Dev.* 56, 58–61. doi: 10.2307/1130173

Nelson, C. A., Morse, P. A., and Leavitt, L. A. (1979). Recognition of facial expressions by 7-month-old infants. *Child Dev.* 50, 1239–1242. doi: 10.2307/1129358

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.6.4

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., et al. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Curr. Biol.* 26, 1334–1338. doi: 10.1016/j.cub.2016.03.047

Oster, H. (1981). ""Recognition" of emotional expression in infancy," in *Infant Social Cognition: Empirical and Theoretical Considerations*, eds M. E. Lamb and L. R. Sherrod (Hillsdale, NJ: Erlbaum), 85–125.

Palama, A., Malsert, J., and Gentaz, E. (2018). Are 6-month-old human infants able to transfer emotional information (happy or angry) from voices to faces? An eye-tracking study. *PLoS One* 13:e0194579. doi: 10.1371/journal.pone.0194579

Peltola, M. J., Hietanen, J. K., Forssman, L., and Leppänen, J. M. (2013). The emergence and stability of the attentional bias to fearful faces in infancy. *Infancy* 18, 905–926. doi: 10.1111/infa.12013

Peltola, M. J., Leppänen, J. M., Palokangas, T., and Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Dev. Sci.* 11, 60–68. doi: 10.1111/j.1467-7687.2007.00659.x

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., and Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion* 9, 560–565. doi: 10.1037/a0015806

Peykarjou, S., Hoehl, S., Pauen, S., and Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* 7:12526. doi: 10.1038/s41598-017-12760-2

Poncet, F., Baudouin, J.-Y., Dzhelyova, M. P., Rossion, B., and Leleu, A. (2019). Rapid and automatic discrimination between facial expressions in the human brain. *Neuropsychologia* 129, 47–55. doi: 10.1016/j.neuropsychologia.2019.03.006

Prunty, J. E., Keemink, J. R., and Kelly, D. J. (2021). Infants scan static and dynamic facial expressions differently. *Infancy* 26, 831–856. doi: 10.1111/infa.12426

Quadrelli, E., Conte, S., Macchi Cassia, V., and Turati, C. (2019). Emotion in motion: facial dynamics affect infants' neural processing of emotions. *Dev. Psychobiol.* 61, 843–858. doi: 10.1002/dev.21860

Quadrelli, E., Roberti, E., Polver, S., Bulf, H., and Turati, C. (2021). Sensorimotor activity and network connectivity to dynamic and static emotional faces in 7-month-old infants. *Brain Sci.* 11:1396. doi: 10.3390/brainsci11111396

Quinn, P. C., Anzures, G., Izard, C. E., Lee, K., Pascalis, O., Slater, A. M., et al. (2011). Looking across domains to understand infant representation of emotion. *Emot. Rev.* 3, 197–206. doi: 10.1177/1754073910387941

Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. New York, NY: Elsevier.

Rekow, D., Baudouin, J.-Y., Poncet, F., Damon, F., Durand, K., Schaal, B., et al. (2021). Odor-driven face-like categorization in the human infant brain. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2014979118. doi: 10.1073/pnas.2014979118

Rekow, D., Leleu, A., Poncet, F., Damon, F., Rossion, B., Durand, K., et al. (2020). Categorization of objects and faces in the infant brain and its sensitivity to maternal odor: further evidence for the role of intersensory congruency in perceptual development. *Cogn. Dev.* 55:100930. doi: 10.1016/j.cogdev.2020.100930

Retter, T. L., Rossion, B., and Schiltz, C. (2021). Harmonic amplitude summation for frequency-tagging analysis. *J. Cogn. Neurosci.* 33, 2372–2393. doi: 10.1162/jocn_a_01763

Reynolds, G. D., and Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: an event-related potential and cortical source localization study. *Dev. Psychol.* 41, 598–615. doi: 10.1037/0012-1649.41.4.598

Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: an ERP study. *Dev. Sci.* 6, 312–328. doi: 10.1111/1467-7687.00287

Rochat, P., and Striano, T. (1999). "Social–cognitive development in the first year," in *Early Social Cognition: Understanding Others in the First Months of Life*, ed. P. Rochat (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 3–34.

Rossion, B., Retter, T. L., and Liu-Shuang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *Eur. J. Neurosci.* 52, 4283–4334. doi: 10.1111/ejn.14865

Rotem-Kohavi, N., Oberlander, T. F., and Virji-Babul, N. (2017). Infants and adults have similar regional functional brain organization for the perception of emotions. *Neurosci. Lett.* 650, 118–125. doi: 10.1016/j.neulet.2017.04.031

Rozin, P., and Fallon, A. E. (1987). A perspective on disgust. *Psychol. Rev.* 94, 23–41. doi: 10.1037/0033-295X.94.1.23

Ruba, A. L., Meltzoff, A. N., and Repacholi, B. M. (2019). How do you feel? Preverbal infants match negative emotions to events. *Dev. Psychol.* 55, 1138–1149. doi: 10.1037/dev0000711

Serrano, J. J. M., Iglesias, J., and Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4- to 6-month-old infants. *Dev. Psychobiol.* 25, 411–425. doi: 10.1002/dev.420250603

Soussignan, R., Dollion, N., Schaal, B., Durand, K., Reissland, N., and Baudouin, J.-Y. (2018). Mimicking emotions: how 3-12 month-old infants use the facial expressions and eyes of a model. *Cogn. Emot.* 32, 827–842. doi: 10.1080/02699931.2017.1359015

Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194. doi: 10.1016/j.neuropsychologia.2006.06.003

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychol. Bull.* 121, 437–456. doi: 10.1037/0033-2909.121.3.437

Walker-Andrews, A. S. (2005). "Perceiving social affordances: the development of emotional understanding," in *The Development of Social Cognition and Communication*, eds B. D. Homer and C. S. Tamis-LeMonda (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 93–116.

Widen, S. C., and Russell, J. A. (2010). The "disgust face" conveys anger to children. *Emotion* 10, 455–466. doi: 10.1037/a0019151

Widen, S. C., and Russell, J. A. (2013). Children's recognition of disgust in others. *Psychol. Bull.* 139, 271–299. doi: 10.1037/a0031640

Xie, W., McCormick, S. A., Westerlund, A., Bowman, L. C., and Nelson, C. A. (2019). Neural correlates of facial emotion processing in infancy. *Dev. Sci.* 22:e12758. doi: 10.1111/desc.12758

Xie, W., and Richards, J. E. (2016). Effects of interstimulus intervals on behavioral, heart rate, and event-related potential indices of infant engagement and sustained attention. *Psychophysiology* 53, 1128–1142. doi: 10.1111/psyp.12670

Young-Browne, G., Rosenfeld, H. M., and Horowitz, F. D. (1977). Infant discrimination of facial expressions. *Child Dev.* 48, 555–562. doi: 10.2307/1128653

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fpsyg.2022.932952

OPEN ACCESS

EDITED BY
Aliette Lochy,
University of Luxembourg,
Luxembourg

REVIEWED BY
Fang Wang,
Stanford University,
United States
Anahita Basirat,
Université de Lille,
France

*CORRESPONDENCE
Valentina N. Pescuma
pescumav@hu-berlin.de

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 30 April 2022
ACCEPTED 08 August 2022
PUBLISHED 07 September 2022

CITATION
Pescuma VN, Ktori M, Beyersmann E, Sowman PF, Castles A and Crepaldi D (2022) Automatic morpheme identification across development: Magnetoencephalography (MEG) evidence from fast periodic visual stimulation. *Front. Psychol.* 13:932952. doi: 10.3389/fpsyg.2022.932952

Automatic morpheme identification across development: Magnetoencephalography (MEG) evidence from fast periodic visual stimulation

Valentina N. Pescuma[1]*, Maria Ktori[1], Elisabeth Beyersmann[2,3], Paul F. Sowman[2], Anne Castles[2,3] and Davide Crepaldi[1]

[1]Cognitive Neuroscience, International School for Advanced Studies (SISSA), Trieste, Italy, [2]School of Psychological Sciences, Macquarie University, Sydney, NSW, Australia, [3]Macquarie University Centre for Reading, Macquarie University, Sydney, NSW, Australia

The present study combined magnetoencephalography (MEG) recordings with fast periodic visual stimulation (FPVS) to investigate automatic neural responses to morphemes in developing and skilled readers. Native English-speaking children (N=17, grade 5–6) and adults (N=28) were presented with rapid streams of base stimuli (6Hz) interleaved periodically with oddballs (i.e., every fifth item, oddball stimulation frequency: 1.2Hz). In a manipulation-check condition, tapping into word recognition, oddballs featured familiar words (e.g., *roll*) embedded in a stream of consonant strings (e.g., *ktlq*). In the experimental conditions, the contrast between oddball and base stimuli was manipulated in order to probe selective stem and suffix identification in morphologically structured pseudowords (e.g., stem+suffix pseudowords such as *softity* embedded in nonstem+suffix pseudowords such as *trumess*). Neural responses at the oddball frequency and harmonics were analyzed at the sensor level using non-parametric cluster-based permutation tests. As expected, results in the manipulation-check condition revealed a word-selective response reflected by a predominantly left-lateralized cluster that emerged over temporal, parietal, and occipital sensors in both children and adults. However, across the experimental conditions, results yielded a differential pattern of oddball responses in developing and skilled readers. Children displayed a significant response that emerged in a mostly central occipital cluster for the condition tracking stem identification in the presence of suffixes (e.g., *softity* vs. *trumess*). In contrast, adult participants showed a significant response that emerged in a cluster located in central and left occipital sensors for the condition tracking suffix identification in the presence of stems (e.g., *softity* vs. *stopust*). The present results suggest that while the morpheme identification system in Grade 5–6 children is not yet adult-like, it is sufficiently mature to automatically analyze the morphemic structure of novel letter strings. These findings are discussed in the context of theoretical accounts of morphological processing across reading development.

KEYWORDS

magnetoencephalography, fast periodic visual stimulation, morphological processing, automatic morpheme identification, reading development, English

Introduction

Morphemes are the smallest linguistic units that bear meaning. For instance, a complex word like *artist* contains a stem, *art*-, and a suffix, -*ist*. Many languages are morphologically rich, meaning that their lexicon includes a great deal of complex words, by derivation, inflection, or compounding; it is estimated that 85% of the English lexicon is made up of complex words (Algeo and Algeo, 1993; Grainger and Ziegler, 2011).

Considering the important role that efficient morphological processing plays in skilled reading (Rastle, 2019), it is unsurprising that many studies in the psycholinguistic domain have focused on the sensitivity to morphological structure during visual word processing (for a review, see Amenta and Crepaldi, 2012). Several theories have been proposed over the years to account for the visual identification, comprehension, and reading aloud of complex words. Some of these theories dispose entirely of explicit morphological representations, and trace back the emergence of morphological effects to the appreciation of statistical regularities in mappings between form, meaning, and phonology (e.g., Seidenberg, 1987; Baayen et al., 2011; for a review, see Stevens and Plaut, 2022). Other *localist* models affirm the existence of morphological representations, either through different, serially arranged stages of processing (e.g., Crepaldi et al., 2010; Taft and Nguyen-Hoan, 2010; Taft, 2015), or along parallel routes (e.g., Grainger and Ziegler, 2011). More recently, Grainger and Beyersmann (2017; see also Beyersmann and Grainger, 2022) proposed that the analysis of the internal structure of words is initiated by the identification of stems as embedded, edge-aligned words. Although the cognitive architecture implied by this model is not substantially different from its predecessors, Grainger and Beyersmann's theoretical account is new in the proposal of a lexical trigger (i.e., the identification of word stems) for morphological analysis, as well as in the different mechanisms it assigns to the identification of stems and affixes. Notably, localist models of morphology build in different ways on the distinction between a level of morphological processing that is mostly based on form and one in which meaning plays a more substantial role.

Indeed, there is wide evidence that skilled reading is characterized by a rapid and automatic process of morphological analysis that operates on any printed word that merely has the orthographic appearance of being complex (Rastle et al., 2004). The main support for this so-called morpho-orthographic processing comes from masked priming studies. This research has found that adult readers routinely show facilitation not only for pairs of words with a semantically-transparent morphological relationship (e.g., *reader* primes the recognition of READ) but also for pairs with a pseudo-morphological relationship (e.g., *corner* primes CORN, relative to a purely orthographic baseline with no apparent morphological structure, e.g., *brothel-BROTH;* e.g., Longtin et al., 2003; Rastle et al., 2004; Beyersmann et al., 2016; see Rastle and Davis, 2003, for a review).

Interestingly, these behavioral findings are bolstered by neurophysiological studies (see Leminen et al., 2019, for an extensive review) examining the time course and neural bases of morphological processing. For example, Whiting et al. (2015), conducted a masked priming MEG study to investigate differences in the processing of simple (*walk*), complex (*farmer*), and pseudocomplex (*corner*) words. For both complex and pseudocomplex items, a similar morphological effect emerged around 330–340 ms in the left middle temporal gyrus (MTG), diverging from noncomplex stimuli. This pattern of findings suggests that both complex and pseudocomplex items undergo a "blind" decomposition process, reflecting morpho-orthographic processing (see also Lavric et al., 2011; Beyersmann and Grainger, 2022, for similar EEG evidence). This is further corroborated by fMRI evidence, such as the masked priming study by Gold and Rastle (2007), in which a similar pattern of reduced activation was observed in the left posterior middle occipital gyrus for pseudomorphologically related pairs (*archer-ARCH*) and for orthographically related ones (*pulpit-PULP*), and reduced activity of the posterior face fusiform gyrus was observed specifically for pseudomorphologically related pairs.

Morphological processing across reading development

Children as young as 7 years show evidence for explicit morphological knowledge. They can successfully manipulate and reflect on the morphological structure of words and novel letter strings, as measured by various types of morphological awareness tasks (e.g., Kirby et al., 2012). Furthermore, there is substantial evidence that young readers' morphological knowledge implicitly influences their online performance on word reading and recognition tasks (Rastle, 2019). For example, Carlisle and Stone (2005) showed that children in the initial (Grades 2 and 3) and later (Grades 5 and 6) years of primary school read aloud real morphologically complex words (e.g., *hilly*) more accurately than pseudo-morphological words, matched on number of syllables, spelling, and frequency (e.g., *silly*). Likewise, Burani et al. (2002) showed that Italian children between Grades 3 and 5 read aloud morphologically structured pseudowords (e.g., *donnista*, "womanist," composed of the root *donn*-, "woman" plus the suffix -*ista*, "-ist") more rapidly and accurately than pseudowords without a morphological structure (e.g., *dennosto*, composed of the non-root *denn*- plus the non-suffix -*osto*). When participating in a lexical decision task, the same children also showed greater difficulty in rejecting morphologically structured pseudowords; a morpheme interference effect that has since been replicated with children of similar school grades in French (Casalis et al., 2015) and in English (Dawson et al., 2018).

But at what stage of reading development does the ability to recognize morphemes rapidly and automatically emerge? To address this question, a series of masked priming studies sought direct evidence for morpho-orthographic processing in developing readers. The evidence they provided, however, is rather mixed. For example, using masked priming, Beyersmann et al. (2012) found that although English third and fifth graders showed facilitation for morphologically related pairs (e.g., *golden-GOLD*), there was no evidence for priming between pairs of words sharing pseudo-morphological (e.g., *mother-MOTH*) or purely orthographic (e.g., *spinach-SPIN*) overlap. But a different pattern of results emerged in a study by Quémart et al. (2011) with French-speaking children. In this experiment, third, fifth, and seventh graders yielded similar priming effects for opaque (*baguette-BAGUE*) and transparent pairs (*tablette-TABLE*), but no priming for orthographic (*abricot-ABRI*) or semantic (*tulipe-FLEUR*) pairs. Yet Schiff et al. (2012) found a different set of results in Hebrew, with fourth and seventh graders showing equally strong priming for prime and target pairs that were morphologically and semantically related, and seventh graders showing additionally a weak priming effect for pairs that were morphologically related and semantically unrelated—a pattern similar to that observed with adult readers of Hebrew (Bentin and Feldman, 1990; Frost et al., 1997).

More recently, Dawson et al. (2018) carried out a more fine-grained investigation into the emergence of adult-like morphological processing in English by including adolescent readers. Using unprimed lexical decision, they showed that although all groups of English-speaking participants rejected pseudo-morphemic pseudowords (e.g., *earist*) less accurately than control pseudowords (e.g., *earilt*), this difference was greater in adults and older adolescents (16–17 years) than in younger adolescents (12–13 years) and children (7–9 years). Furthermore, only adults and older adolescents exhibited a morpheme interference effect in their response times. Together these findings suggest that the way morphological representations are used in visual word recognition continues to undergo important changes during adolescence.

In summary, there is substantial evidence that within only a few years of reading instruction children demonstrate sensitivity to morphological structure during visual word processing. Yet, it remains unclear at what stage in reading development morphological processing is automatized. The available developmental data provide a mixture of results, with some recent evidence indicating that the morpheme recognition processes continue to develop even during adolescence. Admittedly, conclusions are further hindered by the different languages in which this research has been conducted. Indeed, the developmental trajectory of morphological processing appears to differ across languages, a claim that has found support in several recent cross-linguistic investigations (e.g., Beyersmann et al., 2020, 2021b; Mousikou et al., 2020). Another possibility is that the lack of clear evidence is due, at least in part, to issues related to commonly used behavioral paradigms, often requiring children to sit through long sessions and perform a somewhat unnatural task (e.g., primed or unprimed lexical decision), usually yielding quite small effects. One could, of course, take recourse to neurophysiological evidence to resolve this type of inconsistencies. However, to our knowledge, such developmental evidence is nonexistent. To overcome these limitations, the present study seeks to investigate automatic morpheme identification in developing readers by capitalizing on a relatively novel, behavior-free technique that combines Fast Periodic Visual Stimulation (FPVS) with electrophysiological recordings (Rossion, 2014).

Fast periodic visual stimulation (FPVS) and visual word recognition

The FPVS approach is based on the principle of neural entrainment (see Norcia et al., 2015, for a review), and when applied in the context of an oddball paradigm, it relies on frequency tagging to effectively capture visual discrimination processes at the level of the brain. This usually involves presenting sequences of base stimuli at a fast periodic rate (i.e., base stimulation frequency F) with oddball stimuli periodically inserted at fixed intervals within the stream (every n^{th} item), thus resulting in a slower presentation rate (i.e., oddball stimulation frequency F/n). A peak in the neural signal at the oddball stimulation frequency (and its harmonics) indexes the brain's ability to successfully discriminate between oddball and base stimuli. Critically, oddball responses are selective to the dimension that differentiates oddballs from base stimuli.

To date, the FPVS-oddball paradigm has been most commonly used to investigate face processing and recognition (e.g., Dzhelyova and Rossion, 2014; Rossion, 2014; Rossion et al., 2015; Retter and Rossion, 2016; Quek et al., 2018). However, thanks to its versatility, it has gained popularity in many other areas of cognitive processing, including visual word recognition. For example, Lochy et al. (2015) combined FPVS and EEG recordings to probe selective neural representations of words (relative to pseudowords) in skilled adult readers. Even more relevant to the present study, however, the FPVS approach enjoys several advantages that make it ideal for special populations like children. Specifically, the approach is highly sensitive such that only a few minutes of stimulation are sufficient to elicit robust responses with a high signal-to-noise ratio (SNR). This diminishes the need for a large number of experimental trials, especially when small effects are considered. Furthermore, the neural responses elicited by FPVS are clearly and objectively identifiable in the pre-defined base and stimulation frequencies, thus eliminating the subjectivity that can, at times, accompany the detection of event-related components. Finally, the approach does not require participants to actively engage with the experimental stimuli. As such, neural discrimination responses are obtained implicitly and automatically, and are devoid of potential contamination from task-induced cognitive and decisional processes. In this respect, Lochy et al. (2016) already provide us with proof of concept by successfully combining FPVS with EEG recordings to elicit selective neural responses to letter strings in young preschoolers. Here, we pair this technique with magnetoencephalography (MEG) to investigate for the first time an even more fine-grained level of visual word processing, namely the identification of morphemes, across reading development.

The present study

In the present FPVS-MEG study, we presented a group of native English-speaking children (Grades 5 and 6) and a group of native English-speaking adults with rapid sequences of carefully constructed pseudowords in order to examine automatic neural responses to morphemes. By definition, pseudowords are not represented in the lexicon. As such, they constitute ideal linguistic stimuli to explore morpho-orthographic analysis that is considered to operate rapidly and automatically on any printed letter string prior to lexical access (Taft and Forster, 1975). The experimental stimuli consisted of four types of pseudowords. We manipulated the contrast between oddball and base stimuli in order to probe selective stem and suffix identification (see Figure 1). Specifically, to investigate stem identification, two of the experimental conditions featured oddball pseudowords that were either composed of a real English stem and a real English suffix (e.g., *softity*; Condition 1) or of a stem only (e.g., *softert*; Condition 2). To investigate suffix identification, two additional experimental conditions featured oddball pseudowords that were either composed of a real English stem and a real English suffix (e.g., *softity*; Condition 3) or of a suffix only (e.g., *terpity*; Condition 4). In order to elicit a contrast, oddballs were embedded in streams of base stimuli which did not contain the critical morpheme. Namely, in Condition 1, the stem + suffix oddballs (*softity*) were embedded in streams of nonstem + suffix base stimuli (*trumess*), so that the contrast between the two would track stem identification, in the presence of a suffix. In Condition 2, the stem + nonsuffix oddballs (*softert*) were embedded in streams of nonstem + nonsuffix base stimuli (*trumust*), so that the contrast would still track stem identification, this time in the absence of a suffix. Similarly, Conditions 3 and 4 featured, respectively, stem + suffix oddballs embedded in stem + nonsuffix base stimuli (*softity* vs. *stopust*), and nonstem + suffix oddballs embedded in nonstem + nonsuffix base stimuli (*terpity* vs. *trumust*), thus tracking suffix identification, either in an exhaustively decomposable morphological context or not.

In the current design, a robust oddball response across Conditions 1 and 2, and/or across Conditions 3 and 4, would be suggestive of sensitivity to the individual morphemes regardless of whether the oddball can be exhaustively decomposed into a stem and a suffix. On the other hand, an oddball response only in Conditions 1 and 3 would indicate sensitivity to, respectively, stems and suffixes only when presented in an exhaustively decomposable morphological oddball (i.e., in the presence of another morpheme). Adults were administered all four conditions, while children were only administered Conditions 1 and 3, in consideration of the limited time that they could spend in the MEG room.

Materials and methods

Participants

We recruited 32 skilled adult readers (age range: 18–45) and 21 developing readers (enrolled in Grades 5 and 6 of the Australian school system at the time of testing). Data from four adults and four children were eventually removed from the final sample analyzed here, either for excessive head motion (greater than 5 mm for adults; greater than 11 mm for children) or due to an excessive presence of artifacts. This left us with 28 skilled adult readers (age: mean = 22.93 years, sd = 6.38 years) and 17 developing readers (age: mean = 10.59 years, sd = 0.79). Adult participants were recruited through Macquarie University and were offered course credit, where applicable, or monetary compensation. Children were recruited through *Neuronauts*, a dedicated Macquarie University portal, and their families were awarded monetary compensation for their time. Both studies were approved by the Macquarie University Human Research Ethics Committee. All participants were native English speakers and right-handed; none reported neurological or developmental issues, language difficulties, or claustrophobia. They all had normal or corrected-to-normal (through contact lenses) vision.

Stimuli

All conditions consisted of five 60-s trials, for adults, and of six 60-s trials, for children. A within-participant block design was adopted. A manipulation-check condition (Condition 0) probing visual word identification was administered to all participants. Adults were administered four experimental conditions, while children were only administered two of these (Conditions 1 and 3).

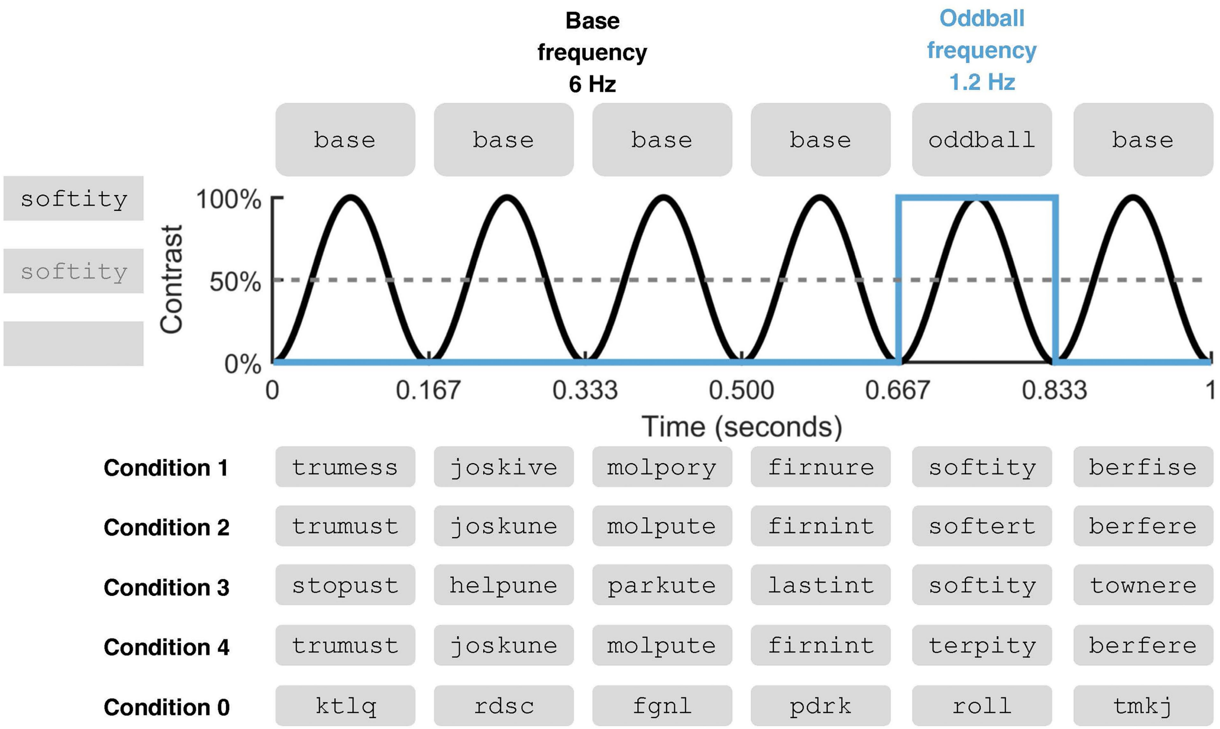


FIGURE 1
Fast Periodic Visual Stimulation (FPVS) in the context of an oddball paradigm (figure adapted from De Rosa et al., 2022). The schematic illustration pertains to one second of stimulation. For a gradual and smooth transition between them, stimuli were presented *via* sinusoidal contrast modulation at 6Hz during 60s, with each cycle reaching full contrast after 83.5ms and lasting a total of 167ms. Oddball stimuli appeared every fifth item (6/5=1.2Hz). Examples are given for the different types of sequences used in the experimental conditions. To examine stem identification, Condition 1 featured stem+suffix oddballs embedded in a stream of nonstem+suffix pseudowords, and Condition 2 (adult sample only) featured stem+nonsuffix oddballs embedded in a stream of nonstem+nonsuffix pseudowords. To examine suffix identification, Condition 3 featured stem+suffix oddballs embedded in a stream of stem+nonsuffix pseudowords and Condition 4 (adult sample only) featured nonstem+suffix oddballs embedded in a stream of nonstem+nonsuffix pseudowords. A manipulation-check condition (Condition 0) was administered to all participants and examined a selective neural response to words that appeared as oddballs within a sequence of unpronounceable letter strings. During stimulation, participants engaged in the independent task of monitoring and responding to changes in the color of a fixation cross at the center of the screen.

In Condition 0, 4-letter English words (oddball stimuli) were embedded in non-pronounceable 4-consonant strings (base stimuli). The purpose of this condition was to ensure that the paradigm worked correctly, as prior FPVS research reports robust oddball responses to words embedded in nonwords (see, e.g., Lochy et al., 2015, 2016). In Condition 1, oddball stimuli were pseudowords made up of a real English stem and a real English suffix (e.g., *softity*), which were embedded in pseudowords made up of a nonstem and a real English suffix (e.g., *trumess*). In Condition 2, pseudowords made up of a real stem and a nonsuffix (e.g., *softert*) were used as oddballs and embedded in pseudowords made up of a nonstem and a nonsuffix (e.g., *trumust*). In Condition 3, oddball pseudowords were made up of a real stem and a real suffix (e.g., *softity*) and were embedded in pseudowords made up of a real stem and a nonsuffix (e.g., *stopust*). Lastly, in Condition 4, pseudowords made up of a nonstem and a nonsuffix (e.g., *terpity*) were embedded in pseudowords made up of a nonstem and a suffix (e.g., *trumust*). Sequence examples for each condition are reported in Figure 1.

Contrasts in each condition were set in order to tap into stem or suffix identification. Specifically, an oddball response in Condition 1 (and 2, in the adult sample only) would index stem identification, and in Condition 3 (and 4, in the adult sample only) it would index suffix identification. The administration of two additional conditions (2 and 4) to the adult participants was intended to shed light on the role of context for the identification of morphemes—that is, whether a robust response to oddballs was present only when they could be fully broken down into the two constituent morphemes (Conditions 1 and 3), or whether morphemes would also be successfully identified when oddballs featured only one morphemic constituent (Conditions 2 and 4).

In the version of the experiment with adult readers, stimuli were composed of 12 items for each type: stems, nonstems, suffixes, and nonsuffixes. Nonstems and nonsuffixes were created from the set of existing stems and existing suffixes, while keeping the same length, Consonant-Vowel structure, and minimising orthographic overlap with existing words (e.g., *terp* was created as a nonstem from *soft*, *ert* was created as a nonsuffix from *ity*). Stem and nonstems were 4 letters in length, while suffixes and nonsuffixes were 3-letter long. The 12 nonsuffixes were non-morphemic endings attested in English. Each set of (non) stems and (non)suffixes was divided into two subsets of 6 items; stimuli were then obtained by combining each element in one subset with each element of another. This procedure generated 72 unique combinations (6 items in the first set, times 6 items in the second set, times 2 subsets) of each type (stem + suffix, nonstem + suffix, stem + nonsuffix, nonstem + nonsuffix), yielding a total of 288 unique stimuli.

In the developmental version of the experiment, the building blocks were reduced to 6, a subset of those used for skilled adult readers (Non)stems and (non)suffixes were combined by groups of 3, to obtain 18 (3*3*2) unique combinations of each type (stem + suffix, nonstem + suffix, stem + nonsuffix), yielding a total of 54 unique stimuli.

All building blocks (stems, nonstems, suffixes, and nonsuffixes) are reported in Table 1. Statistics for stems and suffixes were obtained from two different linguistic databases: SUBTLEX-UK (Van Heuven et al., 2014) and MorphoLex (Sánchez-Gutiérrez et al., 2018). Specifically, while the former frequency database is particularly relevant for its size (over 160,000 types and 200 million tokens from English television show subtitles), the latter resource is a rich morphologically tagged database for English, allowing the extraction of metrics related to the use of items as morphemes in the language.

Stem selection

All selected stems were four-character long and had a CVCC or CCVC consonant-vowel structure. Here, we describe the features of the 12 stems used as constituents in the adult version of the experiment, a subset of which was used in the version with developing readers; the statistics related to the six stems used in the version for children are provided in square brackets. Database exploration, extraction, and calculation of relevant metrics were performed using R (R Core Team, 2021) within RStudio (RStudio Team, 2021). The average SUBTLEX-UK log Zipf frequency was 5.13, with a sd of 0.43 [mean: 4.93, sd: 0.28]; the average stem token frequency in MorphoLex was 155,217, with a sd of 1,44,128.60 [mean: 1,32,510, sd: 1,27,152.30], while the average stem family size in MorphoLex was 15.50, with a sd of 9.64 [mean: 22.83, sd: 8.33]. Finally, the average Levenshtein distance (OLD20; Yarkoni et al., 2008), a lexical density index based on the average distance of the 20 nearest neighbors in the lexicon, was calculated through the R package *vwr* (Keuleers, 2013) using SUBTLEX-UK, the largest resource considered here. The higher the OLD20 value of a stimulus, the lower its orthographic neighborhood. All stems had a mean OLD20 of 1 and a sd of 0 [mean: 1, sd: 0].

Nonstem selection

Nonstems were pseudowords generated with the same length and CV structure types as the real stems, for the items to be orthographically and phonotactically legal, while at the same time minimizing orthographic overlap with the selected stems. The mean OLD20 for our nonstem selection was 1.14, with a sd of 0.21 [mean: 1.07, sd: 0.16].

Suffix selection

Twelve three-letter derivational suffixes were shortlisted from the CELEX database (Baayen et al., 1993). The CV structure types of the selected suffixes were VCC, VCV, CVC, and VVC. A subset of six suffixes was used for the developmental version of the experiment. The same exploration and analysis were performed as for the above-described stem selection. The average SUBTLEX-UK log Zipf frequency was 2.41, with a sd of 0.59 [mean: 2.41, sd: 0.73]. We ensured, through MorphoLex, that all selected items were productive suffixes in the English language. The average suffix token frequency in MorphoLex was 5,14,914.40, with a sd of 4,52,230.50 [mean: 6,43,484.20, sd: 5,23,213.40], while the average suffix family size in MorphoLex was 319.25, with a sd of 226.44 [mean: 431.50, sd: 145.89]. All suffixes had a mean OLD20 of 1 and a sd of 0 [mean: 1, sd: 0].

TABLE 1 Unique stems, nonstems, suffixes, and nonsuffixes combined to generate nonword stimuli, and the respective OLD20 statistics. All listed items were used to construct stimuli in the study with adults, while a subset (underlined items) was used to construct stimuli for the study with children. The full sets of stimuli are reported in the Supplementary Tables S1, S2.

Stem	Nonstem	Suffix	Nonsuffix	OLD20 stem	OLD20 nonstem	OLD20 suffix	OLD20 nonsuffix
help	josk	ity	ert	1.00	1.55	1.00	1.00
soft	terp	ive	une	1.00	1.00	1.00	1.00
last	firn	ory	ute	1.00	1.00	1.00	1.00
ship	bron	ure	int	1.00	1.00	1.00	1.00
stop	trum	ous	ald	1.00	1.00	1.00	1.00
hold	burk	ise	ere	1.00	1.00	1.00	1.00
park	molp	ful	sal	1.00	1.40	1.00	1.00
jump	lort	ist	arn	1.00	1.00	1.00	1.00
town	bemp	ite	ene	1.00	1.40	1.00	1.00
bird	jelt	ish	ult	1.00	1.00	1.00	1.00
farm	culp	ese	oke	1.00	1.35	1.00	1.00
milk	tand	ess	ust	1.00	1.00	1.00	1.00

Nonsuffix selection

We selected 12 three-letter clusters that occur as non-morphological endings in English, with a mean OLD20 of 1 and a sd of 0 [mean: 1, sd: 0].

Manipulation-check condition stimuli

For Condition 0, we selected 72 4-letter words (with various CV structure types, but always ending with a consonant) and 72 4-letter non-pronounceable consonant strings. A subset of 18 words and 18 consonant strings was used for the experiment with children. The average SUBTLEX-UK log Zipf frequency was 4.71, with a sd of 0.54 [mean: 4.91, sd: 0.54].

Stimuli combinations

Statistics for the stimuli used in the developmental version of the experiment, which did not feature nonstem + nonsuffix combinations, are reported in brackets. OLD20 statistics were then computed for all stimuli. Stem + suffix combinations had a mean OLD20 of 2.32 and a sd of 0.30 [mean: 2.43, sd: 0.27], stem + nonsuffix combinations had a mean OLD20 of 2.49 and a sd of 0.37 [mean: 2.55, sd: 0.43], nonstem + suffix combinations had a mean OLD20 of 2.47 and a sd of 0.32 [mean: 2.47, sd: 0.32], and nonstem + nonsuffix combinations had a mean OLD20 of 2.62 and a sd of 0.31. All unique stimuli administered to skilled adult readers can be found in the Supplementary material.

Trial structure

Each trial comprised a 60-s stimulation sequence in which stimuli were presented *via* sinusoidal contrast modulation at 6 Hz (i.e., six stimuli per second)—each individual stimulus appeared gradually, reaching a contrast peak after 83.5 ms (for a schematic illustration, see Figure 1). Each 60-s trial thus contained 360 stimuli overall. Each oddball stimulus appeared every five items (6 Hz/5 = 1.2 Hz); therefore, the stimulation sequence in each trial included 72 oddballs and 288 base items. The oddball stimuli were unique items in the adult design, whereas in the developmental design a greater number of item repetitions was present: in each trial, every oddball was delivered a total of 4 times (18*4 = 72). The sets of stimuli were generated through pseudo-randomization, using in-house R scripts within RStudio for the adult version of the experiment, and using Mix software (Van Casteren and Davis, 2006) for the developmental version. As the process could not be entirely automated, lists were then checked and edited manually when deemed necessary, in order to prevent repetitions of the same combinations within each stimulation sequence. With both skilled and developing readers, we ensured that the same stimulus was not repeated within each 1-s of the stimulation sequence (i.e., the minimum distance between stimulus repetitions was 5). Overlayed to this stimulus sequence, a fixation cross (12 pixels) was constantly present at the center of the screen. The color of the cross changed randomly (from blue to red and vice versa), and participants were instructed to tap a button whenever they detected a color change (Lochy et al., 2015, 2016).

In the experiment with skilled readers, visual stimuli were displayed in black Courier New font, with a font size of 100 pt., within a white bounding box of 500*150 pixels. In the developmental version of the experiment, stimuli were slightly enlarged and they were displayed in black Courier New bold font, with a font size of 110 pt., within a white bounding box of 510*170 pixels. A large font size was adopted for both skilled and developing readers due to their distance from the screen. In both versions of the experiment, stimuli were displayed over a gray background.

Procedure

Responses were recorded through a fiber-optic button box (fORP, Current Designs, Philadelphia, PA, United States). Accuracy in this task was very high for all participants (skilled adult readers: mean = 97.83%, sd = 1.84; developing readers: mean = 95.64%, sd = 4.84). This behavioral task was administered with the mere purpose of ensuring that participants engaged with the area in which the stimuli would be presented. Trials were separated by a 25-s break. The break ended with a 10-s countdown to the new trial. A 2-min break was given twice between recording blocks, to allow head location measurements to be performed; one last measurement was performed at the end of the MEG recording. Overall, the MEG testing in the MSR required 45–50 min with adults and a maximum of 30 min with children.

Apparatus

Data were collected at the KIT-Macquarie Brain Research Laboratory (Sydney, Australia). Participants lay supine in a dimly lit and magnetically shielded room (MSR). Continuous MEG recordings were acquired using a 160-channel whole-head coaxial gradiometer system (KIT, Kanazawa Institute of Technology, Japan) at a sampling rate of 1,000 Hz, with an online bandpass filter of 0.03–200 Hz. Visual stimuli were delivered through a projector (sampling rate: 60 Hz) and mirrored onto a translucent screen mounted above the participant's head, at a distance of approximately 110 cm. The experiment was controlled *via* a Windows desktop computer, using MATLAB 2019a (MATLAB, 2019) and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007). Parallel port triggers were used to mark the beginning and end of each trial, and a photodiode was used to check the correct delivery of oddball stimuli, through a white square in the bottom right corner of the screen. Participants' head shapes were recorded using the Polhemus FASTRAK system and digitizing pen (Colchester, VT, United States). Throughout the MEG recording session, participants wore an elastic cap with five marker coils which allowed tracking the head location relative to the MEG helmet and to measure motion over time.

MEG data preprocessing

Data were preprocessed in MATLAB using the FieldTrip toolbox for EEG/MEG analysis (Oostenveld et al., 2011) as well as in-house functions. A lowpass filter of 100 Hz was applied; continuous MEG recordings were epoched into trials using a custom-made trial function. In trial epoching, a pre-stimulus interval and a post-stimulus interval were set in order to avoid edge artifacts. Respectively, the first two oddball cycles (i.e., the first 1.67 s of stimulation) and the last one (833 ms) were cut from each trial, resulting in trials of 58.33 s each (see Lochy et al., 2015). Recordings were then downsampled to 250 Hz. Data from eight subjects (four adults and four children) with excessive noise artifacts (one adult) or excessive movement artifacts (three adults and four children) were discarded entirely. Noisy channels were removed based on visual inspection, and channel interpolation was performed (neighbors were defined using FieldTrip functions through a triangulation method). One dataset per condition (five trials per condition for adults, six trials per condition for children) per participant was obtained.

Frequency analysis

A very similar procedure to the one used in Lochy et al. (2015, 2016) was adopted. Each participant's trials were averaged by condition and subjected to a Fast Fourier Transform. By calculating the square root of the sum of squares of the real and imaginary parts divided by the number of data points, power spectra were then computed for each sensor. As each epoch was 58.333 s long, the frequency resolution was 1/58.333 = 0.0171 Hz. The spectra were then normalized by dividing the mean power spectrum of each frequency bin by the mean of the surrounding 20 bins (10 on either side, excluding immediately adjacent bins), thus obtaining a signal-to-noise ratio metric (SNR). Oddball response was defined as the average SNR of the response at the oddball (1.2 Hz, precisely 1.1962 Hz as calculated in the collected datasets[1]) stimulation frequency and its corresponding first three harmonics (2.4, 3.6, 4.8 Hz, precisely 2.3924, 3.5886, 4.7848 Hz, as calculated in the collected datasets). Hence, the final dataset consisted of 22,400 data points for the adult sample (28 participants, times 5 conditions, times 160 channels) and 8,160 data points for the children sample (17 participants, times 3 conditions, times 160 channels).

Results

Cluster-based permutation analysis in the sensor space

The present results only pertain to sensor-level analysis, as source-level analysis could not be performed due to technical limitations. A data-driven approach to the analysis was adopted. Although the primary interest is the visual identification of morphemes, and the present paradigm emphasizes quick and automatic visual access, morphological analysis might also trigger higher-level semantic processing. Therefore, we aimed at assessing the existence of any potential tagging of the oddball frequency at the whole-brain level.[2] To this aim, we conducted a cluster-based permutation test at the sensor level (Maris and Oostenveld, 2007), adapted for FPVS-MEG datasets, which span over space (sensors), but not time. Using grand-averaged datasets per participant per condition, cluster-based permutation was performed on the power spectrum at the averaged oddball frequency and first three harmonics (see "Frequency Analysis"), across all 160 sensors. We used a within-subject design and adopted a Montecarlo method for calculating probabilities. A minimum of two neighboring channels were required for a cluster to be defined. A cluster alpha level of 0.05 was set and a one-tailed *t*-test was run (we only contemplated the hypothesis that the SNR was higher than 1). An alpha level of 0.05 was set and 5,000 randomizations were performed. With this configuration, cluster-based permutation was run against an array of ones, representing the noise level in each channel (i.e., the null distribution).

The results for the adult skilled readers are illustrated in Figure 2. In Condition 0, which taps into whole-word identification, we found one large cluster essentially encompassing the whole posterior part of the scalp, with a peak in the left hemisphere [$t(27) = 416.46$, $p < 0.001$, panel A]. A cluster also emerged for Condition 3, which probes suffix identification in the presence of a stem [$t(27) = 113.02$, $p < 0.001$, panel B]. This cluster is much smaller than in Condition 0 and extends along the midline from the vertex to the back of the brain, and then along the left ventral stream. No other significant clusters emerged, there was no robust response to the oddball stimuli in Condition 1 (designed to track stems in the presence of affixes), Condition 2 (stems in the absence of affixes), and Condition 4 (suffixes in the absence of stems).

The results for the developing readers are illustrated in Figure 3. For Condition 0, a cluster emerged [$t(16) = 347.17$, $p < 0.001$, panel A] which is largely left-lateralized and extends over temporo-parieto-occipital sensors. Furthermore, an occipital cluster, mostly located around the midline, emerged in Condition

---

[1] This calculation was performed by means of a custom Matlab function which calculates the observed frequency of stimulus delivery, averaged across participants.

[2] As a sanity check of the effectiveness of the visual stimulation, we furthermore ensured that all participants displayed neural entrainment at the base stimulation frequency (6Hz). See relevant section in the Supplementary material.

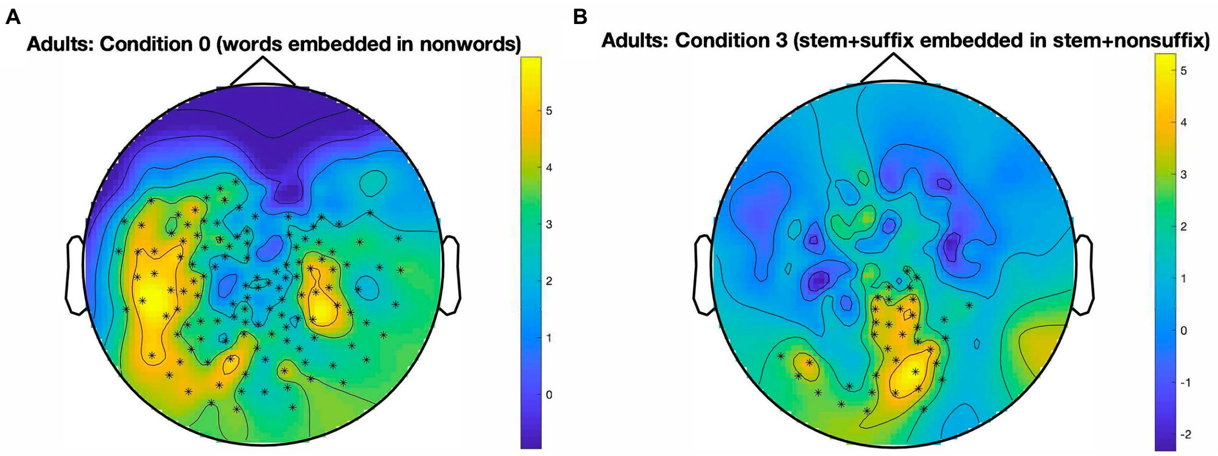


FIGURE 2
Sensor-level clusters in which a significant oddball response emerged, by condition. Data from skilled adult readers. **(A)**: Large temporo-parieto-occipital cluster (mostly left-lateralized, with a right-lateralized part) indicating widespread identification of words in nonwords, in Condition 0; $p<0.001$, cluster alpha level=0.05. **(B)**: Left and central occipital cluster for the identification of stem+suffix oddballs in stem+nonsuffix base stimuli, in Condition 3; $p<0.001$, cluster alpha level=0.05. Color bars represent SNR on a continuous scale (blue=low, yellow=high). Condition 0: *words* in *nonwords* (e.g., *roll* in *kltq*); Condition 3: *stem+suffix* in *stem+nonsuffix* (*softity* in *terpert*).

1 [$t(16) = 74.90$, $p = 0.007$, panel B], in which stem identification in the presence of suffixes is tracked.

Discussion

Using an FPVS-oddball paradigm paired with MEG recordings, the current study sought to examine the automatic identification of morphemes in pseudowords, within a group of developing readers and a group of skilled adult readers of English. Our design also included a manipulation-check condition, that was administered to tap into whole word identification (word oddballs embedded in consonant letter strings), serving as a benchmark for our paradigm. This condition confirmed that the FPVS technique, which has already been quite extensively paired with EEG recordings, can also be successfully employed in MEG studies investigating visual word identification. A cluster-based permutation analysis at the sensor level revealed a large temporo-parieto-occipital cluster at the oddball frequency, that was present in both children and adults. As expected and consistent with previous FPVS studies using psycholinguistic material (e.g., Lochy et al., 2015, 2016), this cluster peaked predominantly in the left hemisphere for both reader groups. Interestingly, however, the cluster spread more anteriorly in adult readers. This could be taken to suggest that even automatic and implicit word identification might trigger full processing, possibly up to the semantic stage.

Although direct parallels with the studies by Lochy et al. (2015, 2016) are not fully warranted due to differences between EEG and MEG with respect to the localization of FPVS responses (see Hauk et al., 2021), one could tentatively explain the more widespread word identification response that we observe here (relative to a localized orthographic response in an occipito-temporo-parietal region) as stemming from two important differences between our studies and Lochy and colleagues'. First, in the experiment with skilled adult participants, we adopted a lower stimulation frequency (6 Hz, vs. 10 Hz in Lochy et al., 2015), resulting in a longer presentation time for each stimulus (167 ms vs. 100 ms). Second, in the developmental version of the experiment, our participants were already quite experienced readers (fifth and sixth graders) compared to the preschooler sample of Lochy et al. (2016).

Remarkably, word identification responses were similar across our developing and adult participants. These findings suggest that fifth and sixth graders, with just a few years of reading instructions, have already built up a highly sophisticated visual word identification system, roughly comparable to that of adults, when it comes to automatic and implicit detection of real words.

In relation to the core research question of the present study, whether automatic identification of morphemes emerges in complex pseudowords, the results revealed several intriguing differences between adults and children. With respect to the identification of morphemes, stems were more reliably identified by developing readers, whereas skilled readers showed sensitivity to suffixes. Suffixes represent salient units in the language, both from a semantic (as they convey systematic meaning) and from a perceptual and orthographic (as frequent chunks) point of view (Lelonkiewicz et al., 2020). Solid signs of suffix identification emerged only in the context of exhaustively decomposable complex pseudowords, which suggests that the process is not simply a catch of frequent and salient units, but involves a comprehensive (morphological) analysis of the whole string. This result aligns with the wealth of studies showing that our cognitive system heavily relies on morphology during reading and visual word identification (e.g., Rastle et al., 2000; Marslen-Wilson and

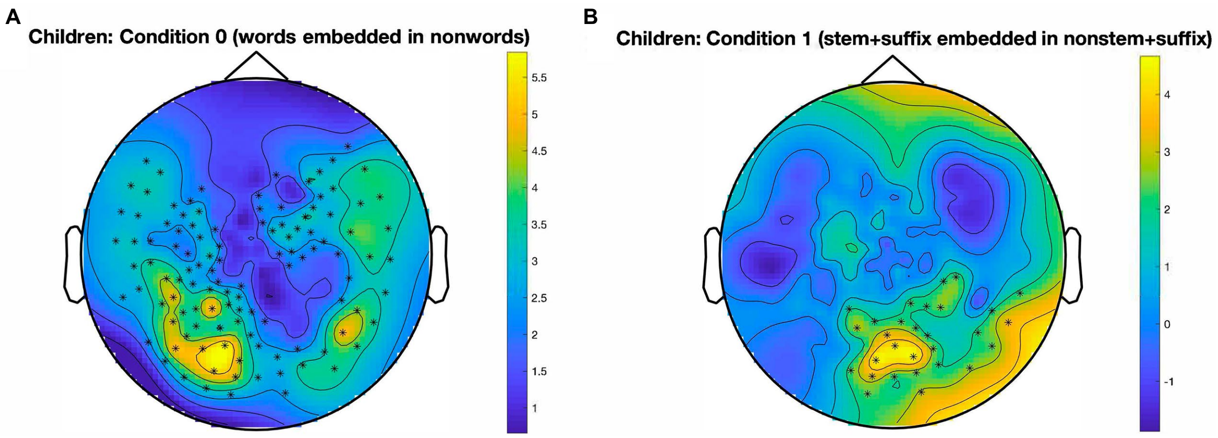


FIGURE 3
Clusters in which a significant oddball response emerged, by condition. Data from developing readers. **(A)**: Temporo-parieto-occipital cluster, largely left-lateralized, for the identification of words in nonwords, in Condition 0; *p*<0.001, cluster alpha level=0.05. **(B)**: Central occipital cluster for the identification of stem+suffix oddballs in nonstem+suffix base stimuli, in Condition 1; *p*=0.007, cluster alpha level=0.05. Color bars represent SNR on a continuous scale (blue=low, yellow=high). Condition 0: *words* in *nonwords* (*roll* in *kltq*); Condition 1: *stem+suffix* in *nonstem+suffix* (*softity* in *terpity*).

Tyler, 2007; Amenta and Crepaldi, 2012; Whiting et al., 2015; Leminen et al., 2019; Beyersmann et al., 2021a). Interestingly, this finding highlights one main limitation of many current models of the visual identification of complex words (e.g., Crepaldi et al., 2010; Taft and Nguyen-Hoan, 2010; Grainger and Ziegler, 2011). These models are fundamentally based on a spreading activation mechanism, and therefore would all predict that a stem is activated any time their constituting letters are present in the input. There is no plausible computational mechanism in those models that would explain how the presence of a suffix vs. a non-suffix might trigger vs. kill the activation of a stem representation.

Our pattern of results is also in tune with findings by Beyersmann et al. (2021b), who provided evidence for a greater steady-state visually evoked potential (SSVEP) magnitude for suffixes, compared to non-suffixes. Such activation boost is taken as an index of an additional semantic feedback mechanism, beyond morphological decomposition, which would instead be sufficient for the identification of stems.

In our developmental sample, we found evidence for stem identification, again, in exhaustively decomposable stimuli (i.e., made up of a real stem and a suffix), suggesting that children in Grades 5–6 have already developed an automatic morpheme identification system, albeit not adult-like. There may be two reasons for the presence of a stem (but not suffix) response. First, stems are often encountered as whole words in English (see, e.g., Grainger and Beyersmann, 2017, 2021); from this point of view, they might be even more perceptually salient than suffixes, given that the surrounding blank spaces might serve as "chunking cues" that help the system identify these items as important functional units. Second, stems are more informative about word identity, allowing to narrow the lexical and semantic interpretation of a word more than a suffix does *per se*. For example, upon encountering *dark*-, a reader can reliably predict the general meaning of the rest of that word; instead, many different words end in *-ness*.

In line with this, Grainger and Beyersmann (2017) suggested that what is typically interpreted as morpho-orthographic processing may in fact reflect a mechanism of embedded word/stem identification that is not, *per se*, genuinely morphological, i.e., it would operate independently of the presence of an affix. This account matches with the recent observation that when lexical competition is partialled out in priming experiments—that is, when pseudowords are used—affixed and non-affixed primes provide the same amount of facilitation (*farmald-FARM*=*farmness-FARM*). Note, however, that this hypothesis could not be fully tested in our developmental sample, as the oddball stimuli in the two experimental conditions administered to children comprised only fully decomposable pseudowords. Moreover, the adult data seems to challenge this assumption, as clear signs of sensitivity to the stems only emerged in the presence of a suffix; this might be due to the intrinsic differences between the priming tasks that contributed most of the experimental basis for Grainger's and Beyersmann's model, and the paradigm we employed here.

Overall, the present findings can be interpreted as corroborating SSVEP evidence by Beyersmann et al. (2021a), where, on the one hand, rapid stem identification was facilitated by the presence of a suffix (or pseudo-suffix), and, on the other, suffixed words received an activation boost relative to non-suffixed ones. This is traced back to the same mechanism: the activation of embedded stems. The observed neural response to stems in children and to suffixes in adults suggests that sensitivity to morphemes differs across reading development, with stems being identified as salient units by the developing reading system (see also Grainger and Beyersmann, 2017, 2021), and suffixes acquiring saliency in a more mature system, due to the higher frequency with which they are encountered in words.

With some caution against making assumptions about potential neural sources, the fact that a response is elicited by morphemes in sensors that span over occipital regions suggests that suffixes are likely processed as visual units, at least at this stage of processing. This aligns with theories positing the existence of a level of morphological analysis that is mostly based on form (e.g., Crepaldi et al., 2010; Grainger and Ziegler, 2011; Xu and Taft, 2014). At this level of analysis, morphemes are primarily seen as frequent, statistically associated clusters of letters, perhaps not so different from what happens in other domains of vision (e.g., Vidal et al., 2021). It is well known that neural circuitry in the ventral stream is particularly apt at finding regularities in the co-occurrence of lower-level units, to then build higher-level representations that exploit such regularities (e.g., Dehaene et al., 2005; Tkačik et al., 2010). This property is particularly prominent in the domain of visual word identification, which is characterized by lower-level units (i.e., letters) that bind together higher-level objects (i.e., morphemes and words). In this context, it should not be surprising that morphemes are captured as chunks of strongly associated letters.

Experimental evidence in support of a view of visual word identification as mostly relying on the detection of (also morphological) regularities is growing. For example, Chetail (2017) asked participants to familiarize themselves with an artificial lexicon made up of pseudo-characters. The lexicon was such that some bigrams were particularly frequent; when participants were involved in a wordlikeness task with entirely novel stimuli, those that contained the frequent bigrams were judged as more word-like. So, even in a completely unfamiliar novel lexicon, made up of completely unfamiliar pseudo-characters, a few minutes of exposure were sufficient for participants to develop sensitivity to small clusters of particularly high frequency. With a similar design and experiment, Lelonkiewicz et al. (2020) were able to reproduce effects that emerged in morphological pseudowords (e.g., Taft and Forster, 1975; Crepaldi et al., 2010) with an artificial lexicon that was entirely devoid of any phonological or semantic ties, that is, a set of purely visual, non-linguistic entities made up of sequences of pseudo-characters. These data suggest that at least part of the morphological effects that we observe with genuine linguistic material can be reproduced in purely visual, non-linguistic systems.

It is less clear why in the children's data (with respect to stem identification), and partially in the adults' data (with respect to suffix identification), a cluster for morpheme identification emerged centrally in occipital sensors. Further FPVS-MEG investigations of the neural source(s) of morpheme identification response would ideally complement the sensor-level findings that we reported. A cautious explanation for the largely central cluster for stem identification observed in the developing readers is that it might reflect a type of processing which is less specific to morpho-orthographic units, perhaps suggestive of a more general lexical/semantic response. This would align with accounts according to which, along reading development, morpho-semantic processing matures earlier than morpho-orthographic processing, which is hypothesized to emerge only at the last stages of reading development (Grainger and Beyersmann, 2017; Beyersmann and Grainger, 2022), and to still be maturing during adolescence (Dawson et al., 2018). Alternatively, the perceptual and automatic nature of a paradigm like FPVS might have boosted those components of morpheme identification that are not specifically linguistic, but more generally visual in nature. This would again be in line with recent evidence showing that several aspects of orthographic and morphological processing can be replicated with exclusively visual material that shares the same statistical features of real language (e.g., Chetail, 2017; Lelonkiewicz et al., 2020; Vidal et al., 2021).

Conclusion

In the present FPVS-MEG study, we showed that in Grade 5-6 children sensitivity to morphological structure, albeit not adult-like, has already sufficiently matured to be captured through an implicit, behavior-free paradigm such as FPVS. Moreover, the present results suggest that morpheme identification is stronger in strings that can be exhaustively decomposed into their constituent morphemes (i.e., when both a real stem and a real suffix are present). Signs of this identification process appeared in sensors that morphological identification as a predominantly visual process, and thus potentially linked to language-agnostic, statistical learning mechanisms (e.g., Rastle and Davis, 2003; Crepaldi et al., 2010; Lelonkiewicz et al., 2020; Vidal et al., 2021). Additionally, our findings make a methodological contribution by providing a further demonstration that the FPVS paradigm can be employed to investigate even more fine-grained processes of visual world recognition than previously explored in the literature.

Data availability statement

Analysis scripts and data can be accessed through the project's OSF repository: https://osf.io/ns93h/.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Macquarie University, Sydney, NSW, Australia. Written informed consent to participate in this study was provided by the participants or, in the case of underage participants, by a parent/legal guardian/next of kin.

Author contributions

The study was conceptualized and designed by DC, EB, MK, and VP, with input from AC. Experimental stimuli and lists were created by VP, in collaboration with DC, EB, MK, and with input from AC. Participant recruitment and MEG data collection were performed by VP. MG data were preprocessed and analyzed by VP, under the guidance of PS, DC, EB, and MK. An initial draft of the manuscript was written by VP, with input from MK, DC, and EB. All authors provided feedback, contributed to the article, and approved the submitted version.

Funding

This research was funded by the European Research Council (ERC) Starting Grant no. 679010 (STATLEARN), awarded to DC.

Acknowledgments

We thank Yamil Vidal for providing MATLAB and Psychtoolbox scripts for the realization of FPVS (subsequently adapted to the specific MEG setup), and Mara De Rosa for helpful advice on the adaptation of the analysis pipeline from EEG to MEG. We furthermore wish to thank Nick Benikos for his technical support at Macquarie University's KIT-Macquarie Brain Research Laboratory Lab, and Lorenzo Vignali for theoretical and experimental advice. Finally, we thank all the university students, children, and parents who contributed to this research with their participation.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.932952/full#supplementary-material

References

Algeo, J., and Algeo, A. S. (eds.). (1993). *Fifty years among the new words: A dictionary of neologisms 1941–1991*. Cambridge: Cambridge University Press.

Amenta, S., and Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Front. Psychol.* 3:232. doi: 10.3389/fpsyg.2012.00232

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438. doi: 10.1037/a0023851

Baayen, R. H., Piepenbrock, R., and Van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM) Philadelphia*. PA: Linguistic Data Consortium, University of Pennsylvania.

Bentin, S., and Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *Q. J. Exp. Psychol.* 42, 693–711. doi: 10.1080/14640749008401245

Beyersmann, E., Castles, A., and Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *Q. J. Exp. Psychol.* 65, 1306–1326. doi: 10.1080/17470218.2012.656661

Beyersmann, E., and Grainger, J. (2022). "The role of embedded words and morphemes in reading," in *Current issues in the psychology of language*. ed. D. Crepaldi (London: Psychology Press)

Beyersmann, E., Montani, V., Ziegler, J. C., Grainger, J., and Stoianov, I. P. (2021a). The dynamics of reading complex words: evidence from steady-state visual evoked potentials. *Sci. Rep.* 11, 1–14. doi: 10.1038/s41598-021-95292-0

Beyersmann, E., Mousikou, P., Javourey-Drevet, L., Schroeder, S., Ziegler, J. C., and Grainger, J. (2020). Morphological processing across modalities and languages. *Sci. Stud. Read.* 24, 500–519. doi: 10.1080/10888438.2020.1730847

Beyersmann, E., Mousikou, P., Schroeder, S., Javourey-Drevet, L., Ziegler, J. C., and Grainger, J. (2021b). The dynamics of morphological processing in developing readers: A cross-linguistic masked priming study. *J. Exp. Child Psychol.* 208:105140. doi: 10.1016/j.jecp.2021.105140

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., and Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychon. Bull. Rev.* 23, 533–539. doi: 10.3758/s13423-015-0927-z

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Burani, C., Marcolini, S., and Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain Lang.* 81, 568–586. doi: 10.1006/brln.2001.2548

Carlisle, J. F., and Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Read. Res. Q.* 40, 428–449. doi: 10.1598/RRQ.40.4.3

Casalis, S., Quémart, P., and Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudo-word processing: Evidence from a cross-language study. *Front. Psychol.* 6:452. doi: 10.3389/fpsyg.2015.00452

Chetail, F. (2017). What do we do with what we learn? Statistical learning of ortho- graphic regularities impacts written word processing. *Cognition* 163, 103–120. doi: 10.1016/j.cognition.2017.02.015

Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *J. Mem. Lang.* 63, 83–99. doi: 10.1016/j.jml.2010.03.002

Dawson, N., Rastle, K., and Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 645. doi: 10.1037/xlm0000485

Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341. doi: 10.1016/j.tics.2005.05.004

De Rosa, M., Ktori, M., Vidal, Y., Bottini, R., and Crepaldi, D. (2022). Frequency-based neural discrimination in fast periodic visual stimulation. *Cortex* 148, 193–203.

Dzhelyova, M., and Rossion, B. (2014). Supra-additive contribution of shape and surface information to individual face discrimination as revealed by fast periodic visual stimulation. *J. Vis.* 14, 15. doi: 10.1167/14.14.15

Frost, R., Forster, K. I., and Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 829. doi: 10.1037/0278-7393.23.4.829

Gold, B. T., and Rastle, K. (2007). Neural correlates of morphological decomposition during visual word recognition. *J. Cogn. Neurosci.* 19, 1983–1993. doi: 10.1162/jocn.2007.19.12.1983

Grainger, J., and Beyersmann, E. (2017). *Edge-aligned embedded word activation initiates morpho-orthographic segmentation. In Psychology of Learning and Motivation* (Vol. *67*, pp. 285–317). Cambridge: Academic Press.

Grainger, J., and Beyersmann, E. (2021). Effects of lexicality and pseudo-morphological complexity on embedded word priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 47, 518–531. doi: 10.1037/xlm0000878

Grainger, J., and Ziegler, J. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2:54. doi: 10.3389/fpsyg.2011.00054

Hauk, O., Rice, G. E., Volfart, A., Magnabosco, F., Ralph, M. L., and Rossion, B. (2021). Face-selective responses in combined EEG/MEG recordings with fast periodic visual stimulation (FPVS). *NeuroImage* 242:118460. doi: 10.1016/j.neuroimage.2021.118460

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. Available at: https://cran.r-project.org/package=vwr

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., and Parrila, R. (2012). Children's morphological awareness and reading ability. *Read. Writ.* 25, 389–410. doi: 10.1007/s11145-010-9276-5

Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception* 36, 1–16. doi: 10.1068/v070821

Lavric, A., Rastle, K., and Clapp, A. (2011). What do fully visible primes and brain potentials reveal about morphological decomposition? *Psychophysiology* 48, 676–686. doi: 10.1111/j.1469-8986.2010.01125.x

Lelonkiewicz, J. R., Ktori, M., and Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *J. Mem. Lang.* 115:104152. doi: 10.1016/j.jml.2020.104152

Leminen, A., Smolka, E., Dunabeitia, J. A., and Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116, 4–44. doi: 10.1016/j.cortex.2018.08.016

Lochy, A., Van Belle, G., and Rossion, B. (2015). A robust index of lexical representation in the left occipito-temporal cortex as evidenced by EEG responses to fast periodic visual stimulation. *Neuropsychologia* 66, 18–31. doi: 10.1016/j.neuropsychologia.2014.11.007

Lochy, A., Van Reybroeck, M., and Rossion, B. (2016). Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proc. Natl. Acad. Sci.* 113, 8544–8549. doi: 10.1073/pnas.1520366113

Longtin, C. M., Segui, J., and Hallé, P. A. (2003). Morphological priming without morphological relationship. *Lang. Cogn. Process.* 18, 313–334. doi: 10.1080/01690960244000036

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W. D., and Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosop. Trans. Royal Soc.: Biolog. Sci.* 362, 823–836. doi: 10.1098/rstb.2007.2091

MATLAB (2019). *Version 9.6.0.1472908 (R2019a)*. The MathWorks Inc., Natick, MA.

Mousikou, P., Beyersmann, E., Ktori, M., Javourey, L., Crepaldi, D., Ziegler, J. C., et al. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Dev. Sci.* 23, 1–19. doi: 10.1111/desc.12952

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *J. Vis.* 15, 4. doi: 10.1167/15.6.4

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:6869. doi: 10.1155/2011/156869

Quek, G. L., Liu-Shuang, J., Goffaux, V., and Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage* 176, 465–476. doi: 10.1016/j.neuroimage.2018.04.034

Quémart, P., Casalis, S., and Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *J. Exp. Child Psychol.* 109, 478–496. doi: 10.1016/j.jecp.2011.02.008

Rastle, K. (2019). The place of morphology in learning to read in English. *Cortex* 116, 45–54. doi: 10.1016/j.cortex.2018.02.008

Rastle, K., and Davis, M. H. (2003). "Reading morphologically-complex words: Some thoughts from masked priming," in *Masked priming: State of the art*. eds. S. Kinoshita and S. J. Lupker (London: Psychology Press)

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., and Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition. *Lang. Cogn. Process.* 15, 507–537. doi: 10.1080/01690960050119689

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098. doi: 10.3758/BF03196742

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Retter, T. L., and Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia* 91, 9–28. doi: 10.1016/j.neuropsychologia.2016.07.028

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Exp. Brain Res.* 232, 1599–1621. doi: 10.1007/s00221-014-3934-9

Rossion, B., Torfs, K., Jacques, C., and Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *J. Vis.* 15, 18. doi: 10.1167/15.1.18

RStudio Team (2021). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA.

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., and Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behav. Res. Methods* 50, 1568–1580. doi: 10.3758/s13428-017-0981-8

Schiff, R., Raveh, M., and Fighel, A. (2012). The development of the Hebrew mental lexicon: When morphological representations become devoid of their meaning. *Sci. Stud. Read.* 16, 383–403. doi: 10.1080/10888438.2011.571327

Seidenberg, M. S. (1987). "Sublexical structures in visual word recognition: Access units or orthographic redundancy?" in *Attention and Performance 12: The Psychology of Reading*. ed. M. Coltheart (New York: Lawrence Erlbaum Associates, Inc), 245–263.

Stevens, P., and Plaut, D. C. (2022). From decomposition to distributed theories of morphological processing in reading. *Psychon. Bull. Rev.* 1–30. doi: 10.3758/s13423-022-02086-0

Taft, M. (2015). "The nature of lexical representation in visual word recognition," in *The Oxford handbook of reading*. eds. A. Pollatsek and R. Treiman (Oxford: Oxford University Press), 99–113.

Taft, M., and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *J. Verbal Learn. Verbal Behav.* 14, 638–647. doi: 10.1016/S0022-5371(75)80051-X

Taft, M., and Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Lang. Cogn. Process.* 25, 277–296. doi: 10.1080/01690960903043261

Tkačik, G., Prentice, J. S., Victor, J. D., and Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. *Proc. Natl. Acad. Sci.* 107, 18149–18154. doi: 10.1073/pnas.0914916107

Van Casteren, M., and Davis, M. H. (2006). Mix, a program for pseudo-randomisation. *Behav. Res. Methods* 38, 584–589. doi: 10.3758/BF03193889

Van Heuven, W. J., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Q. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521

Vidal, Y., Viviani, E., Zoccolan, D., and Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Curr. Biol.* 31, 1261–1267. doi: 10.1016/j.cub.2020.12.017

Whiting, C., Shtyrov, Y., and Marslen-Wilson, W. (2015). Real-time functional architecture of visual word recognition. *J. Cogn. Neurosci.* 27, 246–265. doi: 10.1162/jocn_a_00699

Xu, J., and Taft, M. (2014). Solely soles: Inter-lemma competition in inflected word recognition. *J. Mem. Lang.* 76, 127–140. doi: 10.1016/j.jml.2014.06.008

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979. doi: 10.3758/PBR.15.5.971

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 23 September 2022
DOI 10.3389/fpsyg.2022.920513

EDITED BY
Stefanie Peykarjou,
Heidelberg University,
Germany

REVIEWED BY
Anna Bánki,
University of Vienna,
Austria
Daniela Traficante,
Catholic University of the Sacred Heart,
Italy
Martin I. Antov,
Osnabrück University,
Germany

*CORRESPONDENCE
Ulrike Frischen
ulrike.frischen@uol.de
Gudrun Schwarzer
gudrun.schwarzer@psychol.uni-giessen.de

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 14 April 2022
ACCEPTED 02 September 2022
PUBLISHED 23 September 2022

CITATION
Frischen U, Degé F and Schwarzer G (2022)
The relation between rhythm processing
and cognitive abilities during child
development: The role of prediction.
Front. Psychol. 13:920513.
doi: 10.3389/fpsyg.2022.920513

The relation between rhythm processing and cognitive abilities during child development: The role of prediction

Ulrike Frischen[1]*, Franziska Degé[2] and Gudrun Schwarzer[3]*

[1]Department of Music, University of Oldenburg, Oldenburg, Germany, [2]Music Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany, [3]Department of Developmental Psychology, Faculty of Psychology and Sports Science, University of Giessen, Giessen, Germany

Rhythm and meter are central elements of music. From the very beginning, children are responsive to rhythms and acquire increasingly complex rhythmic skills over the course of development. Previous research has shown that the processing of musical rhythm is not only related to children's music-specific responses but also to their cognitive abilities outside the domain of music. However, despite a lot of research on that topic, the connections and underlying mechanisms involved in such relation are still unclear in some respects. In this article, we aim at analyzing the relation between rhythmic and cognitive-motor abilities during childhood and at providing a new hypothesis about this relation. We consider whether predictive processing may be involved in the relation between rhythmic and various cognitive abilities and hypothesize that prediction as a cross-domain process is a central mechanism building a bridge between rhythm processing and cognitive-motor abilities. Further empirical studies focusing on rhythm processing and cognitive-motor abilities are needed to precisely investigate the links between rhythmic, predictive, and cognitive processes.

KEYWORDS
rhythm processing, prediction, cognitive abilities, musical rhythm development, beat processing

Introduction

Rhythm is a central component of music. The ability to perceive and produce rhythm enables people to make music. The first signs of these rhythmic abilities appear already in infancy (Winkler et al., 2009) and develop into adulthood (Thompson et al., 2015). In recent years, interest in research on rhythm processing, its development, and the connection to cognition has increased. Researchers found positive associations between rhythmic abilities and different cognitive abilities such as language, motor function, or executive functions (e.g., Anvari et al., 2002; Flaugnacco et al., 2014; Degé et al., 2015; Lesiuk, 2015; Slater et al., 2018; Trainor et al., 2018) with some even suggesting potential causal links (e.g., Moritz et al., 2013; Flaugnacco et al., 2015; Frischen et al., 2019; Lê et al., 2020). For example, it has been shown that music training that is highly based on rhythm processing cannot only improve rhythmic, but also benefit language abilities in typical developing children and children with developmental dyslexia (Moritz et al., 2013; Flaugnacco et al., 2015). Additionally, it has been revealed that rhythm-based music training can improve executive functions in preschoolers (Frischen et al., 2019; Williams and Berthelsen, 2019). However, although a lot of research has already been done on the association between rhythmic abilities and cognitive-motor abilities, the precise connections and underlying mechanisms are still unclear in some respects. To gain a better understanding of these relations, we will identify processes that are related to both rhythm processing and cognitive-motor processing. We propose that prediction could be such a process, building a bridge between rhythmic abilities and cognitive-motor abilities. Predictive processes are fundamental for human cognition and highly relevant for early cognitive development (Nagai, 2019). From the very beginning of life, children strive to identify regularities and contingencies in their physical and social environment based on which they make predictions (Köster et al., 2020). In the context of musical rhythm, predictive processes mean the anticipation of temporally ordered sound events which is usually formed with regard to the meter and the rhythm of a tone sequence. As predictive processes are not only of great importance in the context of rhythm processing, but rather play a significant role in a variety of cognitive processes, they can be considered as cross-domain processes that have the potential to connect rhythm and general cognitive-motor processing. Therefore, we hypothesize that prediction is particularly well suited to explain the link between rhythmic and cognitive abilities. In the following, we will focus on the development of rhythm processing, its link to cognitive and motor processing and we will describe the extent to which our hypothesis that predictive processing is partly involved in the relation between rhythmic and various cognitive abilities is in line with extant research findings.

Rhythm processing and its development

Rhythm and meter are key components of music. According to Schulkind (1999) musical rhythm is defined as a serial pattern of variable tone durations in a melody that unfolds within a timing framework of a recurring pattern of beats, the meter. Meter organizes a series of beats into recurring patterns of stressed and unstressed beats. The beat (or tactus) is referred as one count of the meter and the most natural rate at which a listener might tap or clap to music. If you change the tempo of a rhythm (play a rhythm faster or slower), the relative proportions between the individual beats remain the same. Tempo is defined as the pace of music, or the rate at which beats unfold over time (McAuley, 2010).

In the literature we find several abilities that are described as rhythmic abilities, which are divided into rhythm perception and production abilities (Thackray, 1969; Bouwer et al., 2021). As perception abilities we count the ability to discriminate between rhythms and tempi, to memorize rhythms, or to detect the beat in a rhythmic sequence. Rhythm production abilities include tapping to a beat, reproduce rhythmic patterns, or movement to music (e.g., Thackray, 1969; Tierney and Kraus, 2015; Bouwer et al., 2021). Although studies found correlations between different rhythmic abilities, including correlations between rhythm perception and production abilities in adults (e.g., Keele et al., 1985; Fujii and Schlaug, 2013; Tierney and Kraus, 2015), no unitary rhythm ability could be identified. However, according to Tierney and Kraus (2015), rhythm memory and beat-based processing seem to be two central abilities that are dissociable from each other in adult samples. While rhythm memory is more dependent on auditory memory, beat-based processing requires the ability to detect regularities within a temporal sequence, which can be related to an underlying meter [Ozernov-Palchik and Patel, 2018; Note: Beat-based processing is not a uniform term. For the term "beat-based" we also find terms like "metrical" (Essens and Povel, 1985) or "metric simple" (e.g., Grahn and Brett, 2009) in the literature; in this article we stick to the term "beat-based processing"]. Here, we focus on rhythm processing in terms of beat-based processing as predictive processes can be assumed to be a central component of this ability, what we will illustrate in the following sections. Moreover, previous studies already indicated that this could be a bridging principle in the relation to cognitive-motor-abilities (Ozernov-Palchik and Patel, 2018).

Similar to the question of how many and what rhythmic abilities exactly exist, the development of rhythmic abilities in humans is not yet fully explored so that a detailed consideration of the development of rhythmic skills has not yet been done. However, there is a lot of literature on certain time periods of rhythmic development with a focus on infancy and childhood up to about 7 years of age. In infancy, motor control is still under development. Therefore, rhythm perception abilities are ahead of rhythm production abilities. At the perceptual level, first signs of rhythmic ability occur very early. Evidence has been found indicating that few days old newborns already show specific responses in the electroencephalogram (EEG) to changes in sound-durations, which are similar to those EEG-responses in adults suggesting that infants are sensitive to changes in sound durations (Kushnerenko et al., 2001). Moreover, it has been revealed that newborns can detect repetitive sound patterns (Stefanics et al., 2007) and that they are sensitive to omissions of the downbeat within presented sound sequences (Winkler et al., 2009). These findings indicate that the predispositions for more complex rhythmic abilities are already present in newborns. Further studies have shown that those precursor rhythmic abilities continue to develop within infancy. In a study by Phillips-Silver and Trainor (2005) it was demonstrated that infants' encoding of a meter could be influenced by moving them, indicating that they can already distinguish between a double and a triple meter. Moreover, a recent EEG-study by Flaten et al. (2022) showed that 6-month-old infants can extract information about meter from auditorily presented stimuli and transfer it to an auditorily presented ambiguous rhythm. Additionally, results revealed that infants with musically experienced parents showed larger EEG amplitudes indicating that parental musical background influences music perception in infants. Similar results have been reported in a study by Cirelli et al. (2016) with 7- and 15-month-old infants. Their results revealed that infants' music experience and parents' musical background influenced EEG amplitudes corresponding to beat and meter. In sum, results of both studies suggest that already in early development individual differences in music experience and parental musical background affect rhythm processing and music processing in general.

As already mentioned, rhythm production abilities develop later from early childhood on. A study with 5- to 37-month-old children indicated that children's spontaneous motor tempo (SMT) becomes faster with age, because they get better in the ability to make repeated, targeted movements. Additionally, it seems that the SMT is related to the step rate of their parents, suggesting that early rhythm is set by the vestibular stimulation from parental carrying (Rocha et al., 2021). In contrast to freely producing rhythms such as measured by the SMT, synchronizing movements to a rhythm is a more complex skill: During infancy, children are not able to fully synchronize their movements to a musical beat, although they move their arms and legs in response to music (Fujii et al., 2014) and can adapt their movements to tempo changes to some extent (Zentner and Eerola, 2010). A more recent study on children's drumming revealed that synchronization abilities start to develop around the age of 2 years, but only in a tempo which is close to the children's own spontaneous drumming tempo (Yu and Myowa, 2021). The ability to adapt to a slower tempo was only shown in older children from 30 months on. Overall, the study showed that synchronization abilities improved from 18 over 30 to 42 months of age. These results fit well to the results of previous studies, showing that synchronization abilities are still under development in early childhood (e.g., Drake et al., 2000; Provasi and Bobin-Bègue, 2003; McAuley et al., 2006; Kirschner and Tomasello, 2009). For example, the study of McAuley et al. (2006) revealed that children around the age of 2.5 years can manage to tap in synchrony with an isochronous (temporally equidistant) beat, which is close to their own spontaneous drumming tempo, but fail when they are asked to adapt their tapping to different tempi. In contrast, 4-year-old children performed significantly better in adjusting their tapping to different tempi. Further results of this study found in the appendix and reported by Repp and Su (2013) show that 4- and 5-year-olds are still not good at synchronization while 7-year-olds perform almost on an adult level. Similarly, Drake et al. (2000) reported that 4-year-olds already show the ability to synchronize to different tempi and stimuli but that this ability improves further with age.

Taken together, findings from the literature indicate that rhythmic abilities occur already early in life and develop during childhood. While newborn infants are already sensitive to sound durations and rhythmic variations, rhythm production abilities, and especially the ability to synchronize to a beat, which highly relies on beat-processing occurs earliest in young childhood at around 2 years and improves with age.

Rhythmic abilities and cognitive-motor abilities

Previous research showed that rhythmic abilities do not only develop rapidly during early development, but that they are also connected to the development of non-musical abilities such as in the cognitive and motor domain. In this section, we will address this research and focus on the relation between rhythmic abilities and three specific areas within cognitive-motor abilities: language ability, motor skills and executive functions. These relations serve as examples based on which we will develop our hypothesis on predictive processes as a bridge between rhythmic abilities and those three cognitive-motor areas. In the literature we find mainly correlative studies suggesting positive associations between measures of rhythmic abilities and cognitive-motor abilities: It has been shown that rhythmic abilities are associated with language related abilities such as reading, or precursors of reading ability (Anvari et al., 2002; Thomson et al., 2006; Thomson and Goswami, 2008; Huss et al., 2011; Moritz et al., 2013; Tierney and Kraus, 2013; Flaugnacco et al., 2014; Degé et al., 2015; Tierney et al., 2021; Bégel et al., 2022; for a review see, e.g., Ladányi et al., 2020), executive functions (Tierney and Kraus, 2013; Lesiuk, 2015; Slater et al., 2018), and motor abilities. The latter indicated by the finding that populations with motor disabilities show poorer rhythmic abilities than typical participants (Whitall et al., 2008; Grahn and Brett, 2009; Roche et al., 2016; Trainor et al., 2018). Moreover, there are also training studies suggesting causal relationships between rhythmic activities and language (Flaugnacco et al., 2015; Lê et al., 2020), rhythmic activities and executive functions (e.g., Frischen et al., 2019; Williams and Berthelsen, 2019), and rhythmic activities and gait in Parkinson's disease (e.g., Thaut et al., 1996; for a review see Nombela et al., 2013). Theoretically, it would be plausible that such effects could also occur in the other direction, from the training of cognitive-motor skills to rhythmic skills. However, we do not know of any studies on this topic so far.

In the following, we will describe some of these studies on the relation between rhythmic abilities and the three non-musical areas of language ability, motor skills and executive functions in detail to analyse and show that one particular aspect of rhythmical experience, beat-based processing, is particularly related to these non-musical abilities.

A study conducted with infants underlined the strong link between musical rhythm and language processing. In a training study, Zhao and Kuhl (2016) presented short waltz-like musical pieces to 9-month-old infants in 12 training sessions to familiarize them to the temporal structure of a triple meter. The infants were tested before and after the training sessions with respect to their neural responses to violations of this temporal structure using musical and non-native syllable-like sequences. The authors found that infants in the music-intervention group showed a larger neural response to a violation of the triple meter in musical sequences and, crucially, also in the speech-like stimuli compared to infants in a control group who did not participate in the music intervention. The detection of the temporal structure of the musical pieces has thus generalized to the detection of a comparable sequence of speech-like stimuli.

Regarding older children, the study of Degé et al. (2015) found out that rhythm perception, and rhythm production tasks are related to measures of phonological awareness in pre-schoolers. Phonological awareness refers to the ability to recognise, analyse and manipulate sounds in oral language (Stahl and Murray, 1994; Lonigan, 2006) and is an important predictor of reading and writing abilities (e.g., Marx, 2007). In a correlational approach, they tested different measures of musical ability and phonological awareness while controlling for IQ and socioeconomic status in pre-schoolers. First, results showed that several musical abilities were related to measures of phonological awareness. However, after controlling for the first type of error, only the music production tasks related to rhythmic skills still showed a significant association with measures of phonological awareness. Flaugnacco et al. (2014) reported similar results in a sample of children diagnosed with dyslexia. Their findings showed that measures of rhythmic abilities (tapping to a metronome, rhythm reproduction, and meter perception) are related to reading skills. Additionally, in a following training study it has been shown that a rhythm intervention can improve reading abilities in children with dyslexia (Flaugnacco et al., 2015). Ozernov-Palchik et al. (2018) investigated the previously reported relation between rhythmic abilities and literacy skills in more detail and pursued the question whether there is a specific feature in rhythm that is especially linked to literacy skills. A specific characteristic of most rhythms is that they have a regular underlying structure of recurring beats, what they term as "beat-based." Beat-based rhythms can be grouped into equal time units whereas in non-beat-based rhythms the beats cannot be grouped in equal time units. Since there is a regular underlying structure in beat-based rhythms, the occurring beats can be anticipated. In a correlational approach, the authors tested beat-based as well as non-beat-based rhythm perception in a sample of 5- to 6-year-old children and found relations between both the beat-based and the non-beat-based rhythm task and different early literacy skills. However, they found especially the beat-based task being a unique predictor of one measure of early literacy skills (letter-sound knowledge) above general cognitive abilities, phonological awareness, and non-beat-based processing.

Taken together, research on rhythmic abilities and language abilities shows significant associations that occur already in infancy and could be demonstrated in typically developing children and children with dyslexia even with a first indication of a causal link. Additionally, first evidence emerges that especially beat-based processing plays a significant role in the relation between rhythm abilities and language abilities.

Regarding the association between rhythm abilities and motor function, studies revealed that children and adults with motoric disorders show difficulties in rhythm processing as indicated through studies with children with Developmental Coordination Disorder (DCD; e.g., Whitall et al., 2008; Roche et al., 2016) as well as adults affected by Parkinson's disease (PD, a neurodegenerative condition including symptoms of problems in walking and gait; Knutsson, 1972; Grahn and Brett, 2009; Nombela et al., 2013). Whitall et al. (2008) showed that children with a diagnosis of DCD have poor rhythmic skills. In their study they compared children diagnosed with DCD with gender and age-matched typical controls as well as with typical adults in a finger-tapping paradigm. The results demonstrated that children with DCD are broadly able to match their tapping to the different metronome tempi. However, children with DCD have particular problems to match the beat in slow tempi. Moreover, the tapping of the children with DCD was more variable compared to age-matched controls and adults. While adults mostly show to be a bit before the beat, children tend to tap behind the beat. The children with DCD, however, did not show any consistent relation with the beat. These results indicate that children with DCD have problems in identifying and anticipating the single beats within a rhythm sequence, so that they seem to have difficulties in beat-based processing. Similar results have been revealed from studies with older people diagnosed with PD indicating that people with PD have poorer rhythmic abilities compared to typical controls (e.g., Grahn and Brett, 2009; Hsu et al., 2022). For example, Grahn and Brett (2009) tested older people affected by PD and typical controls in a rhythm discrimination task similar to the task described in Ozernov-Palchik et al. (2018) consisting of two conditions: a beat-based and a non-beat-based condition. The results revealed that people diagnosed with PD do not show differences in both conditions of the task, while typical controls showed a better performance in the beat-based task compared to the non- beat-based task. Moreover, older adults affected by PD performed worse in the beat-based task compared to typical controls. The results indicate that people affected by PD also have problems in the detection and anticipation of the beat structure indicating problems in beat-based processing. In sum, the reported studies on the relation between rhythmic abilities and motor function also indicate that beat-based processing seems to be a fundamental ability which is linked to motor skills.

In addition to associations between rhythmic abilities and more specific non-musical abilities such as language and motor function, associations between rhythmic abilities and general cognitive abilities, such as executive functions, have also been revealed: Frischen et al. (2019) found out that a rhythm-based music intervention can enhance inhibition (a measure of executive functions) in pre-schoolers. In a randomized controlled training study, children from different kindergartens received a 6-months rhythm-based music intervention, a pitch-based music intervention, or a sports intervention three times a week. Before and after the intervention children's executive functions were assessed. The results showed that only the rhythm group improved significantly from pre- to post-test in inhibition, suggesting that rhythm training can improve inhibition skills in young childhood. Further measures of executive functions (working memory, flexibility) were not significantly affected by any training. While this study suggests that rhythmic practice promotes inhibition, it leaves open which rhythmic abilities exactly were trained and how they are linked to inhibition. Related to this issue the study of Tierney and Kraus (2013) indicated that rhythm production ability, measured by tapping to a beat, was positively correlated with inhibition in adolescents. In this study, tapping performance was assessed in two conditions: tapping to the beat (paced condition) and tapping in silence (unpaced condition). Inhibition was measured with a test including an auditory and a visual condition. The results showed that tapping to the beat (paced condition) was positively correlated with both auditory as well as visual inhibition. However, the unpaced tapping condition was not correlated to inhibition in any way. Moreover, it has been found that less tapping variability was associated with better performance in the auditory and the visual inhibition task. These results have been confirmed in a following study with a similar design revealing that less variable drumming was positively associated with inhibition in young adults (Slater et al., 2018). Since tapping or drumming to a beat is highly dependent on beat-based processing, it seems that especially this rhythmic ability is of great importance in the relation to inhibition as one measure of executive functions, which is in line with the previously reported findings on language abilities and motor function. However, since the reported studies did not report on relations between beat-based processing and further executive functions apart from inhibition, it remains unclear whether there is a specific relation solely to inhibition or whether we can assume a relation between beat-based processing and executive functions in general.

Taken together, there is a substantial amount of literature suggesting correlations between rhythmic and cognitive-motor skills. In addition, there are individual studies that suggest causal relationships. Interestingly, studies on the relation between rhythmic abilities and non-musical abilities from all three reported areas suggest that the ability to extract regularities from rhythms, such as in beat-based processing (e.g., Ozernov-Palchik et al., 2018) is a highly relevant aspect of rhythmic experience that is connected to cognitive-motor abilities.

Processes involved in the association between rhythmic abilities and cognitive-motor abilities

Despite the reported findings indicating a close association between beat-based rhythm processing and cognitive-motor abilities, it is still unanswered how these rhythm processes possibly exert an effect on the above-mentioned non-musical abilities. We propose that one possible process involved in this relation could be the process of making predictions, which is a key process in beat-based rhythm processing and cognitive-motor processing as well.

The ability to recognize a beat-based rhythm requires that an internal representation or model of the rhythm with the corresponding beats has been generated. Such an internal representation is necessary as it provides the basis for anticipating the upcoming beats. Thus, the formation of expectations and the reduction of expectation errors represent fundamental processes in the perception and recognition of a rhythm. The rhythm in a piece of music "plays" with predictions to generate and convey tension and tension release through the violation and fulfilment of expectations in musical sequences. The predictive character of a rhythm also becomes particularly obvious when a rhythm is to be played or clapped synchronously. Imagine this process would be a reactive instead of anticipatory process, then one would always be a little behind the beat and a precise timing would not be possible. Thus, synchronous production of a rhythm is only possible because we are anticipating the beat we are about to produce. Therefore, prediction can be understood as an inherent component of perceiving and producing a rhythm.

However, prediction is not only a key process of beat-based rhythm processing but also a key process of human perception and cognition in general. Continuously making predictions is indispensable and vital for survival. Making predictions can be considered as a working principle that aims to constantly adapt mental representations of the environment to its requirements by making predictions and learning from prediction errors. This in turn leads to a minimizing of prediction errors and to an increase of successful interactions with an ever changing environment (see Köster et al., 2020). The origin of research on predictive processes can be traced back to basic motor learning principles as already described by Helmholtz (1867) or see (e.g., Schubotz, 2015) which have recently been transferred as a basic learning principle of the human brain (e.g., Friston, 2005, 2010). Viewed from this theoretical framework, predictive processes occur on various, hierarchically organized levels: from basic, often automatic motor responses to controlled, higher reasoning. Feedback about prediction errors is sent back to the levels in the hierarchy that are involved in the prediction process to adjust existing predictions and internal models. This can then lead to changes on the motor up to the cognitive level (see Köster et al., 2020). Thus, predictive processing is not a single cognitive ability, but a cross-domain working principle that encompasses perception, motor skills (action), and cognition.

Regarding predictive processes during development, previous research has mainly focused on such processes outside the domain of music or rhythm processing. Here, it has been demonstrated that from early infancy on, children have the strong motivation to detect regularities in their environment using statistical information in stimulus sequences from which they form predictions (Bulf et al., 2011). Such statistical information, e.g., specific frequencies, redundancies, or transitional probabilities of stimuli in a sequence can be found in almost all natural events in our auditory and visual environment and can be perceived without any instructions or feedback. Perceiving statistical information is considered a mandatory process of human information processing (e.g., Gómez, 2017) which allows the detection of patterns in structured inputs that can also serve as a basis for predicting subsequent events. Evidence of this mechanism was provided in the seminal work of Saffran et al. (1996) on the statistical learning of continuous speech patterns in 8-month-old infants, which led to an explosion of research on this topic. This body of research showed that infants and children recognize statistical information in a variety of sensory domains, not only in language and other auditory stimuli, but also in sequences of visual and haptic stimuli (e.g., Kirkham et al., 2007; Fassbender et al., 2014; Aslin, 2017).

Previous research on non-musical predictive processes during development is also based on the idea that children's predictions are more and more controlled by their sensorimotor experiences. This basic idea originates in Piaget's work on children's sensorimotor development (Piaget, 1952), and has been convincingly confirmed by Nagai (2019) work on predictive learning. Nagai proposes two modules which represent the architecture of infant predictive learning. The first module comprises the sensorimotor system which has the role of executing actions, interacting with the environment, and recording the resulting sensory feedback from the environment. The second module represents the so-called predictor, which comprises the internal model of the sensorimotor system. The aim of the predictor is to accurately simulate the sensorimotor system by learning to minimize the so-called predictive error. Nagai assumes that an infant's predictor needs to develop and constantly improves with increasing sensorimotor experience which is in agreement with the conception of so-called forward models of motor control in children and adults. These models also emphasize the extremely close connection between motor performance and prediction. Forward models are used by the Central Nervous System (CNS) to internally simulate the behavior of the motor system in planning, control and learning (Wolpert and Miall, 1996). When a motor signal from the CNS is sent to the periphery, a copy of this motor outflow (i.e., reference copy) is generated. This reference copy inputs to the internal model which can estimate the sensory consequences of the motor command, thus generating the predicted sensory feedback. This forward mechanism is used to anticipate the sensory effects of movement. Thus, the sensory consequences of self-generated movements can be accurately predicted. This mechanism clearly shows the high extent to which motor behavior and prediction are interwoven with each other. In our own research we could demonstrate how increasingly correct predictions in infancy correlate with the increase in infants' motor experience (e.g., Schwarzer, 2014). The prediction of visual–spatial object relations was improved in infants with advanced crawling and manual object exploration skills, compared to infants with low motor skills. In particular, our results suggest that infants with different types of locomotion and manual object exploration experiences differ in their visual processes based on which they generate their visual–spatial predictions (Gerhard-Samunda et al., 2021; Kelch et al., 2021). It can therefore be stated that the developing motor system acts as a control mechanism which promotes correct predictions or reduces prediction errors. Motor experiences allow children to identify regularities in their environment based on which they generate internal models of their environment from which they form predictions leading to increasingly lower prediction errors.

Another line of research on the development of general predictive processes focuses on the idea that children base their predictions on their increasing prior knowledge. Stahl and Feigenson (2015), for example, demonstrated that 11-month-old infants responded to a violation of their expectations regarding their physical core knowledge, and thus showed that the infants had made a particular prediction based on that knowledge. A study by Senju et al. (2011) provided evidence that children's predictions based on their own theory of mind enable them to attribute mental states to others. Thus, it is obvious that all the acquired knowledge of children forms an essential basis for their predictions. In addition, especially in older children, metacognitive knowledge, knowledge about one's own cognitive processes, can also serve as a basis for improving predictions. It can be assumed that applying predictions to one' s own thinking in the sense of comparing learning goals to what has been achieved can improve higher-level cognitive outcomes. It is thought that predictive internal representations of the future are constantly compared with the actual perceived outcome of internal mental and external events. In this respect, making predictions allows to learn from previous experiences, a process that can be applied to various domains.

Overall, it can be summarized that children's general predictive processes continuously improve with age and are mainly based on the detection of regularities in terms of statistical information, sensorimotor experiences, and acquired knowledge from which they generate their predictions. This is in line with Köster et al. (2020) who argued that a predictive-processing framework may provide a unifying umbrella of these at first sight unrelated cognitive processes. They considered prediction of future events as a general, early learning goal which is coupled with the ongoing motivation to reduce experienced uncertainty and to extract predictive structure from physical and social events. So far, only little research exists on the development of predictive processes in rhythm processing. However, we assume that the regularities from which predictions are formed in the course of general cognitive development can also be used for the formation of predictions in the processing of rhythms. For example, Trainor et al. (2003) and Trainor (2012) could show that even infants detect deviations in regular sequences of tone lengths and thereby showed their recognition of the temporal statistic within a rhythmical stimulus sequence. Interestingly, Markova et al. (2019) provided evidence that infants can detect rhythmical information in social interactions such as affective touch or singing with adults. They assume that entrainment (the process when neural oscillations couple to an external rhythm) to these social rhythms underlies the formation of interpersonal synchrony and thus stimulates reciprocal interactions between infants and their caregivers (see also Pereira et al., 2019; Nguyen et al., 2021). With respect to the role of sensorimotor experiences in prediction processing, Phillips-Silver and Trainor (2005) impressively demonstrated that bouncing movements influenced the encoding of meters in infants indicating a close connection between sensorimotor stimulation and temporal processing. Regarding the impact of prior knowledge on rhythm-based predictions, Vuust and Witek (2014) showed that adults' rhythmic predictions are inferred from their previous musical experience and described that the processing system is always in a relation between bottom-up and top-down processes. They provided evidence that, for example, during syncopation – a rhythmic structure that violates metric expectations – the listener's previous musical training determined the accuracy of the participant's predictions.

Thus, we assume that children apply their general drive to make predictions about rhythmic events as well, using similar cognitive processes as they do for events outside the music domain and hypothesize that predictive processing in the general cognitive domain and in the domain of processing rhythms develops along similar types of regularities. We also presume that predictive processes from the rhythmic and cognitive-motor domains can influence each other, as the basic striving to make predictions can be similarly manifested in both domains. Nevertheless, it could be speculated that at least for children relatively simple rhythms or meters are one of the best examples by which predictions and their confirmation can be experienced, which is why they could have a unique, prediction-stimulating effect.

The role of predictive processes in the relation between rhythm and cognitive-motor processing

As predictive processes are crucially involved in the domain of beat-based rhythm processing and cognitive and motor processing, we suppose that predictive processes could serve as mechanisms which could partly explain the association between rhythmic processing and non-musical cognitive-motor abilities. We believe that predictive processes in rhythm processing can stimulate similar processes in the cognitive-motor domain and thereby have the potential to build a bridge between the domains. In the following, we will describe such a potential bridge in more detail using findings from some of the previously mentioned studies. For example, the study by Zhao and Kuhl (2016) demonstrated that infants detected the metrical statistic in a musical piece and transferred it to the same metrical statistics, now however presented in speech-like stimuli and were able to make predictions on such a statistical regularity. These results demonstrate that already in infancy predictive processes based on the recognition of statistical regularities can be used from one into another domain. Moreover, we assume that prediction might be the process that builds the bridge between rhythm processing and linguistic abilities. In a study with pre-schoolers, Ozernov-Palchik et al. (2018) found out that especially the processing of beat-based rhythms is linked to precursors of reading abilities. Interestingly, Ozernov-Palchik and Patel (2018) describe that prediction is involved in beat-based processing and how this is linked to literacy abilities. They point out that beat-based rhythms consist of recurring temporal statistics that can be predicted and that also language, or literacy skills in specific, are based on predictive processes. While in beat-based rhythms, the process of prediction lies specifically in being able to predict future musical rhythmic events based on the rhythmic statistics that are emitted, in the case of literacy skills, the prediction process lies more in being able to make predictions about future linguistic material on the basis of the linguistic structure (e.g., phonological structure, syntactic structure). For example, through statistical learning, children gain knowledge about frequently occurring phonetic combinations and can make predictions about how words are put together from phonemes. The same applies to the translation of phonemes into graphemes: Through statistical learning, children gain knowledge about how phonemes are often transferred into graphemes. Based on this knowledge, they can make predictions about how words are going to be written down. Studies found that in children with developmental dyslexia predictive rhythm processing is disturbed, which is not only reflected in poor rhythm perception and production ability (e.g., Flaugnacco et al., 2014; Bégel et al., 2022) but also on a neural level: Children with developmental dyslexia show atypical neural rhythmic entrainment during beat perception and production (Colling et al., 2017). Thus, the results of the studies with typical developing children and with children with developmental dyslexia are in line with our hypothesis that prediction may be involved in the rhythmic-language relation and may build a bridge between the two abilities.

Also, studies that focused on the relation between rhythmic processing and motor abilities are in line with the idea that prediction could be a bridging process: In the previous section, we already explained the strong link between the development of the motor system and predictive processing. This strong link becomes also evident through the above-mentioned example of adults and children living with motor disabilities such as in DCD or PD. As reported above children with DCD and older adults suffering from PD both have particular problems in processing beat-based rhythms (Grahn and Brett, 2009; Roche et al., 2016), which – as already explained – heavily rely on predictive processes and especially when a rhythmic motor skill is required on generating effective forward models. For example, the study of Whitall et al. (2008) revealed that children with DCD have poorer tapping performance compared to typical children and adults. The finding that the tapping of children with DCD is more variable without a consistent relationship to the beat indicate that impaired forward models with correspondingly impaired prediction processes account for the more variable tapping behavior. Predictive processes are especially essential for motor control. Without an internal model of a planned motor action and the prediction of the outcome of a movement, it is not possible to adapt, or correct rapidly this movement sequence, if required (e.g., to correct tapping behavior when it is not perfectly matched to the beat, or when the beat is changing). In fact, it has already been proposed that children with DCD have problems in motor control due to impaired forward models and therefore impaired prediction of motor sequences. Interestingly, it has been found, that these impaired predictive processes in DCD are not only found in relation to motor function, but also in relation to other cognitive domains (Opitz et al., 2020). This finding supports the idea that predictive processing can function as a cross-domain mechanism that is relevant for several cognitive-motor tasks and can potentially provide a link between rhythm processing and cognitive-motor processing. Regarding children with DCD, it can be assumed that the weak tapping performance can be explained through a general impairment of predictive processes that influence both, the exact anticipation of the beats and the execution and adaptation of movements.

Lastly, predictive processes could serve as a linking process in the relation between rhythmic processing and inhibition (as one measure for executive functions). As reported above, the results by Frischen et al. (2019) demonstrated an enhancement in children's inhibition skills after a rhythm training intervention and the results of Tierney and Kraus (2013) showed associations between beat-based processing and inhibition. As already mentioned, especially metacognitive knowledge or processes about one's own cognitive functioning can also serve as a basis from which predictions can be made. It can be assumed that the predictive internal representations of the to be solved rhythm tasks (tapping/ clapping/ drumming along different beats) in the studies of Frischen et al. (2019) and Tierney and Kraus (2013) would stimulate other cognitive processes linked with prediction such as attention and cognitive control (including inhibition). These processes allow participants to constantly compare the actual perceived outcomes with the intended outcomes. Thus, we assume that predictive processes involved in rhythm processing do also stimulate other metacognitive processes which in turn have a facilitating effect on executive functions such as inhibition. This assumption goes well with the finding that children and adults with Attention Deficit Hyperactivity Disorder (ADHD) have problems in synchronizing movements to a beat (Puyjarinet et al., 2017). Since ADHD is associated with poor executive functions, this finding could probably highlight the association between predictive rhythm abilities and executive functions.

Taken together, these studies give a first indication that predictive processes could play a key role in the association between rhythm processing and cognitive-motor abilities. The fact that predictive processes are involved in various abilities across domains points in the direction that this process could serve as a cross-domain mechanism explaining the link between rhythm processing and cognitive-motor abilities. Moreover, the reported studies suggest that predictive processes in general play a key role during early learning and development as it is suggested by several researchers (e.g., Nagai, 2019; Köster et al., 2020). However, we are aware that our assumption is based on only a few studies, and that this idea should be investigated by empirical studies specifically designed for this purpose. Such studies need to assess rhythm processing as such, so that it is controlled to which extent predictive processes are stimulated. This component is missing in many previous studies. Moreover, the study of populations with atypical developmental (or developmental decline in late adulthood) could give more insights in these associations. Specifically, it could be interesting to further investigate the neural processes associated with predictive rhythm processing as demonstrated in children with dyslexia (Colling et al., 2017). Future results from studies with clinical samples showing atypical (or declining) development could provide information on whether the malfunction is only found in predictive rhythm processing or whether prediction is generally affected across domains (as it was addressed by Opitz et al. (2020) in children with DCD). Additionally, further experimental studies could focus on the causal links and investigate whether rhythm training can benefit predictive processes related to rhythm and further predictive processes in other domains. Also, it would be interesting to evaluate whether this also works the other way around (e.g., in how far a training in predictive processes within another domain also improves predictive processes related to musical rhythm).

Conclusion

With our article we developed and substantiated the hypothesis that predictive processes could be considered as a potential explanation for the link between rhythmic abilities, especially beat-based processing and various cognitive-motor abilities. Because predictive processes are a crucial element of beat-based rhythm processing as well as of cognitive-motor abilities, it is possible that prediction as a cross-domain working principle is a central mechanism to explain the connections between rhythmic abilities and cognitive-motor abilities found in several studies. Our analyses of existing findings on associations between rhythm and language, motor development and executive functions are in line with our assumption. Moreover, results from samples with atypical development indicate that a malfunction in predictive rhythm processing can be associated with significant limitations in cognitive-motor processing. To better investigate these relations, further empirical studies are needed that capture predictive processes while processing rhythms (e.g., through EEG) and cognitive-motor abilities. Randomized controlled studies can give further insights into potential causal relations. Furthermore, future studies could address the question of whether certain rhythms (e.g., highly familiar vs. unfamiliar) or certain meters (e.g., duple meter vs. triple meter) particularly stimulate predictive processes and, for example, examine the extent to which the complexity of a beat-based rhythm matters.

Data availability statement

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

Author contributions

UF, FD, and GS contributed to the conception of the manuscript. UF and GS wrote the first draft of the manuscript.

All authors contributed to the article and approved the submitted version.

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

Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8

Aslin, R. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. Wiley Interdiscip rev. *Cogn. Sci.* 8:e1373. doi: 10.1002/wcs.1373

Bégel, V., Dalla Bella, S., Devignes, Q., Vandenbergue, M., Lemaître, M.-P., and Dellacherie, D. (2022). Rhythm as an independent determinant of developmental dyslexia. *Dev. Psychol.* 58, 339–358. doi: 10.1037/dev0001293

Bouwer, F. L., Nityananda, V., Rouse, A. A., and Ten Cate, C. (2021). Rhythmic abilities in humans and non-human animals: A review and recommendations from a methodological perspective. *Philosop. Transac. Royal Society B: Biolog. Sci.* 376:20200335. doi: 10.1098/rstb.2020.0335

Bulf, H., Johnson, S. P., and Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition* 121, 127–132. doi: 10.1016/j.cognition.2011.06.010

Cirelli, L. K., Spinelli, C., Nozaradan, S., and Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: effects of music background. *Front. Neurosci.* 10, 1–11. doi: 10.3389/fnins.2016.00229

Colling, L. J., Noble, H. L., and Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: an EEG study. *Front. Neurosci.* 11:360. doi: 10.3389/fnins.2017.00360

Degé, F., Kubicek, C., and Schwarzer, G. (2015). Associations between musical abilities and precursors of reading in preschool aged children. *Front. Psychol.* 6, 1–10. doi: 10.3389/fpsyg.2015.01220

Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences. Attunement, referent period, focal attending. *Cognition* 77, 251–288. doi: 10.1016/S0010-0277(00)00106-2

Essens, P. J., and Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Percept. Psychophys.* 37, 1–7. doi: 10.3758/BF03207132

Fassbender, I., Lohaus, A., Thomas, H., Teubert, M., Vierhaus, M., Lamm, B., et al. (2014). African versus Caucasian faces in a visual expectation paradigm: A longitudinal study with German and Cameroonian infants. *J. Cross-Cult. Psychol.* 45, 1273–1287. doi: 10.1177/0022022114537555

Flaten, E., Marshall, S., Dittrich, A., and Trainor, L. (2022). Evidence for imagined meter in infancy as shown by primed neural responses to an ambiguous rhythm. *Eur. J. Neurosci.* 55, 2003–2023. doi: 10.1111/ejn.15671

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., and Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One* 10, 1–17. doi: 10.1371/journal.pone.0138715

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., et al. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Front. Hum. Neurosci.* 8, 1–14. doi: 10.3389/fnhum.2014.00392

Frischen, U., Schwarzer, G., and Degé, F. (2019). Comparing the effects of rhythm-based music training and pitch-based music training on executive functions in preschoolers. *Front. Integr. Neurosci.* 13, 1–11. doi: 10.3389/fnint.2019.00041

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Fujii, S., and Schlaug, G. (2013). The Harvard beat assessment test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Front. Hum. Neurosci.* 7, 1–16. doi: 10.3389/fnhum.2013.0077

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., and Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS One* 9, 1–12. doi: 10.1371/journal.pone.0097680

Gerhard-Samunda, T. M., Jovanovic, B., and Schwarzer, G. (2021). Role of manually-generated visual cues in crawling and non-crawling 9-month-old infants' mental rotation. *Cogn. Dev.* 59:101053. doi: 10.1016/j.cogdev.2021.101053

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philos. Trans. R. Soc. B* 372:20160054. doi: 10.1098/rstb.2016.0054

Grahn, J. A., and Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson' s disease. *Cortex* 45, 54–61. doi: 10.1016/j.cortex.2008.01.005

Helmholtz, H. V. (1867). *Handbuch der physiologischen Optik*. Leipzig, Germany: Leopold Voss.

Hsu, P., Ready, E. A., and Grahn, J. A. (2022). The effects of Parkinson's disease, music training, and dance training on beat perception and production abilities. *PloS One.* 17:e0264587. doi: 10.1371/journal.pone.0264587

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689. doi: 10.1016/j.cortex.2010.07.010

Keele, S. W., Pokorny, R. A., Corcos, M., and Ivry, R. (1985). Do perception and motor production share common timing mechnisms: a correlational analysis. *Acta Psychol.* 60, 173–191. doi: 10.1016/0001-6918(85)90054-X

Kelch, A., Schwarzer, G., Gehb, G., and Jovanovic, B. (2021). How 9-month-old crawling infants profit from visual-manual rotations in a mental rotation task. *Infant Behav. Dev.* 65:101642. doi: 10.1016/j.infbeh.2021.101642

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., and Johnson, S. P. (2007). Location, location, location: development of spatial temporal sequence learning in infancy. *Child Dev.* 78, 1559–1571. doi: 10.1111/j.1467-8624.2007.01083.x

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Knutsson, E. (1972). An analysis of Parkinsonian gait. Brain: a. *J. Neurol.* 95, 475–486. doi: 10.1093/brain/95.3.475

Köster, M., Kayhan, E., Langeloh, M., and Hoehl, S. (2020). Making sense of the world: infant learning from a predictive processing perspective. *Perspect. Psychol. Sci.* 15, 562–571. doi: 10.1177/1745691619895071

Kushnerenko, E., Ceponiene, R., Fellman, V., Huotilainen, M., and Winkler, I. (2001). Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *Neuro Report* 12, 3777–3781. doi: 10.1097/00001756-200112040-00035

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., and Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdiscip. Rev. Cogn. Sci.* 11, e1528–e1532. doi: 10.1002/wcs.1528

Lê, M., Quémart, P., Potocki, A., Gimenes, M., Chesnet, D., and Lambert, E. (2020). Rhythm in the blood: the influence of rhythm skills on literacy development in third graders. *J. Exp. Child Psychol.* 198:104880. doi: 10.1016/j.jecp.2020.104880

Lesiuk, T. (2015). Music perception ability of children with executive function deficits. *Psychol. Music* 43, 530–544. doi: 10.1177/0305735614522681

Lonigan, C. J. (2006). Development, assessment, and promotion of preliteracy skills. *Early Educ. Dev.* 17, 91–114. doi: 10.1207/s15566935eed1701_5

Markova, G., Nguyen, T., and Hoehl, S. (2019). Neurobehavioral interpersonal synchrony in early development: the role of interactional rhythms. *Front. Psychol.* 10:78. doi: 10.3389/fpsyg.2019.02078

Marx, P. (2007). *Lese- und Rechtschreiberwerb*. Paderborn, Germany: Schöningh UTP.

McAuley, J. D. (2010). "Tempo and rhythm" in *Music perception*. eds. M. R. Jones, R. R. Fay and A. N. Popper (Germany: Springer Science + Business Media), 165–199.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *J. Exp. Psychol. Gen.* 135, 348–367. doi: 10.1037/0096-3445.135.3.348

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., and Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Read. Writ.* 26, 739–769. doi: 10.1007/s11145-012-9389-0

Nagai, Y. (2019). Predictive learning: its key role in early cognitive development. *Philos. Trans. R. Soc. B* 374:20180030. doi: 10.1098/rstb.2018.0030

Nguyen, T., Abney, D. H., Salamander, D., Bertenthal, B. I., and Hoehl, S. (2021). Proximity and touch are associated with neural but not physiological synchrony in naturalistic mother-infant interactions. *Neuro Image* 244:118599. doi: 10.1016/j.neuroimage.2021.118599

Nombela, C., Hughes, L. E., Owen, A. M., and Grahn, J. A. (2013). Into the groove: can rhythm influence Parkinson' s disease? *Neurosci. Biobehav. Rev.* 37, 2564–2570. doi: 10.1016/j.neubiorev.2013.08.003

Opitz, B., Brady, D., and Leonard, H. C. (2020). Motor and non-motor sequence prediction is equally affected in children with developmental coordination disorder. *PLoS One* 15, e0232562–e0232519. doi: 10.1371/journal.pone.0232562

Ozernov-Palchik, O., and Patel, A. D. (2018). Musical rhythm and reading development: does beat processing matter? *Ann. N. Y. Acad. Sci.* 1423, 166–175. doi: 10.1111/nyas.13853

Ozernov-Palchik, O., Wolf, M., and Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *J. Exp. Child Psychol.* 167, 354–368. doi: 10.1016/j.jecp.2017.11.009

Pereira, M. R., Barbosa, F., de Haan, M., and Ferreira-Santos, F. (2019). Understanding the development of face and emotion processing under a predictive processing framework. *Dev. Psychol.* 55, 1868–1881. doi: 10.1037/dev0000706

Phillips-Silver, J., and Trainor, L. J. (2005). Psychology: feeling the beat: movement influences infant rhythm perception. *Science* 308:1430. doi: 10.1126/science.1110922

Piaget, J. (1952). *The origins of intelligence in children*. Madison, CT: International Universities press.

Provasi, J., and Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 21/2- and 4-year-old children. *Int. J. Behav. Dev.* 27, 220–231. doi: 10.1080/01650250244000290

Puyjarinet, F., Bégel, V., Lopez, R., Dellacherie, D., and Dalla Bella, S. (2017). Children and adults with attention-deficit/hyperactivity disorder cannot move to the beat. *Sci. Rep.* 7, 11550–11511. doi: 10.1038/s41598-017-11295-w

Repp, B. H., and Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychon. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2

Rocha, S., Southgate, V., and Mareschal, D. (2021). Infant spontaneous motor tempo. *Dev. Sci.* 24, e13032–e13037. doi: 10.1111/desc.13032

Roche, R., Viswanathan, P., Clark, J. E., and Whitall, J. (2016). Children with developmental coordination disorder (DCD) can adapt to perceptible and subliminal rhythm changes but are more variable. *Hum. Mov. Sci.* 50, 19–29. doi: 10.1016/j.humov.2016.09.003

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Schubotz, R. I. (2015). "Prediction and expectation," in *Brain Mapping: An Encyclopedic Reference*. *Vol. 3*. ed. A. W. Toga (Academic Press: Elsevier), 295–302.

Schulkind, M. D. (1999). Long-term memory for temporal structure: evidence from the identification of well-known and novel songs. *Mem. Cogn.* 27, 896–906. doi: 10.3758/BF03198542

Schwarzer, G. (2014). How motor and visual experience shape infants' processing of objects and faces. *Child Dev. Perspect.* 8, 213–217. doi: 10.1111/cdep.12093

Schwarzer, G., and Jovanovic, (2015). "Entwicklungspsychologie der Kindheit" in *Standards Psychologie*. eds. M. Hasselhorn, W. Kunde and S. Schneider (Stuttgart: Kohlhammer).

Senju, A., Southgate, V., Snape, C., Leonard, M., and Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychol. Sci.* 22, 878–880. doi: 10.1177/0956797611411584

Slater, J., Ashley, R., Tierney, A., and Kraus, N. (2018). Got rhythm? Better inhibitory control is linked with more consistent drumming and enhanced neural tracking of the musical beat in adult percussionists and nonpercussionists. *J. Cogn. Neurosci.* 30, 14–24. doi: 10.1162/jocn_a_01189

Stahl, A. E., and Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science* 348, 91–94. doi: 10.1126/science.aaa3799

Stahl, S. A., and Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *J. Educ. Psychol.* 86, 221–234. doi: 10.1037//0022-0663.86.2.221

Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., et al. (2007). Auditory temporal grouping in newborn infants. *Psychophysiology* 44, 697–702. doi: 10.1111/j.1469-8986.2007.00540.x

Thackray, R. (1969). Rhythmic abilities and their measurement. *J. Res. Music. Educ.* 17, 144–148. doi: 10.2307/3344201

Thaut, M. H., McIntosh, G. C., Rice, R. R., Miller, R. A., Rathbun, J., and Brault, J. M. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. *Mov. Disord.* 11, 193–200. doi: 10.1002/mds.870110213

Thompson, E. C., White-Schwoch, T., Tierney, A., and Kraus, N. (2015). Beat synchronization across the lifespan: intersection of development and musical experience. *PLoS One* 10, 1–13. doi: 10.1371/journal.pone.0128839

Thomson, J. M., Fryer, B., Maltby, J., and Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *J. Res. Read.* 29, 334–348. doi: 10.1111/j.1467-9817.2006.00312.x

Thomson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol. Paris* 102, 120–129. doi: 10.1016/j.jphysparis.2008.03.007

Tierney, A., Gomez, J. C., Fedele, O., and Kirkham, N. Z. (2021). Reading ability in children relates to rhythm perception across modalities. *J. Exp. Child Psychol.* 210:105196. doi: 10.1016/j.jecp.2021.105196

Tierney, A., and Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain Lang.* 124, 225–231. doi: 10.1038/jid.2014.371

Tierney, A., and Kraus, N. (2015). Evidence for multiple rhythmic skills. *PLoS One* 10, 1–14. doi: 10.1371/journal.pone.0136645

Trainor, L. J. (2012). Predictive information processing is a fundamental learning mechanism present in early development: evidence from infants. *Int. J. Psychophysiol.* 83, 256–258. doi: 10.1016/j.ijpsycho.2011.12.008

Trainor, L. J., Chang, A., Cairney, J., and Li, Y. C. (2018). Is auditory perceptual timing a core deficit of developmental coordination disorder? *Ann. N. Y. Acad. Sci.* 1423, 30–39. doi: 10.1111/nyas.13701

Trainor, L. J., McFadden, M., Hodgson, L., Darragh, L., Matsos, L., and Sonnadara, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *Int. J. Psychophysiol.* 51, 5–15. doi: 10.1016/S0167-8760(03)00148-X

Vuust, P., and Witek, M. A. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Front. Psychol.* 5:1111. doi: 10.3389/fpsyg.2014.01111

Whitall, J., Chang, T. Y., Horn, C. L., Jung-Potter, J., McMenamin, S., Wilms-Floet, A., et al. (2008). Auditory-motor coupling of bilateral finger tapping in children with and without DCD compared to adults. *Hum. Mov. Sci.* 27, 914–931. doi: 10.1016/j.humov.2007.11.007

Williams, K. E., and Berthelsen, D. (2019). Implementation of a rhythm and movement intervention to support self-regulation skills of preschool-aged children in disadvantaged communities. *Psychol. Music* 47, 800–820. doi: 10.1177/0305735619861433

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2468–2471. doi: 10.1073/pnas.0809035106

Wolpert, D. M., and Miall, R. C. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279. doi: 10.1016/s0893-6080(96)00035-4

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: a study of 18- to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/infa.12403

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5768–5773. doi: 10.1073/pnas.1000121107

Zhao, T. C., and Kuhl, P. K. (2016). "Musical intervention enhances infants' neural processing of temporal structure in music and speech." in *Proceedings of the National Academy of Sciences of the United States of America.* 113, 5212–5217.

Glossary

Beat (Tactus): One count of the meter, the most natural rate at which a listener might tap or clap to music.
Beat-based processing: The processing of beat-based stimuli; beat-based stimuli show a regular structure of reoccurring beats.
Childhood: The period from the fifth to the tenth year of life (Schwarzer and Jovanovic, 2015).
Early childhood: The period from the second to the fourth year of life (Schwarzer and Jovanovic, 2015).
Entrainment: A neural process; when neural oscillations couple to an external rhythm.
Infancy: The first year of a child's life (Schwarzer and Jovanovic, 2015).
Newborns: The first two weeks of life (Schwarzer and Jovanovic, 2015).
Meter: Organizes a series of beats into recurring patterns of stressed and unstressed beats.
Rhythm: Serial pattern of variable tone durations in a melody that unfolds within a timing framework of a recurring pattern of beats.
Tempo: The pace of music, or the rate at which beats unfold over time.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 29 September 2022
DOI 10.3389/fpsyg.2022.907834

OPEN ACCESS

EDITED BY
Viola Macchi Cassia,
University of Milano-Bicocca, Italy

REVIEWED BY
Chiara Cantiani,
Eugenio Medea (IRCCS), Italy
Joëlle Provasi,
Université de Sciences Lettres de Paris,
France

*CORRESPONDENCE
Lira Yu
lirayu@gmail.com

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 30 March 2022
ACCEPTED 13 September 2022
PUBLISHED 29 September 2022

CITATION
Yu L, Todoriki K and Myowa M (2022) From spontaneous rhythmic engagement to joint drumming: A gradual development of flexible coordination at approximately 24 months of age.
Front. Psychol. 13:907834.
doi: 10.3389/fpsyg.2022.907834

From spontaneous rhythmic engagement to joint drumming: A gradual development of flexible coordination at approximately 24 months of age

Lira Yu[1]*, Kaho Todoriki[2] and Masako Myowa[2]

[1]Graduate School of Arts and Sciences, The University of Tokyo, Meguro, Japan, [2]Graduate School of Education, Kyoto University, Kyoto, Japan

Humans have a flexible and accurate ability to coordinate their movement in time with external rhythms. However, it remains unclear when and how, during their development, human children acquire the ability to adjust tempo and control the timing of their movement toward others. A previous study suggested that such self-regulation of coordination develops at around 18 and 30 months after birth. In this study, we investigated the performance of 24-month-old children and compared their data with those of 18- and 30-month-olds provided in our previous study. In the joint-drumming task, each child was enticed to drum under four conditions [partner: mother or robot; speed: 400 or 600 ms inter-stimulus-interval (ISI)]. The most pivotal test condition was the 600 ms ISI speed condition (slower than children's spontaneous motor tempo in these age groups). We found that from the age of 24 months, children try to slow down their drumming tempo toward the 600 ms ISI speed condition, regardless of the drumming partner. On the other hand, significant timing control toward the onset of the 600ms ISI condition was observed from the age of 30 months. This implies that both motor and cognitive mechanisms are required for flexible tempo adjustment and accurate synchronization and that these develop gradually among 18-, 24-, and 30-month-olds.

KEYWORDS

joint drumming, spontaneous motor tempo, tempo adjustment, synchronization, entrainment, inhibitory control, action prediction

Introduction

Humans have a flexible and accurate ability to coordinate their movement to match external rhythms. For example, while listening to music, we often produce finger/foot-taps to the music beat. When walking with another person, we tend to match our stride with others without attentive effort. Such an ability to coordinate one's movement in time with external rhythms is known to be one of the most important social-cognitive abilities in humans, since it enables us to establish smooth social interaction, communication, and cooperation with others (McNeill, 1997).

When and how do infants coordinate their movements in time with their external rhythms? In infants under 2 years of age, spontaneous rhythmic engagement with music, also referred to as *entrainment*, has been observed in a musical context (e.g., Zentner and Eerola, 2010; Fujii et al., 2014; Rocha and Mareschal, 2017; Cirelli and Trehub, 2019). The infants in these studies mostly demonstrated bouts of repetitive movements using the limbs, torso, or head and some extent of tempo flexibility. Conversely, in children over 2 years of age, studies that investigated specific target movements, such as tapping or drumming behaviors, demonstrated that children begin to show persistent rhythmic movements and flexible and accurate coordination with auditory rhythms (e.g., Drake et al., 2000; Provasi and Bobin-Bègue, 2003) or to a drumming partner (e.g., Kirschner and Tomasello, 2009; Endedijk et al., 2015). These studies suggest that flexible tempo adjustment and accurate synchronization appear in children as young as 2.5 years. However, since there are few studies examining children that are both under and above 2 years in the same experimental conditions (but see Kragness et al., 2022), it is still unclear whether, and to what extent, children under 2 years have coordination abilities compared to older age groups.

Our previous study extended the testable minimum age by adapting a joint-drumming task from Kirschner and Tomasello (2009) as follows: (1) introducing a drumstick to be used for drumming, (2) setting the participants' mother as a social partner, and (3) playing a well-known song ("Twinkle Twinkle Little Star") during the test phase to elicit the participants' repetitive drumming (see Yu and Myowa, 2021 for more detail). Those adaptations enabled us to examine 18-, 30-, and 42-month-old children's abilities under the same experimental conditions. The results showed that tempo adjustment toward the 400 ms ISI condition, which is close to the spontaneous motor tempo (SMT) for these age groups (Provasi and Bobin-Bègue, 2003; Bobin-Bègue and Provasi, 2008; Rocha et al., 2021), was observed in children as young as 18 months old. In contrast, tempo adjustment and synchronization ability toward the 600 ms ISI condition were observed from the age of 30 months, regardless of the drumming partner (i.e., mother or robot). A subject's SMT slows with age during childhood, and as the SMT in adults is approximately 600 ms (Fraisse, 1982; McAuley et al., 2006), movement coordination under the 600 ms ISI condition may be difficult for children. Overall, the findings suggest that flexible tempo adjustment and accurate synchronization develops between the ages of 18 and 30 months.

Flexible coordination is also necessary for turn-taking. Meyer et al. (2015) assessed the ability to predict others' actions (action prediction) and control one's own actions (inhibitory action control) in 30-month-old children in a turn-taking game. They demonstrated that action prediction was related to turn-timing variability (i.e., time interval between the last button press by the adult experimenter and the button press of the child), whereas inhibitory action control was related to turn-taking accuracy (i.e., a correct execution of button pressing in alternation between the adult partner and the child). This suggests that both motor and cognitive abilities play distinct roles in early joint action coordination.

Based on these findings, the current study aimed to reveal when children develop the ability of flexible tempo adjustment, as well as how motor development and cognitive ability to predict others' actions affect development. We examined 24-month-olds under the joint-drumming task and compared their data with those of 18- and 30-month-olds from our previous study (Yu and Myowa, 2021). Regarding flexible tempo adjustment, we examined whether children's drumming tempo significantly changed depending on the speed condition (400 or 600 ms ISI condition). For the children's motor development, we examined their SMT from their free-drumming responses during the familiarization phase. Regarding the ability to predict others' actions, we examined the children's synchronization ability by performing Rayleigh's test. We hypothesized the following: 1) a tempo adjustment toward a slower tempo than the children's SMT will develop gradually in 18-, 24-, and 30-month-olds; 2) the ability to perform a flexible tempo adjustment will require a slowdown of the SMT; and 3) the ability to perform a flexible tempo adjustment will require accurate synchronization.

Materials and methods

Participants

Nineteen 24-month-old children participated in the experiment (12 girls and 7 boys; mean age = 24.21 months; range = 23.54 to 25.02 months). Three additional children were excluded from the analysis because of excessive fussiness. The children's parents were asked to provide written informed consent before participation. The study protocol adhered to the Declaration of Helsinki and was approved by the ethics review board of the Kyoto University Unit for the Advanced Study of the Mind. Data for 18-month-olds (10 girls and 8 boys; mean age = 18.21 months; range = 17.52 to 18.74 months) and 30-month-olds (9 girls and 9 boys; mean age = 30.17 months; range = 28.60 to 31.89 months) were taken from our previous study (Yu and Myowa, 2021).

Apparatus and stimuli

We used the same apparatus and stimuli as in our previous study (Yu and Myowa, 2021). For the drumming behavior, two sets of toy drums (24.5 cm diameter, 12 cm high) and toy drumsticks (15 cm length) were used. A vibration sensor (Piezo film sensor) was attached beneath each drumming surface and a PC data logger (U3HV-LJ, LabJack Co.) was used to record the signals from both drums. A preprogrammed drumming robot was introduced to examine the partner's effect (i.e., mother or robot).

A digital video camera (HDR-CX670, SONY) was used to film the participants and their mothers throughout the experiment. For each trial, a speaker (Fostex PM0.1) played background music, namely the melody of "Twinkle Twinkle Little Star" and Audacity 2.4.2 was used to create versions of the two speed conditions (i.e., 400 and 600 ms ISI conditions). Sixty-four beats were presented regardless of the speed condition. Thus, the durations of the 400 and 600 ms ISI conditions were 25 s and 38 s, respectively.

Procedure

This study was conducted with the cooperation of the participants' mothers. After receiving informed consent, the experimenter (author LY) explained further procedures to the mother, including the verbal instructions to be given to the child, depending on the condition. A practice for matching drumming in time with the presented music was also conducted. During this time, a research assistant (author KT) interacted with the child with a few toys and books, beside the parent and experimenter.

Following a brief warm-up period in the waiting area that contained the PC and monitors used in the study, the child and mother moved to the experimental space with the experimenter. The experimental space was separated from the waiting area using a curtain (Figure 1). The child and mother sat next to each other on a floor mat, and the experimenter sat in front of them on another floor mat. After the experimenter placed a drum on each floor mat and handed one drumstick to the child, a familiarization phase consisting of three sequences started as follows: (1) the child was encouraged to drum freely while the experimenter sat in front of the child; (2) the child and mother were allowed to drum freely while facing each other, each using one drum; and (3) the child was introduced to the drumming robot, named *Shikaku-chan*, and was allowed to drum freely with the robot or touch it. Each situation lasted no longer than 1 min to avoid the child's loss of interest in drumming. No music was played throughout the familiarization phase.

Following the familiarization phase, the experimenter left the experimental space, drew the curtain, remained in the waiting area, and started the test phase. Each participant was tested under four conditions in a single day (partner: mother or robot; speed: 400 or 600 ms ISI). The order of speed and partner under the same speed condition was counterbalanced across participants. Each condition was repeated only once. The second trial was conducted only when the child did not drum at all in the first trial. Each trial started with verbal instructions provided by the child's mother. In the mother condition, the mother said, "Shall we play the drum together now?" while sitting in front of the child. A few seconds later, the melody of "Twinkle Twinkle Little Star" was played, and the mother began drumming along with every beat of the music. In the robot condition, the mother said, "Can you drum with *Shikaku-chan* now?" while sitting beside the robot. A few seconds later, the robot began drumming and music was played to accompany the movement of the robot. In all conditions, the mother unconditionally praised the child after the trial ended. Between the conditions, the children were given stickers to play with for approximately 3 min.

Data analysis

To assess the children's ability to make tempo adjustments, the median of inter-response-intervals (IRIs) was calculated for each trial. To test whether children changed the drumming tempo depending on the speed condition (400 or 600 ms ISI), the Kruskal–Wallis rank sum test was conducted for each age group in both mother and robot conditions. We calculated the median IRIs only when the trial included more than four repeated

FIGURE 1
Waiting area (front) and experimental space (back) of this study.

drumming hits (i.e., more than three IRIs), without unusually large IRIs. If we detected an unusually large IRI that exceeded 2 s, we checked the videos. We eliminated large IRIs if the child changed the drumming hand, took a break (i.e., hands free from drumming), or drummed without hitting the drum surface. As in our previous study, we checked the mothers' drumming tempo for the mother condition. In the 24-month-old group, no data were excluded due to the mother's too fast or too slow drumming tempo (±5% from the designated speed).

To measure the children's SMT, the free-drumming responses during the familiarization phase were examined. The SMT of the 24-month-old group was extracted from the responses observed when the children drummed alone in front of the experimenter.

To assess the children's ability to synchronize, Rayleigh's test was performed for each trial. This tests the null hypothesis of circular uniformity. Rejection of the null hypothesis indicated that the children controlled their drumming at a specific timing in response to the onset of their partner's drumming. We performed Rayleigh's test only when the trial included more than 10 drumming hits (Zar, 2019). All statistical tests were conducted using R software (R Core Team, 2020).

Results

Tempo adjustment

To test whether the children showed different drumming responses depending on the speed condition (400 or 600 ms ISI), we examined the median IRIs of each age group in the mother and robot conditions, respectively. In the mother condition (Figure 2A), the three age groups showed marginal or significant differences in drumming tempo between the 400 and 600 ms ISI conditions (Kruskal–Wallis rank sum test: 18-month-olds, $p = 0.057$; 24-month-olds, $p < 0.01$; 30-month-olds, $p < 0.001$). In contrast, in the robot condition (Figure 2B), while both the 24- and 30-month-olds showed a significant difference in drumming tempo depending on the speed condition ($p < 0.001$), no significant difference was observed in the 18-month-olds ($p = 0.683$). As shown in Supplementary Data S1, we further examined whether the variability of the drumming tempo or drumming frequency changed depending on the speed condition.

SMT and tempo adjustment

To test whether the children who shifted to slower SMT were better at tempo adjustment, the relationship between the children's SMT and the median IRI was examined in each condition (Figure 3). In the 24- and 30-month-olds, a positive correlation was found in the 600 ms ISI condition in both the mother and robot conditions (mother condition: 24-month-olds, $r(8) = 0.77$, $p < 0.01$; 30-month-olds, $r(11) = 0.70$, $p < 0.01$; robot condition: 24-month-olds, $r(4) = 0.82$, $p < 0.05$; 30-month-olds, $r(11) = 0.69$, $p < 0.01$). On the other hand, in the 18-month-olds, a positive correlation was found in the 400 ms ISI condition when they drummed in the mother condition. In Supplementary Data S2, we show the changes in the children's SMT, the variability of the SMT, and the total number of drumming hits used for the SMT measurements across the three age groups. We found no significant correlation between SMT variability and tempo adjustment ability in any of the three age groups across the four test conditions.

Synchronization and tempo adjustment

Table 1 shows the absolute number (n) and percentage (%) of children reaching significance in Rayleigh's test ($p < 0.05$). The

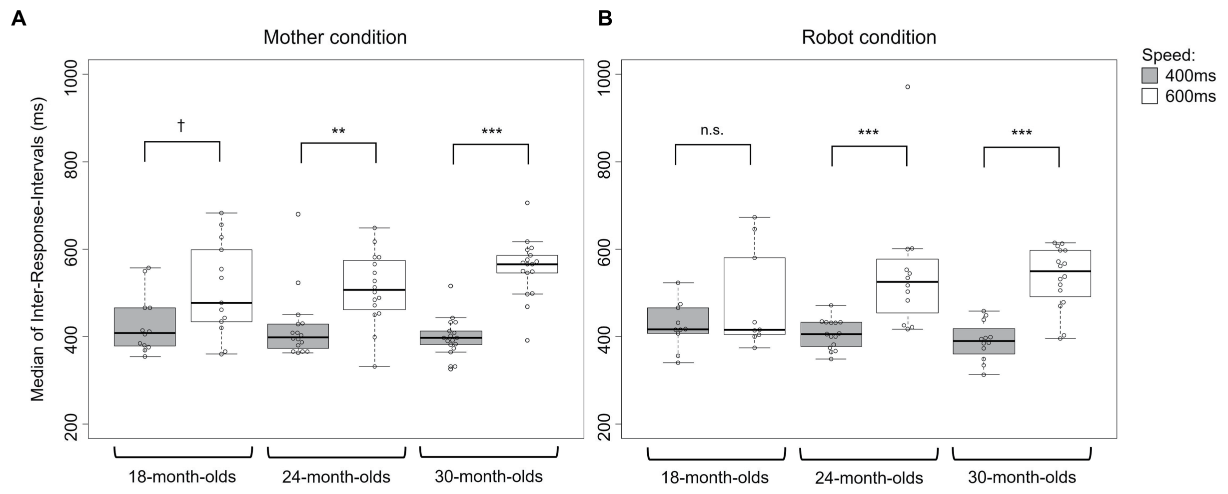


FIGURE 2
Median of inter-response intervals (IRIs) depending on the two speed conditions under **(A)**, mother condition and **(B)**, robot condition.

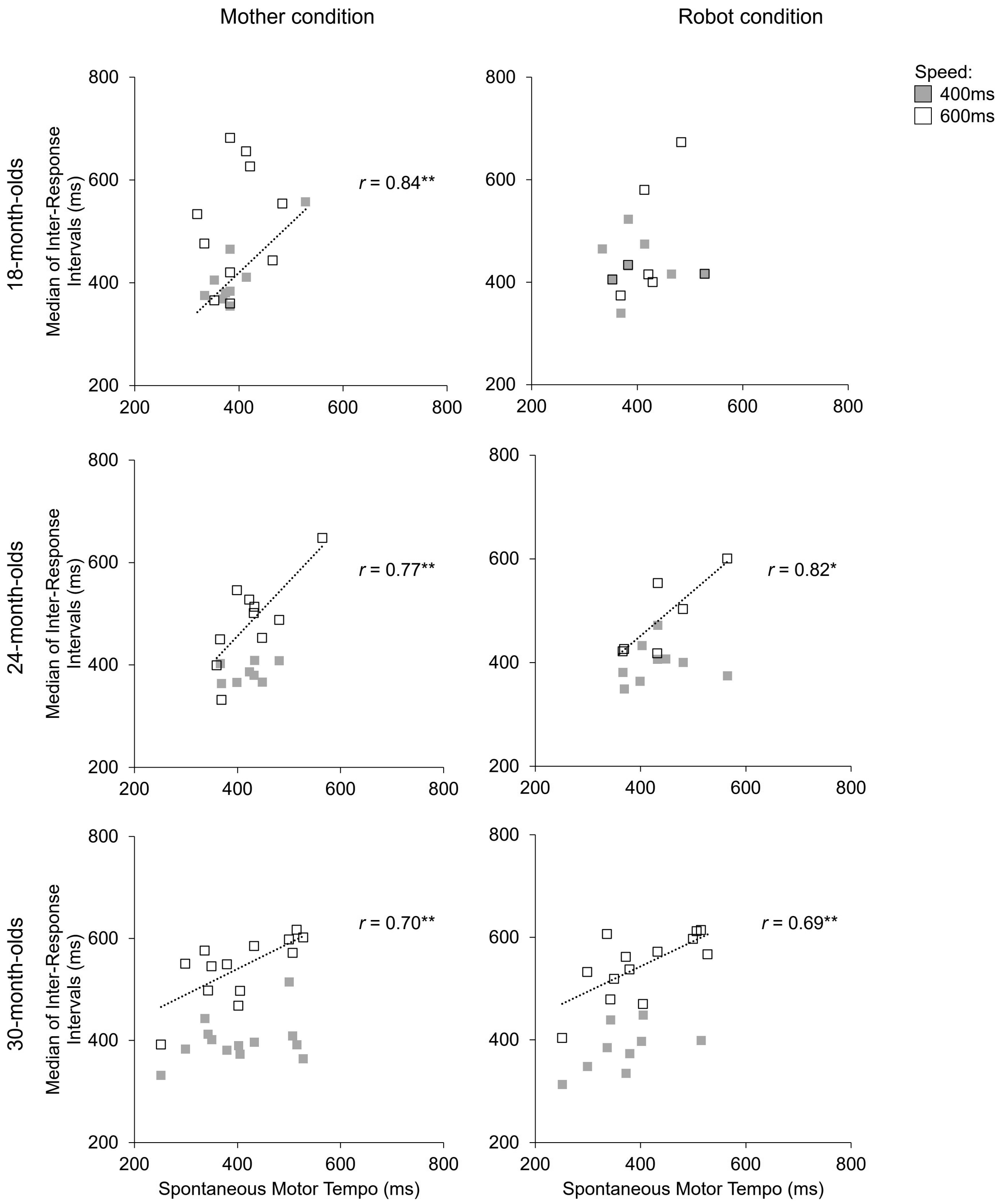


FIGURE 3
Correlation between the children's spontaneous motor tempo (SMT) and the median of inter-response-intervals (IRIs) during the four test conditions.

percentages were calculated as a ratio of the number of children who reached significance to the number of children tested. We found a developmental change in the absolute number of children who reached significance across the four conditions. In the 400 ms ISI condition, the number of children that reached significance tended to increase between the ages of 18 and 24 months. Conversely, in the 600 ms ISI condition, a prominent increase was observed between the ages of 24 and 30 months. These developmental trajectories were common in both the mother and robot conditions.

To examine the relationship between the abilities of accurate synchronization and flexible tempo adjustment, we first categorized the children into two groups: (1) the synchronous group, which included the children who reached significance in

TABLE 1 Absolute number and percentage of children reaching significance in Rayleigh's test.

Partner	Speed	18-month-olds			24-month-olds			30-month-olds		
		Tested	*n*	%	Tested	*n*	%	Tested	*n*	%
Mother	400 ms	8	3	37.5	15	7	46.67	17	8	42.06
	600 ms	7	3	42.86	16	3	18.75	17	11	64.71
Robot	400 ms	6	0	0	15	4	26.67	12	4	33.33
	600 ms	3	0	0	10	4	40	16	9	56.25

Rayleigh's test and showed a phase preference between −90° and + 90° (i.e., in-phase synchrony); (2) the "other" group, which included the rest of the children. We then compared the tempo adjustment abilities between the two groups under the same speed condition. In both 24- and 30-month-olds, we found a marginal difference in the tempo adjustment ability between the synchronous and other groups in the 600 ms ISI condition when drumming in the mother condition (Kruskal–Wallis rank sum test: 24-month-olds, $p = 0.068$; 30-month-olds, $p = 0.056$; Figure 4).

Discussion

The current study aimed to clarify the developmental process of flexible and accurate rhythmic coordination by examining when and how tempo adjustment and synchronization develop between 18-, 24-, and 30-month-old children. The most pivotal test condition in the current joint-drumming task was the 600 ms ISI speed condition. To coordinate with the 600 ms ISI, the children needed to slow down and control the timing of their drumming because the SMT for children around the ages of our participants was close to the 400 ms ISI (see also Provasi and Bobin-Bègue, 2003; Bobin-Bègue and Provasi, 2008; Rocha et al., 2021). The analysis of tempo adjustment indicated that children from the age of 24 months were able to slow down their drumming toward the 600 ms ISI condition, regardless of drumming partner. On the other hand, the analysis of synchronization indicated that children from the age of 30 months were able to control the timing with the onset of their partner's drumming in the 600 ms ISI condition. These findings suggest that flexible tempo adjustment and accurate synchronization abilities develop gradually among 18-, 24-, and 30-month-olds.

The relationship between the children's SMT and their tempo adjustment abilities demonstrated a significant correlation in 24- and 30-month-olds. This finding suggests that, as the children shifted to slower SMT (i.e., close to 600 ms), they were better at tempo adjustment during the 600 ms ISI condition. This further suggests that flexible tempo adjustments toward slower tempo than one's own SMT develops later than tempo adjustments toward faster tempo in children. Regarding the positive correlation observed in the 18-month-olds, we assumed this result to be a false positive, as a child who showed an SMT close to 600 ms produced a drumming tempo of approximately 600 ms during the 400 ms ISI speed condition.

As previous studies have reported (e.g., Provasi and Bobin-Bègue, 2003; Bobin-Bègue and Provasi, 2008), the current study found that inter-individual differences in the children's SMT tended to increase among the 18-, 24-, and 30-month-olds. This may be due to the divergence of the individual differences in the children's motor development as their SMT shifts toward 600 ms, which is the SMT of adults (Fraisse, 1982; McAuley et al., 2006). Interesting further investigations may include clarifying the biological foundations of SMT, if an SMT of approximately 600 ms is a universal in adults of all cultures.

Although it was a marginal effect, the relationship between synchronization and tempo adjustment demonstrated that the children's ability to synchronize accurately might have facilitated their tempo adjustment, especially in the 600 ms ISI condition. In the literature on human adults, two processes for error correction—period and phase correction—are known to act jointly to sustain accurate synchronization (Repp, 2005). Accordingly, during development, it is plausible that children use their ability to predict others' actions to synchronize their own movements with those of another (i.e., phase correction), and this inevitably results in accurate tempo adjustments (i.e., period correction).

Compared to the two older age groups, children aged 18 months showed difficulty in flexible tempo adjustment and accurate synchronization toward the 600 ms ISI condition. Moreover, they showed few drumming hits across the four conditions. Our post-hoc video analysis indicated that most of the 18-month-olds exhibited rhythmic movements other than drumming behavior, such as head bobbing, body swaying, or bouncing (see Supplementary Data S3). As previous studies have reported (e.g., Zentner and Eerola, 2010; Fujii et al., 2014; Rocha and Mareschal, 2017; Cirelli and Trehub, 2019), it is plausible that the 18-month-olds were showing spontaneous rhythmic engagement with the music rather than attending to the partner to drum together. However, from the age of 24 months, the proportion of children showing other rhythmic movements decreased, and these children demonstrated more drumming hits compared to those of 18-month-olds. This suggests that the inhibition of other rhythmic movements, as well as attentional shift toward the partner (i.e., joint attention, see Siposova and Carpenter, 2019 for a review), were necessary for the children to produce drumming behavior in the current task.

In summary, we found that flexible tempo adjustment and accurate synchronization abilities develop gradually among 18-, 24-, and 30-month-olds in the joint-drumming task. Moreover, the findings demonstrated that both children's motor development and cognitive ability to predict others' actions function jointly in the development of flexible coordination in early childhood.

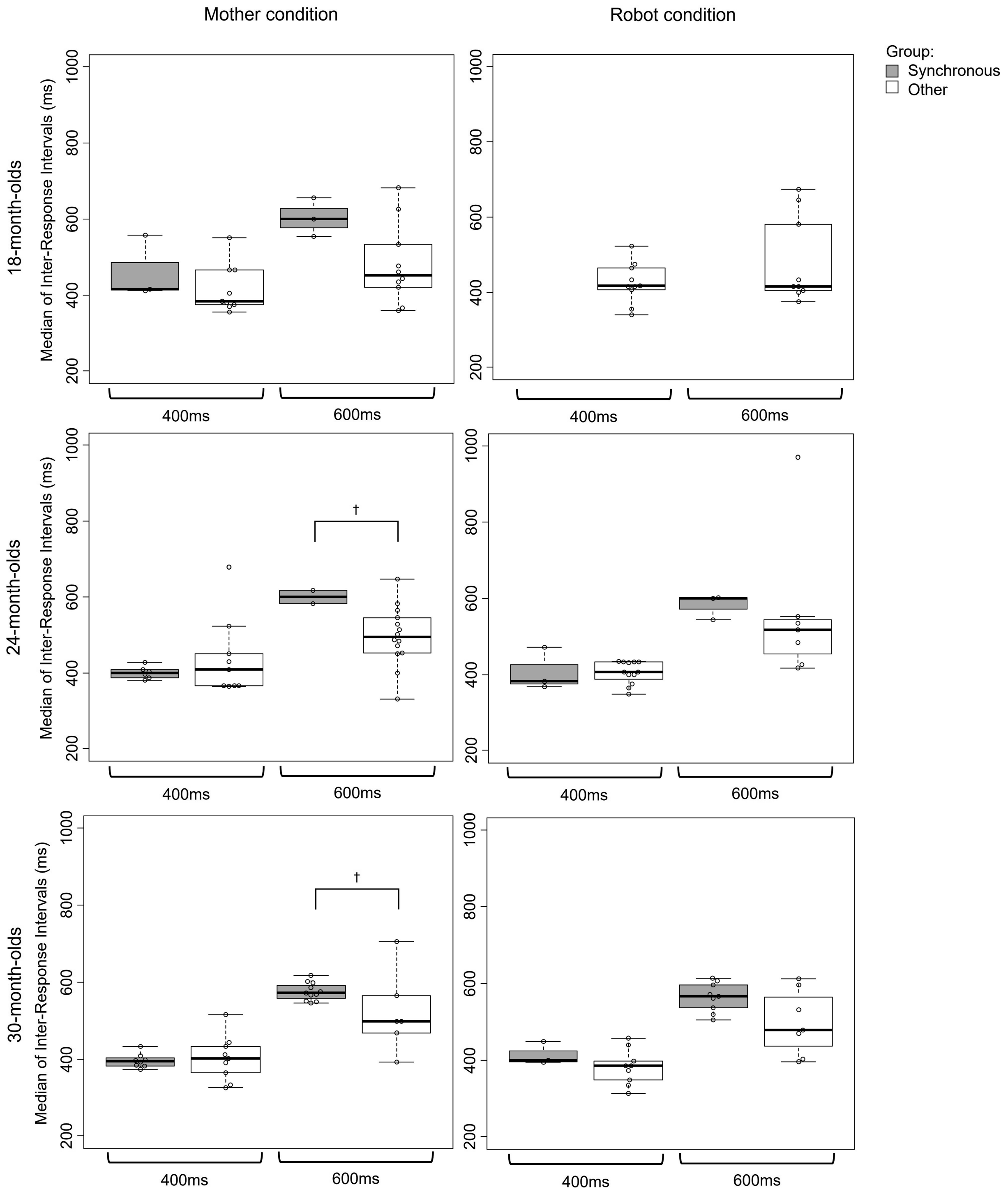


FIGURE 4
Relationship between the abilities of synchronization and tempo adjustment.

Data availability statement

The data analyzed in this study are available from the corresponding author upon reasonable request.

Ethics statement

The studies involving human participants were reviewed and approved by Kyoto University Unit for the Advanced Study of the

Mind. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LY and MM designed the experiments and wrote the paper. LY and KT conducted the experiments. LY analyzed data. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the JSPS Grants-in-Aid for Scientific Research (17H01016, 18H05524, 16F16001, 19K20646).

Acknowledgments

We thank all members at Myowa Lab for their assistance with this experiment. We also thank Takashi Nakagawa for his technical support. We thank James Anderson and Wataru Yano for their invaluable comments on an earlier draft.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https: //ww.frontiersin.org/articles/10.3389/fpsyg.2022.907834/full#supplementary-material

References

Bobin-Bègue, A., and Provasi, J. (2008). Régulation rythmique avant 4 ans: effet d'un tempo auditif sur le tempo moteur. *Année Psychol.* 108:631. doi: 10.4074/S000350330800403X

Cirelli, L. K., and Trehub, S. E. (2019). Dancing to Metallica and Dora: case study of a 19-month-old. *Front. Psychol.* 10:1073. doi: 10.3389/fpsyg.2019.01073

Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 251–288. doi: 10.1016/S0010-0277(00)00106-2

Endedijk, H. M., Ramenzoni, V. C., Cox, R. F., Cillessen, A. H., Bekkering, H., and Hunnius, S. (2015). Development of interpersonal coordination between peers during a drumming task. *Dev. Psychol.* 51, 714–721. doi: 10.1037/a0038980

Fraisse, P. (1982). Rhythm and tempo. *Psychol. Music* 1, 149–180. doi: 10.1016/B978-0-12-213562-0.50010-3

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., and Taga, G. (2014). Precursors of dancing and singing to music in three-to four-months-old infants. *PLoS One* 9:e97680. doi: 10.1371/journal.pone.0097680

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Kragness, H. E., Ullah, F., Chan, E., Moses, R., and Cirelli, L. K. (2022). Tiny dancers: effects of musical familiarity and tempo on children's free dancing. *Dev. Psychol.* 58, 1277–1285. doi: 10.1037/dev0001363

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *J. Exp. Psychol. Gen.* 135, 348–367. doi: 10.1037/0096-3445.135.3.348

McNeill, W. H. (1997). *Keeping Together in Time: Dance and Drill in Human History*. Cambridge Harvard University Press.

Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., and Hunnius, S. (2015). The role of action prediction and inhibitory control for joint action coordination in toddlers. *J. Exp. Child Psychol.* 139, 203–220. doi: 10.1016/j.jecp.2015.06.005

Provasi, J., and Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2½-and 4-year-old children. *Int. J. Behav. Dev.* 27, 220–231. doi: 10.1080/01650250244000290

R Core Team (2020). R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria. available at: https://www.R-project.org/

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992. doi: 10.3758/BF03206433

Rocha, S., and Mareschal, D. (2017). Getting into the groove: the development of tempo flexibility between 10 and 18 months of age. *Infancy* 22, 540–551. doi: 10.1111/infa.12169

Rocha, S., Southgate, V., and Mareschal, D. (2021). Infant spontaneous motor tempo. *Dev. Sci.* 24:e13032. doi: 10.1111/desc.13032

Siposova, B., and Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition* 189, 260–274. doi: 10.1016/j.cognition.2019.03.019

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: a study of 18- to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/infa.12403

Zar, J. H. (2019). *Biostatistical Analysis. 5th Edn.* Pearson Education. Upper Saddle River, New Jersey.

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5768–5773. doi: 10.1073/pnas.1000121107

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 21 October 2022
DOI 10.3389/fpsyg.2022.896319

OPEN ACCESS

EDITED BY
Stefanie Peykarjou,
Heidelberg University, Germany

REVIEWED BY
Martin Hartmann,
University of Jyväskylä, Finland
Sinead Rocha,
Birkbeck College, University of
London, United Kingdom

*CORRESPONDENCE
Zuzanna Laudańska
zlaudanska@psych.pan.pl
David López Pérez
d.lopez@psych.pan.pl
Przemysław Tomalski
ptomalski@psych.pan.pl

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 14 March 2022
ACCEPTED 04 October 2022
PUBLISHED 21 October 2022

CITATION
Laudańska Z, López Pérez D, Kozioł A,
Radkowska A, Babis K,
Malinowska-Korczak A and Tomalski P
(2022) Longitudinal changes in infants'
rhythmic arm movements during
rattle-shaking play with mothers.
Front. Psychol. 13:896319.
doi: 10.3389/fpsyg.2022.896319

Longitudinal changes in infants' rhythmic arm movements during rattle-shaking play with mothers

Zuzanna Laudańska*, David López Pérez*, Agata Kozioł,
Alicja Radkowska, Karolina Babis, Anna Malinowska-Korczak
and Przemysław Tomalski*

Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Warsaw,
Poland

From early on, infants produce a variety of rhythmic behaviors—an ability that likely supports later social communication. However, it is unclear, how this rhythmic motor production changes with age. Here, we investigated the coupling between infants' arm movements across the first year of life in a social context of a rattle-shaking play with their mothers. Through longitudinal measurements at 4, 6, 9, and 12 months of age using wearable motion trackers placed on infants' arms, we show that infants (N = 40) are similarly motivated to attempt rattle-shaking across the first year of life. However, with age, they make more rattling movements with an increased frequency. Their left and right arm movements become more coupled during rattle-shaking, as shown by an increase in wavelet coherence. Infants produced more rattling movements when they were rattling alone than when their mothers were rattling or singing simultaneously. There were no differences between infants' individual and social rattling in between-arms coherence. Our results may help to understand rhythmic arm movements as precursors of motor social coordination.

KEYWORDS

rhythm, motor development, infants, rhythmic arm movements, wearables, inertial motion units, wavelet coherence

1. Introduction

Humans have the ability to produce rhythmic actions and coordinate their movements to external rhythms. Generally, a rhythm can be defined as a sequence of short and repeated intervals, with regularities that allow us to build expectancies when the next beat arrives (Jones, 1976), or as a recurrent non-random temporal pattern of actions that may not be strictly regular (Jaffe et al., 2001). Previous studies suggest that infants have the innate ability to process rhythms, since already newborns can detect on a cortical level the violation of the beat of a rhythmic sound sequence (Winkler et al., 2009) and the onset and offset of sound trains and changes in the presentation rate (Háden et al., 2015). In addition, it was shown that beat perception abilities are culture-specific (Hannon and Trehub, 2005). Furthermore, behavioral experiments demonstrated that 2-month-olds could discriminate between different musical rhythm patterns (Demany et al., 1977), even when the beat difference was small (Baruch and Drake, 1997), whereas 5-month-olds discriminated between different temporal groupings of audio stimuli (Chang and Trehub, 1977). Thus, it seems that rhythm perception and processing occurs from early on and become more specialized with age.

Overall, the majority of previous research was focused on children's perception and processing of rhythms. In contrast, the production of rhythmic actions by infants and children has been far less investigated. So far, it is known that across early development, infants produce various rhythmic behaviors (e.g., kicking, rocking, waving) with a peak period of rhythmic hand-banging around 6–7 months of age (Thelen, 1979, 1981). The ability to keep a steady beat and produce a spontaneous motor tempo emerges earlier than the ability to synchronize to an external beat (Provasi and Bobin-Bègue, 2003; Zentner and Eerola, 2010; Provasi et al., 2014). Infants' spontaneous motor tempo during drumming was observed from 5 months of age. It is slower than the adult one and it becomes faster and more regular with age (Rocha et al., 2021a,b). However, it is unknown whether the production of rhythmic movements changes during interactions with social partners.

The social context seems to modulate infants' production of some motor actions but only at a later age. For example, bouncing and rocking were displayed by 18-month-olds more often in the absence of a social partner (in a condition where they were presented with a non-social visual animation). In younger infants, at 10 months of age, these behaviors were not modulated by the presence or absence of a social partner (Rocha and Mareschal, 2017). A similar pattern was observed by Rocha and collaborators (Rocha et al., 2021a) during the drumming task—infants spent a higher proportion of time in rhythmic movement during the non-social trials. The social context seems to also facilitate joint drumming synchronization in preschool children (Kirschner and Tomasello, 2009; Yu and Myowa, 2021). Children's ability to coordinate their rhythmic activities with a partner develops between 18 and 30 months of age and studies with 18-month-olds have shown the crucial, facilitating role of a social partner's actions as opposed to those of a robot (Yu and Myowa, 2021).

Studying infants' rhythmic actions in the context of social interactions could be key to better understand whether these rhythmic actions form a foundation for later social communication (Jaffe et al., 2001). Infants early on start to engage in proto-conversations with their caregivers, in which infants' movement patterns are very responsive to the time structure of their mothers' movements (Trevarthen, 1979). Communication with others and verbal dialogues are also rhythmic activities, where both timing and synchronization of own actions with the interlocutor's actions are crucial (Jaffe et al., 2001). Production of rhythmic movements—especially the ones that result in multimodal feedback such as drumming or rattling—may be an opportunity for infants to learn about contingencies between their actions and outcomes of those actions. Repetitive and recurrent movements are also an opportunity to practice specific types of limb movements and master their execution. Since motor coordination is another important aspect of dyadic interactions, it seems key to better understand its early precursors during rhythmic actions. Altogether, more studies are needed to describe the role of a social partner in rhythmic activities and how these rhythmic actions may form the foundation for later social communication (Jaffe et al., 2001).

In the present study, we investigated how infants' spontaneous rhythmic behavior in the social context of play changes in development. Our main goal was to study the developmental changes in motor coordination between arm movements during rattle-shaking. Furthermore, we also studied whether infants produce more rhythmic arm movements as they grow older and whether they do it at a higher frequency. Additionally, we explored the role of the social partner in infants' rattling. We studied the changes in rhythmic arm movements in a naturalistic set-up: mother-infant dyads were invited to play together in the lab. Their interactions were video-recorded, which enabled us to annotate, during which episodes infants were rattling alone and during which mothers were rattling or singing alongside each other. We have compared these categories in exploratory analyses to see whether there are potential differences between infants' individual and social rattling.

To this end, we first recorded infants' arm movements using wearable motion trackers (Inertial Motion Units, IMUs) in a rattle-shaking task during parent-infant interactions when infants were around 4, 6, 9 and 12 months of age. Secondly, we identified and manually annotated the episodes when infants were rattling to include only this type of activity in further analyses. Thirdly, we classified episodes of infant rattling into two categories: "Mother Not Providing Rhythm" in instances where the infant was rattling alone and "Mother Providing Rhythm" in instances where during the infant's rattling, the parent was providing them with auditory stimulation by rattling or singing. Fourthly, we calculated the number of rattle-shakes (i.e., infant arm movements with a rattle) in a data-driven way. This, in turn, allowed us to calculate the rattling frequency and the coordination between the movements of both arms. To assess the degree of coordination between the infants' two arms, we used wavelet coherence, which captures information on a range of constituent frequencies of the signals across the recorded interaction (e.g., Grinsted et al., 2004; Hale et al., 2020).

We hypothesized that (1) infants would be able to produce more rhythmic arm movements with age (Rocha et al., 2021a,b), (2) they would rattle at a higher frequency with age, and (3) their between-arms coordination (measured with wavelet coherence) would increase with age. The analyses regarding the effect of rattling alone vs. rattling with a mother were exploratory and we did not have any a priori hypotheses.

TABLE 1 Sample characteristics.

Time point	*N*	Number of girls	Number of boys	Mean age in months (SD)	Min age in months	Max age in months
T1	31	10	21	4.35 (0.29)	3.90	5.20
T2	35	13	22	6.55 (0.36)	6.00	7.40
T3	39	13	26	9.14 (0.39)	8.60	10.20
T4	21	4	17	12.05 (0.37)	11.60	13.10

2. Methods

2.1. Participants

Participants were 40 Polish mother-infant dyads from an ongoing longitudinal study on infant limb movement during social interactions and language development. Participants were invited to the lab when the infants were around 4 (T1), 6 (T2), 9 (T3), and 12 (T4) months old. Six infants contributed data at all four time points, whereas 34 infants missed one visit (mostly due to Covid-19 related restrictions). Therefore, 20 infants contributed data at T1, T2 and T3, 9 at T2, T3, and T4, and 5 at T1, T3 and T4 (see Table 1 for an overview of sample characteristics). Participants were from predominantly middle-class families living in the Warsaw metropolitan area. The majority of mothers had completed higher education: 35 held a master's degree, 2 held a bachelor's and 2 completed high school (1 missing data). For their participation, infants received a diploma and a small gift (a baby book). The study received clearance from the Ethics Committee at the Institute of Psychology, Polish Academy of Sciences.

2.2. Procedure

Infant-parent interactions were recorded in a laboratory room, in a carpeted play area. Upon the family's arrival, an experimenter explained the study protocol and obtained parental written consent. Once the infant was familiarized with the laboratory, the wearable motion trackers attached to elastic bands were put on the infants' and caregivers' bodies. Then, the infant-parent dyads took part in a series of interactive games with different sets of age-appropriate toys. There were 6–7 different tasks during each meeting, but here we report data only from the rattle-shaking task. In this task, which lasted approximately 5 min, the caregivers were instructed to play with their infants using the provided rattles in their preferred way. They were given two maracas rattles and two other rattles of different types (smaller and lighter barbell rattles at T1 and T2 and bigger teddybear rattles at T3 and T4, see Figure 1). At the beginning of each game, the caregivers were asked to clap several times to mark the start of the procedure to synchronize wearable sensors with video recordings. The infants' body position was not constrained and both the mother and the infant were free to move around the room. Therefore, the sitting arrangement varied between visits and could change during each visit. The most common body position at T3 and T4 was independent sitting, whereas for T1 and T2 was lying either in a prone or a supine position.

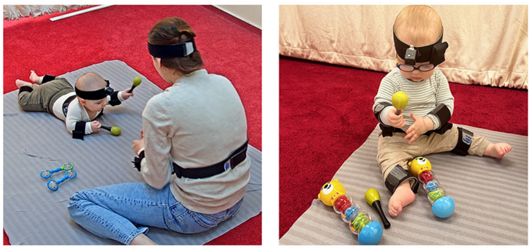


FIGURE 1
Photos of the toys used in the rattle-shaking play at T1 and T2 **(left)**, T3 and T4 **(right)**. Signed permission of the caregiver was acquired for the publication of the images.

2.3. Equipment

Infants' and mothers' body movements were recorded at 60 Hz using wearable motion trackers (MTw Awinda, Xsens Technologies B.V.) connected wirelessly through an Awinda station receiver (Xsens Technologies B.V.) and synchronized in real-time with MT Manager Software (Xsens Technologies B.V.). Overall, 12 sensors were used (on the infant's arms, legs, head, and torso, see Figure 1; and on the caregiver's arms, head, and torso), but in this paper, we report data only from two sensors placed on infant's arms.

The interactions were recorded with three remote-controlled CCTV color cameras in HD quality. During the interaction, an experimenter operated the cameras (this included zooming in and out as well as moving them vertically and horizontally) to ensure that at least one camera captured the infant's behavior.

2.4. Manual annotation of rattling

In each video recording, the episodes when infants were rattling as well as mothers clapping for the purpose of synchronization of videos with wearable data were manually annotated by a trained coder in ELAN 6.3 (Sloetjes and Wittenburg, 2008; ELAN 6.3, 2022). Firstly, the onset and offset of each clap were identified in a frame-by-frame manner to precisely include the moment of acceleration before joining hands. Secondly, the onset and offset of each infant rattling episode were annotated. We defined a rattling episode as a period when an infant was holding at least one rattle and made at least one movement that produced the rattling noise. Instances of an infant generating the rattling sound unintentionally (e.g., while holding a rattle during crawling or throwing it) were not annotated. Each episode ended if (1) the infant dropped the rattle or (2) was holding the rattle but not making any arm movements. Periods when an infant did not wear motion trackers on both arms were not annotated and excluded from the analyses. Periods when the mother was moving infant's arms were not annotated. In the second pass of coding, each rattling episode was assigned to a category: either (1) Mother Not Providing Rhythm (infant's rattling alone while the mother was not providing an auditory rhythm) or (2) Mother Providing Rhythm (this included rattling at the same time as the mother was rattling, singing, or both simultaneously, and rattling directly after the mother had finished rattling or singing).

In total, 126 videos were annotated. Videos during which mothers did not clap were excluded from further analyses ($N = 3$, two at T1, one at T3) due to problems with synchronizing motion trackers' data with video recording. Similarly, videos during which the infant did not make any rattling movements were excluded from further analyses ($N = 5$, two at T1, two at T2, one at T3). In order to establish the inter-rater reliability, 26 randomly selected videos (20%) were annotated separately by two trained coders. Inter-rater reliability was performed in ELAN and estimated using Cohen's κ statistic, which takes into account chance agreement. The mean Cohen's κ for rattling episodes was 0.79, which can be interpreted as substantial agreement (Landis and Koch, 1977).

2.5. Data pre-processing

Acceleration data from sensors placed on both wrists of an infant were processed in Matlab (Mathworks, Inc., Natick, USA) using in-house scripts. First, missing samples were identified and interpolated using the *interp1* function with cubic spline interpolation of the values at neighboring grid points. Then we collapsed the kinematic vectors obtained from the IMUs into a unique normalized dimension (a one-dimensional overall acceleration time series) as follows:

$$Acc = \sqrt{x(t)^2 + y(t)^2 + z(t)^2} \tag{1}$$

where Acc is the normalized acceleration

$$x, y, z \in \mathbb{R}^{1\times N} \tag{2}$$

and a_x, a_y, and a_z are the kinematic acceleration vectors in x, y, and z dimensions respectively at each time point t. Next, data were smoothed using the *medfilt1* function that applies a third-order median filter to remove one-point outliers by replacing each value with the median of three neighboring entries (see Supplementary Figure 1 for an example of the sensor time series).

2.6. Synchronization of sensor data and annotated videos

Video and sensor data for each infant and visit were later synchronized using the mothers' clapping (see Supplementary Section 2 for an example). To this end, a graphical user interface (GUI) loaded the sensor data to manually select the period when the clapping occurred (Supplementary Figures 1, 2). Then, we categorized the manually selected sensor periods from the GUI as "1" and "0", where 1 indicated movements that were one standard deviation above the mean acceleration in that period and 0 otherwise. Next, the time series outside the selected clapping period was set to 0. Finally, we merged those automatically detected claps separated by 50 ms or less to avoid artifactual claps due to extremely short claps or claps close together. This process resulted in a time series that contained only the mothers' claps. In the next step, this was used to find the delay between the IMUs data and the manually coded video data. To find this delay, we used diagonal cross-recurrence quantification analysis (DCRP) (e.g., Richardson and Dale, 2005) using two different time windows (a shorter window of 6 s and a longer one of 15 s). We calculated the lag profile using a Matlab version of the R function *drpdfromts* (CRQA R-package) (Coco and Dale, 2014). Generally, the experimenter initiated video and sensor recordings closely in time, so the lag between them usually was not longer than 6 s. Initially, the algorithm estimated the delay using the 6 s time window and loaded a GUI plotting both the sensor data and the manually coded data (see Supplementary Figure 3). This process asked the user to visually inspect and validate the proper alignment of the data. In 7% of cases, the lag between sensor and video data was longer than 6 s. Therefore, in these cases, we repeated the previous step, using a 15 s-long time window. Again, the alignment was visually inspected. Further analyses were performed on the temporally aligned time series.

2.7. Wavelet coherence analysis of arm movements

Wavelet coherence (WC) is a relative measure of how well-correlated the power and phase of two signals are at a given frequency and time (Grinsted et al., 2004) and it is defined as the squared absolute value of the smoothed cross-wavelet spectrum normalized by the product of the smoothed individual wavelet power spectra, as follows:

$$WC = \frac{|S(C_x^*(a,b)C_y(a,b))|^2}{S(|C_x(a,b)|^2) \cdot S(|C_y(a,b)|^2)} \tag{3}$$

where *Cx(a,b)* and *Cy(a,b)* denote the continuous wavelet transforms of *x* and *y* (with x and y indicating time series of an infant's left and right arm movements) at scales *a* in frequency and positions *b* in time. The superscript $*$ is the complex conjugate and *S* is a smoothing operator in time and scale. The dot in the denominator indicates a product between the individual wavelet spectra of both time series. Wavelet coherence has a value between 0 and 1, where 0 means that no coherence is present between signals and 1 means that both signals are fully coherent at any given time and frequency. Wavelet coherence closely resembles a traditional correlation coefficient, and it can be interpreted as a localized correlation coefficient in time-frequency space (Grinsted et al., 2004).

Here, we estimated the wavelet coherence between movements of both hands using the *wcoherence* function in Matlab. To this end, manually annotated episodes of rattling were used to estimate the average duration of each rattling episode and to segment the wearable data (see Figure 2A for an example and Figure 2C for its computed wavelet coherence spectra) and to identify the number of rattling movements using an in-house Matlab script. We estimated rattling movement events following the same approach we used to calculate the clapping events. We categorized the rattling periods as "1" and "0", where 1 indicated movements that were one standard deviation above the mean acceleration and 0 otherwise. Then

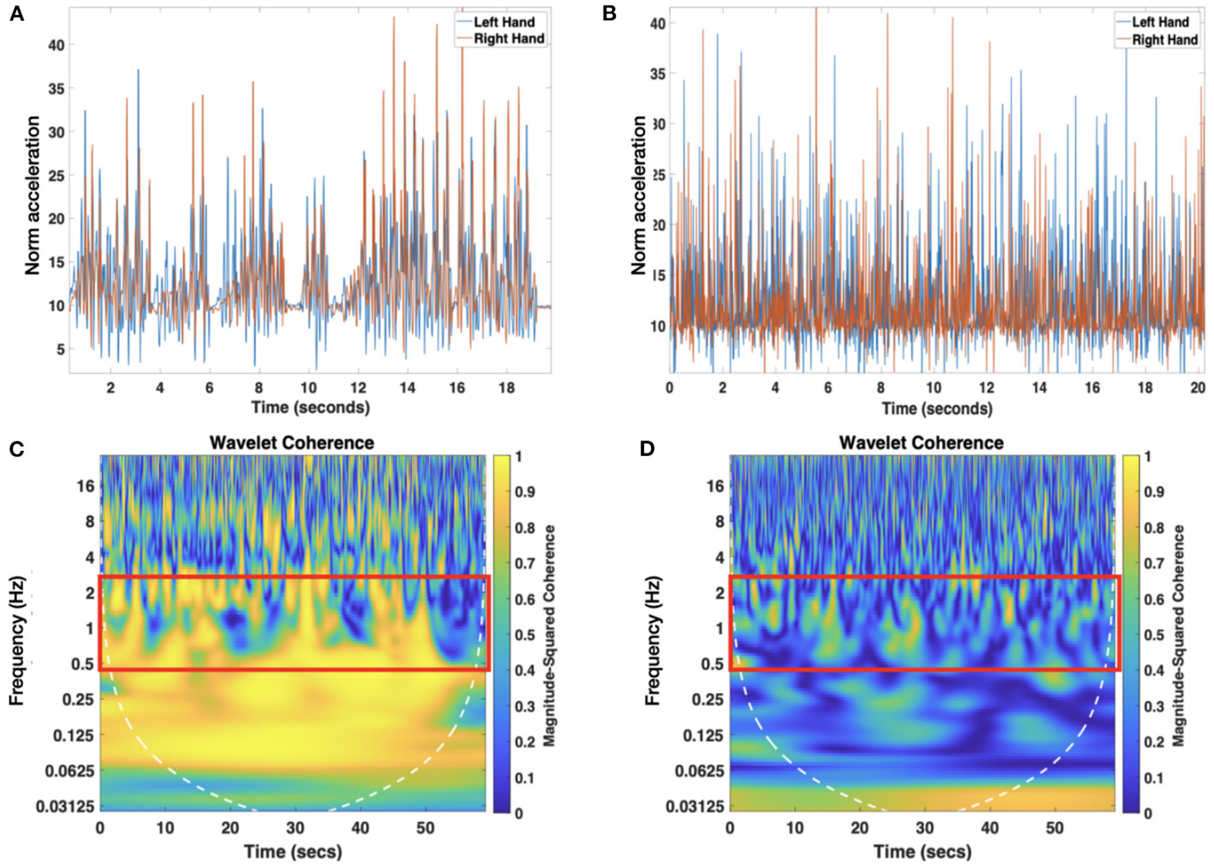


FIGURE 2
Example of the time series created by joining together the rattling episodes of both arms using the manually annotated data **(A)**. Panel **(B)** represents the randomized version of the rattling time series. Only the first 20 s are shown to ease representation. Panels **(C,D)** represent the wavelet coherence spectra of movements of both arms using the original rattling time series and the randomized version, respectively. Highlighted with a red rectangle are the areas where the average wavelet coherence was computed.

we merged those automatically detected movements separated by 50ms or less to avoid artifactual rattling events due to extremely short movements or movements close together. Next, the rattling frequency was calculated as the number of rattling movements divided by the total duration of rattling time derived from the video annotation data (see Table 2 for descriptives). In all but two visits, infant rattling was within the range of 0.5 and 2.5 Hz. Two visits (one at T2 and one at T4) that had the rattling frequency above 2.5 Hz were considered outliers and excluded from the analysis. Given the range of rattling frequencies (see (2) for descriptives), we calculated the average wavelet coherence coefficient within the range of 0.5 and 2.5 Hz for each visit.

Finally, we conducted a control analysis by calculating wavelet coherence between the right and the left arm on the shuffled time-series data from each participant and comparing the mean coherence values of the shuffled data with the original data from all participants. The procedure was iterated 1000 times. This allowed us to show that the wavelet coherence between hand movements did not arise randomly (see Figure 2B for an example of the randomized time series and Figure 2D for its wavelet coherence spectra).

In addition, to investigate developmental changes in movements of a single hand we calculated the continuous wavelet transform spectra (see Supplementary Information 2).

2.8. Statistical analysis

First, to investigate the developmental changes in the number of rattling episodes, their mean duration, the number of rattling movements, the frequency of rattling and the between-hands coherence we ran General Estimating Equations (GEEs) with a Bonferroni correction for pairwise comparisons with age as a repeated measure (T1, T2, T3, T4). Second, to explore potential differences between infants' spontaneous and social rattling, we ran GEEs with age (T1, T2, T3, T4) and condition (Mother Not Providing Rhythm vs. Mother Providing Rhythm) as repeated measures. GEEs are particularly adequate for longitudinal data because they take into account the dependency and ordering of the data within subjects in repeated-measures designs. Data analysis was conducted in IBM SPSS Statistics 26, Figures 1–7 were created using (R Core Team, 2020) and RStudio, version 1.4.1106 (RStudio Team, 2020), and ggplot2 package (Wickham, 2016).

Finally, for control purposes, we run two control analyses. In the first one, we excluded infants who had the lowest numbers of rattling episodes (7 rattling episodes or less) to see whether the infrequent rattlers affected the pattern of results. The significance of all main effects remained unchanged apart from the effect of age on the number of rattling episodes (see Supplementary Information 1 for the full overview). In the second one, we have re-coded our video data to include only those rattling episodes during which infants consecutively performed at least 4 arm movements in a row that produced a rattling sound. Again, the significance of main effects remained unchanged apart from the effect of age on the average duration of a rattling episode (see Supplementary Information 3 for the full overview).

3. Results

3.1. Number of rattling episodes and the average duration of an episode

The number of rattling episodes (annotated periods when an infant was holding at least one rattle and made at least one movement that produced rattling noise) slightly increased with age [Wald $\chi^2(3)= 10.448$, $p = 0.015$, see Figure 3 and Table 2 for descriptive statistics] as the number of episodes increased between T1 and T4 ($p = 0.026$). There was also a main effect of age in the analysis of the average duration of a rattling episode [Wald $\chi^2(3) = 38.450$, $p < 0.001$, see

TABLE 2 Descriptive statistics at each time point.

	T1			T2			T3			T4		
	Mean (SD)	**Min**	**Max**	**Mean (SD)**	**Min**	**Max**	**Mean (SD)**	**Min**	**Max**	**Mean (SD)**	**Min**	**Max**
Number of rattling episodes during play	10.65 (7.16)	3.00	27.00	12.66 (8.39)	2.00	30.00	15.57 (8.31)	2.00	33.00	16.90 (7.11)	4.00	30.00
Mean duration of rattling episode [s]	1.51 (0.85)	0.46	4.20	2.59 (1.14)	0.81	6.03	3.08 (1.25)	0.94	6.14	2.90 (1.30)	1.33	5.59
Number of rattling movements during play	17.77 (10.75)	2.00	39.00	38.69 (33.10)	5.00	141.00	62.54 (31.48)	6.00	124.00	83.60 (45.08)	17.00	195.00
Rattling frequency [Hz]	1.09 (0.38)	0.49	2.43	1.10 (0.28)	0.55	1.72	1.22 (0.31)	0.62	1.93	1.43 (0.34)	0.95	2.18
Wavelet coherence	0.38 (0.12)	0.21	0.73	0.31 (0.10)	0.12	0.58	0.42 (0.13)	0.19	0.79	0.55 (0.15)	0.30	0.78

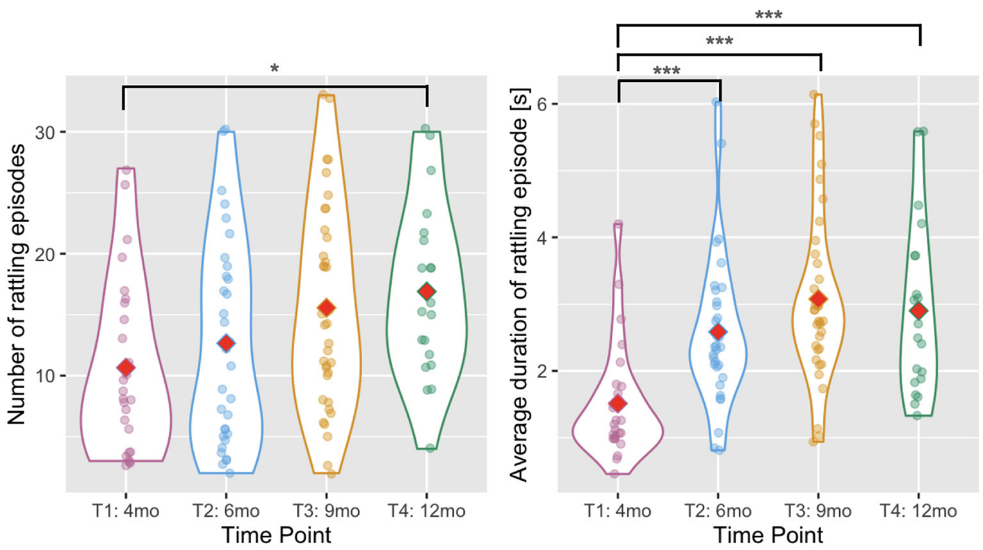


FIGURE 3
Violin plots showing the number of rattling episodes **(left)** and the average duration of rattling episode **(right)** across time points. Red diamonds indicate mean values. A single asterisk indicates significance at $p < 0.05$, two asterisks indicate $p < 0.01$, and three indicate $p < 0.001$.

Figure 3]. The duration was shorter at T1 than at T2, T3, and T4 (all *ps* < 0.001).

3.2. Number of rattling movements

We predicted that infants would be able to produce more rhythmic arm movements with age. To test this hypothesis, we took the number of rattling movements detected automatically in the movement time series during annotated rattling episodes. The number of rattling movements increased with infants' age [Wald $\chi^2(3) = 129.804$, $p < 0.001$, see Figure 4], and pairwise comparisons showed that there were significantly fewer rattling movements at T1 than at T2 ($p = 0.002$), T3 ($p < 0.001$), and T4 ($p < 0.001$); and fewer at T2 than at T3 ($p = 0.018$) and T4 ($p = 0.001$). The difference between T3 and T4 was not significant ($p = 0.465$).

3.3. Rattling frequency

The rattling frequency (i.e., number of rattling movements divided by the total duration of rattling time) increased with infants' age [Wald $\chi^2(3) = 20.498$, $p < 0.001$, see Figure 4] and it was higher at T4 than at T1 ($p = 0.007$) and T2 ($p < 0.001$). The difference between T4 and T3 did not reach significance ($p = 0.058$).

3.4. Between-arms coherence

Average wavelet coherence increased with age [Wald $\chi^2(3) = 49.795$, $p < 0.001$, see Figure 5] between T2 and T3 ($p < 0.001$) and between T3 and T4 ($p = 0.009$). It was higher at T4 than at T1 ($p = 0.001$) or T2 ($p < 0.001$). The difference between T1 and T2 was not significant ($p = 0.224$), similarly there was no difference between T1 and T3 ($p = 0.725$).

3.5. Mother providing vs. not providing rhythm

To explore whether there are any differences in the rhythmic movements that infants produce on their own without an external beat (Mother Not Providing Rhythm) and movements that they produce while being accompanied by their mother through rattling or singing (Mother Providing Rhythm), we have compared these two categories in additional analyses.

3.5.1. Number of rattling episodes and the average duration of an episode across categories

Overall, 69.44% of rattling episodes were classified as Mother Not Providing Rhythm and 30.55% as Mother Providing Rhythm. This proportion was similar across time points (Mother Not Providing Rhythm at T1: 71.03%, T2: 73.02%, T3: 68.92%, T4: 64.53%). Difference in the number of rattling episodes between categories was statistically significant [Wald $\chi^2(1) =$

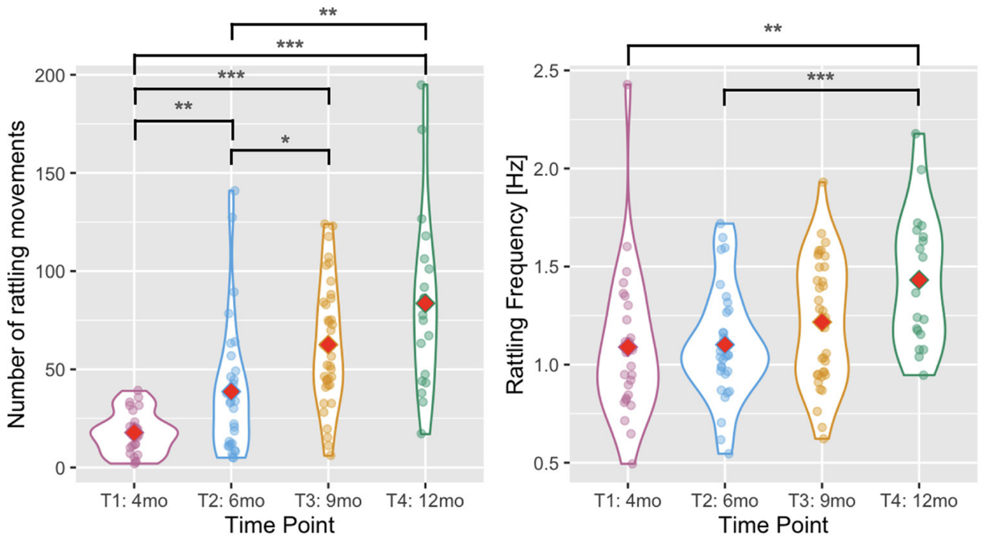


FIGURE 4
Violin plots showing the number of rattling movements **(left)** and the rattling frequency **(right)** across time points. Red diamonds indicate mean values. A single asterisk indicates significance at $p < 0.05$, two asterisks indicate $p < 0.01$, and three indicate $p < 0.001$.

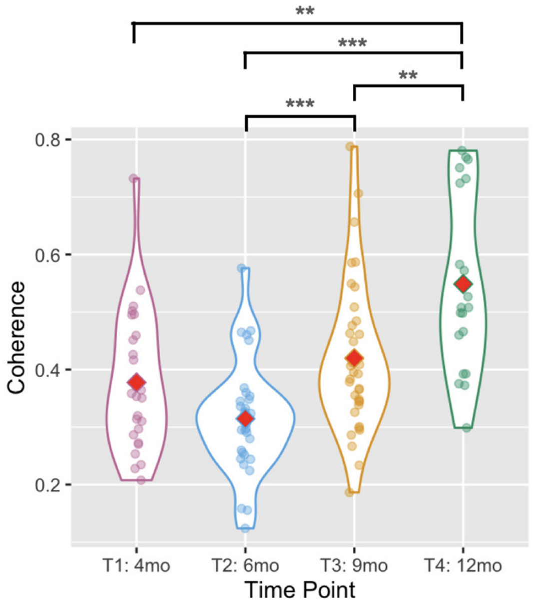


FIGURE 5
Violin plots showing the between-arms coherence. Red diamonds indicate mean values. Two asterisks indicate $p < 0.01$, and three indicate $p < 0.001$.

30.249, $p < 0.001$, see Figure 6]. There was no main effect of age [Wald $\chi^2(3) = 4.893$, $p = 0.180$] and the interaction between age and condition was also not significant [Wald $\chi^2(3) = 6.854$, $p = 0.077$].

In the average duration of a rattling episode (see Figure 6), there was no effect of condition [Wald $\chi^2(1) = 2.072$, $p = 0.150$] and the age × condition interaction was also not significant [Wald $\chi^2(3) = 2.077$, $p = 0.556$]. There was, however, a main effect of age [Wald $\chi^2(3) = 25.112$, $p < 0.001$], as the average duration of a rattling episode was shorter at T1 than at T2 ($p = 0.033$), T3 ($p < 0.001$), and T4 ($p = 0.001$).

3.5.2. Number of rattling movements across categories

There were more rattling movements in the Mother Not Providing Rhythm category than in the Mother Providing Rhythm category [Wald $\chi^2(1) = 17.880$, $p < 0.001$, see Figure 7]. The number of rattling movements increased with infants' age [Wald $\chi^2(3) = 100.117$, $p < 0.001$] as it was higher at T4 than at T1 ($p < 0.001$) and T2 ($p = 0.001$); and higher at T3 than at T1 ($p < 0.001$) and T2 ($p = 0.012$), and higher at T2 than at T1 ($p < 0.001$). There was also a significant interaction effect of age × condition [Wald $\chi^2(3) = 13.548$, $p = 0.004$] as there were more rattling movements in the Mother Not Providing Rhythm condition than in the Mother Providing Rhythm condition at T2 ($p = 0.013$) and T3 ($p = 0.003$).

3.5.3. Rattling frequency across categories

In the rattling frequency (see Figure 7), there was no effect of condition [Wald $\chi^2(1) = 3.578$, $p = 0.059$] and the age × condition interaction was not significant [Wald $\chi^2(3) = 1.860$, $p = 0.602$]. There was a main effect of age [Wald $\chi^2(3) = 25.168$,

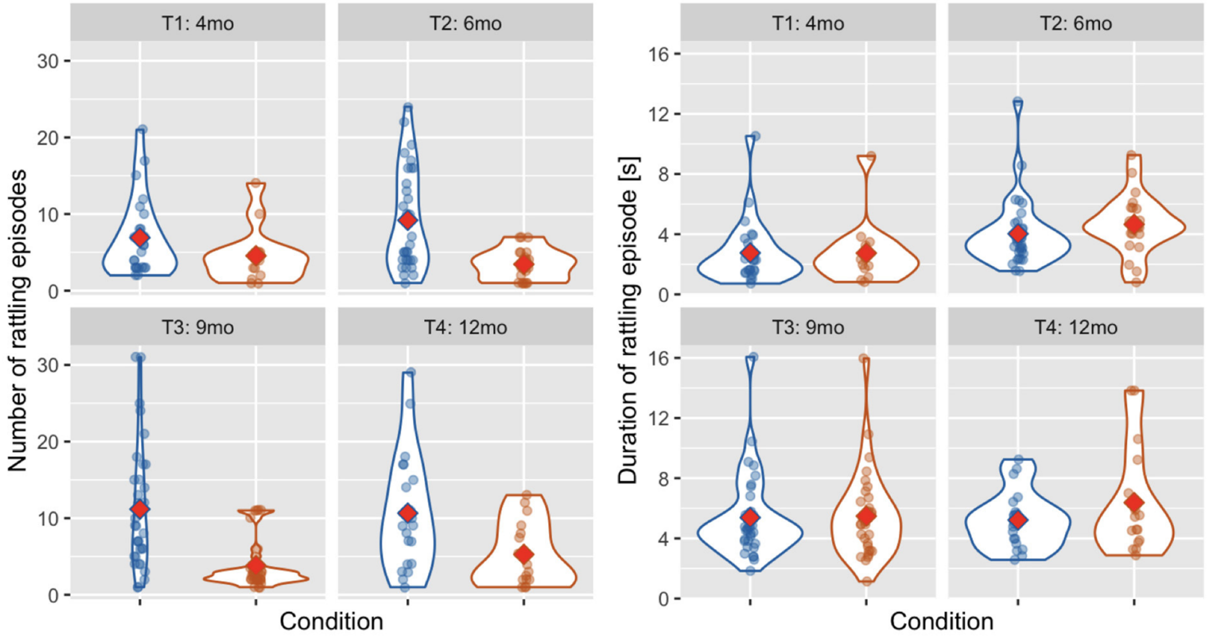


FIGURE 6
Violin plots showing the number of rattling episodes **(left)** and the average duration of rattling episode **(right)** across time points and conditions: mother Not Providing Rhythm (blue) and Mother Providing Rhythm (orange). Red diamonds indicate mean values.

$p < 0.001$] as the rattling frequency was higher at T4 than at T1 ($p < 0.001$), T2 ($p < 0.001$), and T3 ($p = 0.021$).

3.5.4. Between-arms coherence across categories

There was no difference in the average between-arms coherence across conditions [Wald $\chi^2(1) = 0.608$, $p = 0.435$, see Figure 7] and the interaction of age × condition was also not significant [Wald $\chi^2(3) = 6.436$, $p = 0.092$]. The average coherence increased with age [Wald $\chi^2(3) = 38.427$, $p < 0.001$] as it was higher at T4 than at T1 ($p = 0.014$) and T2 ($p < 0.001$); and higher at T3 than at T2 ($p < 0.001$).

3.6. Control comparisons with shuffled time series

To show that the wavelet coherence of between-arm movements did not arise randomly, we conducted a control analysis by calculating wavelet coherence between the right and the left arm on the shuffled time-series data from each participant and comparing the mean coherence values of the shuffled data with the original data from all participants. These comparisons showed that at T2 coordination between both arms is not different from noise [T2: $t_{(32)} = 0.043$; $p = 0.966$]. For T1, T3, and T4 the coherence for observed data was significantly higher than their corresponding shuffled data [T1: $t_{(25)} = 2.555$; $p = 0.017$; T3: $t_{(36)} = 5.800$; $p < 0.001$; T4: $t_{(20)} = 6.904$; $p < 0.001$]. The difference between observed and shuffled data at T1 was not significant in the control analysis with excluded infrequent rattlers (see Supplementary Analysis 1).

4. Discussion

Here we investigated how infants' spontaneous rhythmic behavior in the social context of play with rattles changes across the first year of life. Through precise longitudinal measurements using wearable motion trackers, we show that infants are highly motivated to produce rhythmic manual actions that generate multimodal feedback (rattle-shaking). The number of rattling episodes (periods when the infant was holding at least one rattle and made at least one movement that produced rattling noise) is similar across all visits, suggesting that infants are similarly motivated to attempt rattle-shaking. The mean duration of rattling episodes increases in subsequent months in comparison to the first visit at 4 months as infants' motor control and grasp strength increase. As infants grow older, they also make more rattling movements and their frequency of rattling increases. Furthermore, infants' arm movements become more coupled during rattle-shaking, as shown by the age-related increase in wavelet coherence - and this effect was consistent across control analyses with stricter inclusion criteria. In an additional analysis

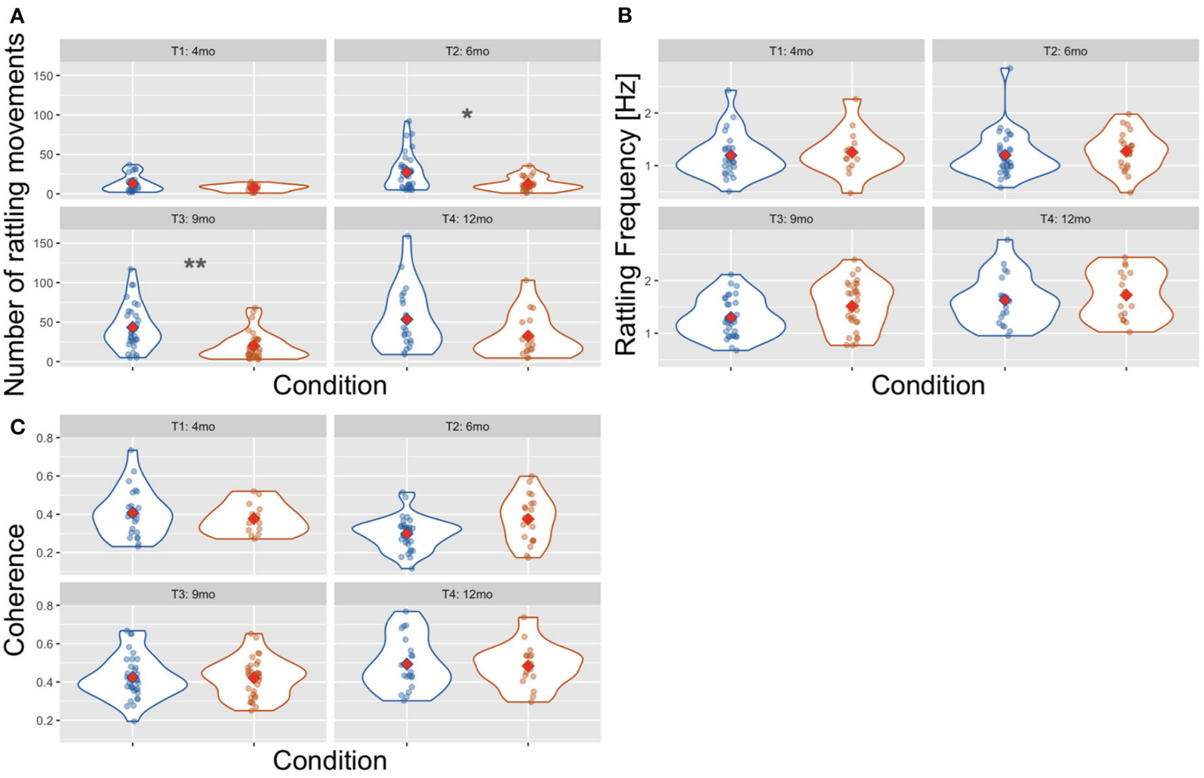


FIGURE 7
Violin plot showing the number of rattling movements **(A)**, the frequency of rattling **(B)**, and the between-arms coherence **(C)** across time points and conditions: Mother Not Providing Rhythm (blue) and Mother Providing Rhythm (orange). Red diamonds indicate mean values, a single asterisk indicates significance at $p < 0.05$, and two asterisks indicate $p < 0.01$.

we also observed the developmental increase in power of wavelet spectra of movements of a single hand, with power being highest in the frequency range between 2 and 3 Hz consistently at all time points. This suggests that across the first year of life it is not the frequency of rattling that changes, but the organization of rhythmicity within the same frequency range. Finally, infants also produced more rattling movements (at the age of 6 and 9 months) when they were rattling alone compared to when mothers provided them with auditory stimulation of rattling or singing. This effect could reflect the caregivers approach to give their infants time for individual exploration of a new interesting objects (rattles) in a novel situation (laboratory). There were no differences between infants' rattling alone and infants' rattling with their mothers in the average between-arms coherence, rattling frequency, or average duration of rattling episode.

In our study, we investigated motor aspects of infants' rattle-shaking in the context of interactions with their mothers. Overall, our results suggest a developmental increase in arm movements during rattle-shaking play with a mother. Younger infants, at 4 and 6 months of age, seem to make fewer rhythmic arm movements which could be explained by their immature motor control (Goldfield, 1995). Motor control at the subcortical level of the central nervous system emerges and matures mainly during the first year of life, allowing for essential trunk stabilization and body positioning, a prerequisite for reaching and grasping arm movements (Westcott et al., 1997; Dusing and Harbourne, 2010; Kobesova and Kolar, 2014), both of which are necessary for the execution of rhythmic rattling. With emerging postural control, arms can also be less involved in stabilizing the body posture and used more in skilled manual reaching (Hadders-Algra, 2005). Our finding of an increase in the frequency of rattling in the second half of the first year of life suggests that older infants can execute rattling movements with more ease. This is in line with a previous study, which recently showed that infants' movements during drumming become faster and more regular with age (Rocha et al., 2021a). We also observed a developmental increase in the infants' between-arms coherence, which shows that arm movements become more coupled during rattle-shaking across the first year of life. On the one hand, this could be explained by the fact that older infants are able to play comfortably in a position and do not need one hand to support themselves while sitting or lying in a prone position. On the other hand, this could be related to an increase in the overall spontaneous rhythmicity of movement. As Hoehl et al. (2021) stated in their review, rhythmic synchronization is usually not limited to a movement of a single limb, but it diffuses throughout the body.

All in all, our results shed more light on the development of the infants' spontaneous rhythmic actions during play with the caregiver. We show that infants are motivated to play with rattles already at 4 months and they keep trying to produce rhythmic arm movements despite constraints related to their limited strength and ability to stay comfortably in a given body position. The number of rattling movements that they produce and the frequency of those movements increase across the first year of life. Similarly, the level of their between-arms coordination also increases with age.

Our findings on spontaneous rhythmic actions during early interactions are highly relevant to the understanding of early underpinnings of conversational skills. First, the motivation to keep producing spontaneous rhythmic movements that result in multimodal feedback may be beneficial for learning about the structure of early social interactions and proto-conversations. Interpersonal communication shows many rhythmic properties. Rhythmic patterns provide information necessary to predict and anticipate the other person's actions (Warner, 1992). Effective communication requires reciprocity, being responsive to the interlocutor (Sebanz et al., 2006) and becoming in-sync on many different levels (e.g., Feldman, 2007; Feldman et al., 2011; Dumas and Fairhurst, 2021). It also requires an understanding of timing to be able to participate in vocal turn-taking (e.g., Gratier et al., 2015). Second, infants' rhythmic movements are considered by parents as communicative signals. Caregivers respond to them frequently, especially when they co-occur with infants' signals from other modalities (vocalizations or gaze toward parents; Moreno-Núñez et al., 2021). Third, rhythmic arm movements are postulated to be the precursor of vocal-entangled gestures that accompany day-to-day adult communication (Pouw and Fuchs, 2022). As was shown in previous studies (Thelen, 1979; Locke et al., 1995; Ejiri, 1998; Ejiri and Masataka, 2001; Iverson and Fagan, 2004; Iverson and Wozniak, 2007; Burkhardt-Reed et al., 2021), rhythmic manual movements often co-occur with infants' vocalizations and this co-occurrence is observed at much earlier developmental stages than other types of gestures—such as pointing (emerging around 12 months of age; Colonnesi et al., 2010; Murillo et al., 2021) or iconic gestures (emerging around 26 months of age, Ozcaliskan and Goldin-Meadow, 2011). Thus, it seems that rhythmic arm movements that accompany vocal learning may serve as a precursor to the gesture-speech system (Iverson and Fagan, 2004; Pouw and Fuchs, 2022).

As rattling is a multimodal signal, it creates a unique opportunity to practice rhythmic arm movements and motor control with encouragement from the parent's side. Caregivers can facilitate infants' rattling by handing in the rattles but also by rattling in-sync or producing an external beat that infants can entrain to (singing or reciting nursery rhymes). This allows infants to practice coordinating their movements with the movements produced by the caregiver, yet these aspects of motor social coordination have been scarcely investigated in infancy (e.g., Scola et al., 2015). Overall, improvements in spontaneous rhythmic production, especially one that provides auditory feedback (such as rattling and drumming), may benefit social communication and in future studies, infants' and caregivers' rattling should be analyzed in the broader context of communicative behaviors.

This study was focused on how infants produce rattling-movements during play with their caregivers. Our set-up was aimed to give participating infant-mother dyads much freedom and to allow them to play in their preferred way. Therefore, we acknowledge several limitations resulting from this trade-off between a more naturalistic play set-up and controlled conditions. We have not controlled infant posture, and our results include rattling episodes produced across many varied body positions (e.g., lying prone and supine, sitting with and without support, standing) and even during locomotion (rattling while walking). Future studies should further investigate whether between-arms coherence during rhythmic actions is dependent on body positioning. Furthermore, mothers were instructed to play with their infants as they usually do, which means they displayed many different behaviors that may have affected infants' rattling on several levels. First, on a low level, caregivers differed in the structural support they offered to their infants (e.g., for wobbly sitters) and in the encouragement toward rattle-shaking (e.g., handing in a rattle when the infant could not reach for it independently). On a higher level, mothers differed in the prompts for joint-play (rattling together at the same time) and in the number of social cues such as singing or reciting baby rhymes. Our exploratory results do not show significant differences between infants' rattling alone and infants' rattling with their mothers in the average between-arms coherence, rattling frequency or average duration of a rattling episode. However, it may not be possible to fully isolate infants' spontaneous rattling activity from the early attempts at rhythmic coordination with a partner within our experimental set-up. Thus, future research is necessary to measure between-arms coherence by directly testing the same dyads with differing task instructions (asking parents to rattle simultaneously vs. not rattle at all).

5. Conclusions

In conclusion, our current findings characterize the longitudinal changes in infants' rhythmic arm movements during rattle-shaking play with their mothers. Infants are similarly motivated to attempt rattle-shaking across the first year of life, but with age, they make more rattling movements and their frequency of rattling increases. Their left and right arm movements become more coupled during rattle-shaking, as shown by the increase in wavelet coherence. Infants also produced more rattling movements and across more rattling episodes in their own rhythm than while being provided with auditory stimulation of mothers' rattling or singing. There were no differences between infants' rattling alone and infants' rattling with their mothers in the average between-arms coherence, rattling frequency or average duration of rattling episode. Our results might shed more light on how spontaneous rhythmic movements can act as precursors of motor social coordination.

Data availability statement

The datasets presented in this article are not readily available because they will be available upon request from the corresponding authors following an embargo period from the date of publication to allow for the finalization of the ongoing longitudinal project. The computer code used in this study is openly available in GitHub: https://github.com/Mirandeitor/frontiersRhythmicPaper. Requests to access the datasets should be directed to ptomalski@psych.pan.pl.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee at the Institute of Psychology, Polish Academy of Sciences. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

KB, AM-K, ZL, AR, and AK collected data. ZL conducted behavioral coding and performed statistical analyses. AK double-coded 20% of videos. DL performed sensor data pre-processing and wavelet coherence analysis. ZL and DL wrote the first draft of the manuscript. PT acquired funding, contributed main theoretical framework, designed and supervised the overall longitudinal study. All authors contributed to conception and design of the study. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

The data collection and analyses, as well as the APC were funded by the Polish National Science Centre grant number 2018/30/E/HS6/00214 to PT.

Acknowledgments

We would like to thank all participating infants and parents for their generous contribution and two reviewers for their valuable comments and suggestions.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.896319/full#supplementary-material

References

(2022). *ELAN Version 6.3.* Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen.

Baruch, C., and Drake, C. (1997). Tempo discrimination in infants. *Infant Behav. Dev.* 20, 573–577. doi: 10.1016/S0163-6383(97)90049-7

Burkhardt-Reed, M. M., Long, H. L., Bowman, D. D., Bene, E. R., and Oller, D. K. (2021). The origin of language and relative roles of voice and gesture in early communication development. *Infant Behav. Dev.* 65, 101648. doi: 10.1016/j.infbeh.2021.101648

Chang, H.-W. and Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child Dev.* 48, 1666. doi: 10.2307/1128532

Coco, M. I., and Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Front. Psychol.* 5, 510. doi: 10.3389/fpsyg.2014.00510

Colonnesi, C., Stams, G. J. J., Koster, I., and Noom, M. J. (2010). The relation between pointing and language development: a meta-analysis. *Dev. Rev.* 30, 352–366. doi: 10.1016/j.dr.2010.10.001

Demany, L., McKenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature* 266, 718–719. doi: 10.1038/266718a0

Dumas, G., and Fairhurst, M. T. (2021). Reciprocity and alignment: quantifying coupling in dynamic interactions. *R. Soc. Open Sci.* 8, *210138*. doi: 10.1098/rsos.210138

Dusing, S. C., and Harbourne, R. T. (2010). Variability in postural control during infancy: implications for development, assessment, and intervention. *Phys. Therapy* 90, 1838–1849. doi: 10.2522/ptj.2010033

Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55, 226–237. doi: 10.1159/000028434

Ejiri, K., and Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Dev. Sci.* 4, 40–48. doi: 10.1111/1467-7687.00147

Feldman, R. (2007). Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *J. Child Psychol. Psychiatry Allied Disciplines* 48, 329–354. doi: 10.1111/j.1469-7610.2006.01701.x

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., and Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behav. Dev*. 34, 569–577. doi: 10.1016/j.infbeh.2011.06.008

Goldfield, E.C. (1995). *Emergent Forms: Origins and Early Development of Human Action and Perception*. Oxford: Oxford University Press.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., and Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Front. Psychol*. 6, 1167. doi: 10.3389/fpsyg.2015.01167

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process. Geophys*. 11, 561–566. doi: 10.5194/npg-11-561-2004

Hadders-Algra, M. (2005). Development of postural control during the first 18 months of life. *Neural Plasticity*. 12, 99–108. doi: 10.1155/NP.2005.99

Háden, G. P., Honing, H., Török, M., and Winkler, I. (2015). Detecting the temporal structure of sound sequences in newborn infants. *Int. J. Psychophysiol.* 96, 23–28. doi: 10.1016/j.ijpsycho.2015.02.024

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., and Hamilton, A. F. C. (2020). Are you on my wavelength? Interpersonal coordination in dyadic conversations. *J. Nonverbal Behav*. 44, 63–83. doi: 10.1007/s10919-019-00320-3

Hannon, E. E., and Trehub, S. E. (2005). Tuning in to musical rhythms: infants learn more readily than adults. *Proc. Natl. Acad. Sci. U.S.A*. 102, 12639–12643. doi: 10.1073/pnas.0504254102

Hoehl, S., Fairhurst, M., and Schirmer, A. (2021). Interactional synchrony: signals, mechanisms and benefits. *Soc. Cogn. Affect. Neurosci*. 16, 5–18. doi: 10.1093/scan/nsaa024

Iverson, J. M., and Fagan, M. K. (2004). Infant vocal – motor coordination: precursor to the gesture – speech system? *Child Dev*. 75, 1053–1066. doi: 10.1111/j.1467-8624.2004.00725.x

Iverson, J. M., and Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *J. Autism Dev. Disord*. 37, 158–170. doi: 10.1007/s10803-006-0339-z

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., and Jasnow, M. D. (2001). Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev*. 66, 1–132. doi: 10.1111/1540-5834.00137

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev*. 83, 323–355. doi: 10.1037/0033-295X.83.5.323

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol*. 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Kobesova, A., and Kolar, P. (2014). Developmental kinesiology: three levels of motor control in the assessment and treatment of the motor system. *J. Bodywork Movement Therap*. 18, 23–33. doi: 10.1016/j.jbmt.2013.04.002

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33, 159–174. doi: 10.2307/2529310

Locke, J., Bekken, K., Mcminnlarson, L., and Wein, D. (1995). Emergent control of manual and vocal-motor activity in relation to the development of speech. *Brain Lang*. 51, 498–508. doi: 10.1006/brln.1995.1073

Moreno-Núñez, A., Murillo, E., Casla, M., and Rujas, I. (2021). The multimodality of infant's rhythmic movements as a modulator of the interaction with their caregivers. *Infant Behav. Dev*. 65:101645. doi: 10.1016/j.infbeh.2021.101645

Murillo, E., Montero, I., and Casla, M. (2021). On the multimodal path to language: the relationship between rhythmic movements and deictic gestures at the end of the first year. *Front. Psychol*. 12, 616812. doi: 10.3389/fpsyg.2021.616812

Ozcaliskan, S., and Goldin-Meadow, S. (2011). "Is there an iconic gesture spurt at 26 months?" in *Integrating Gestures: The Interdisciplinary Nature of Gesture*, eds G. Stam and M. Ishino (Amsterdam: John Benjamins Publishing Company), 163–174. doi: 10.1075/gs.4.14ozc

Pouw, W., and Fuchs, S. (2022). Origins of vocal-entangle gesture. Neurosci. *Biobehav. Rev*. 141:104836. doi: 10.1016/j.neubiorev.2022.104836

Provasi, J., Anderson, D. I., and Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Front. Psychol*. 5, 1048. doi: 10.3389/fpsyg.2014.01048

Provasi, J., and Bobin-Bégue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 21/2- and 4-year-old children. *Int. J. Behav. Dev*. 27, 220–231. doi: 10.1080/01650250244000290

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Richardson, D. C., and Dale, R. (2005). Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cogn. Sci*. 29, 1045–1060. doi: 10.1207/s15516709cog0000_29

Rocha, S., Attaheri, A., Choisdealbha, I., Brusini, P., Flanagan, S. A., Mead, N., et al. (2021a). Infant sensorimotor synchronisation to speech and non-speech rhythms: a longitudinal study. *psyArXiv [Preprint]*. doi: 10.31234/osf.io/jbrga

Rocha, S., and Mareschal, D. (2017). Getting into the groove: the development of tempo-flexibility between 10 and 18 months of age. *Infancy* 22, 540–551. doi: 10.1111/infa.12169

Rocha, S., Southgate, V., and Mareschal, D. (2021b). Infant spontaneous motor tempo. *Dev. Sci*. 24:e13032. doi: 10.1111/desc.13032

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Scola, C., Bourjade, M., and Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffect. Neurosci. Psychol*. 5, 28256. doi: 10.3402/snp.v5.28256

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci*. 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sloetjes, H., and Wittenburg, P. (2008). "Annotation by category - ELAN and ISO DCR," in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)* (Marrakech: European Language Resources Association), 816–820.

Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Anim. Behav*. 27(Pt 3), 699–715. doi: 10.1016/0003-3472(79)90006-X

Thelen, E. (1981). Kicking, rocking, and waving: contextual analysis of rhythmical stereotypies in normal human infants. *Anim. Behav*. 29, 3–11. doi: 10.1016/S0003-3472(81)80146-7

Trevarthen, C. (1979). "Communication and cooperation in early infancy: A description of primary intersubjectivity," in *Before Speech: The Beginning of Human Communication*, ed M. Bullowa (Cambridge: Cambridge University Press), 321–347.

Warner, R. M. (1992). Cyclicity of vocal activity increases during conversation: support for a nonlinear systems model of dyadic social interaction. *Behav. Sci*. 37, 128–138. doi: 10.1002/bs.3830370204

Westcott, S. L., Lowes, L. P., and Richardson, P. K. (1997). Evaluation of postural stability in children: current theories and assessment tools. *Phys. Therapy* 77, 629–645. doi: 10.1093/ptj/77.6.629

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U.S.A*. 106, 2468–2471. doi: 10.1073/pnas.0809035106

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: a study of 18- to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/infa.12403

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A*. 107, 5768–5773. doi: 10.1073/pnas.1000121107

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 21 October 2022
DOI 10.3389/fpsyg.2022.944670

OPEN ACCESS

EDITED BY
Stefanie Peykarjou,
Heidelberg University, Germany

REVIEWED BY
Claire Kabdebon,
UMR 8554 Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), France
Amirhossein Ghaderi,
York University, Canada

*CORRESPONDENCE
Chiara Cantiani
chiara.cantiani@lanostrafamiglia.it

SPECIALTY SECTION
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

RECEIVED 15 May 2022
ACCEPTED 22 September 2022
PUBLISHED 21 October 2022

CITATION
Cantiani C, Dondena C, Molteni M, Riva V and Piazza C (2022) Synchronizing with the rhythm: Infant neural entrainment to complex musical and speech stimuli. *Front. Psychol.* 13:944670. doi: 10.3389/fpsyg.2022.944670

Synchronizing with the rhythm: Infant neural entrainment to complex musical and speech stimuli

Chiara Cantiani[1]*, Chiara Dondena[1], Massimo Molteni[1], Valentina Riva[1] and Caterina Piazza[2]

[1]Child Psychopathology Unit, Scientific Institute, IRCCS Eugenio Medea, Lecco, Italy, [2]Bioengineering Lab, Scientific Institute, IRCCS Eugenio Medea, Lecco, Italy

Neural entrainment is defined as the process whereby brain activity, and more specifically neuronal oscillations measured by EEG, synchronize with exogenous stimulus rhythms. Despite the importance that neural oscillations have assumed in recent years in the field of auditory neuroscience and speech perception, in human infants the oscillatory brain rhythms and their synchronization with complex auditory exogenous rhythms are still relatively unexplored. In the present study, we investigate infant neural entrainment to complex non-speech (musical) and speech rhythmic stimuli; we provide a developmental analysis to explore potential similarities and differences between infants' and adults' ability to entrain to the stimuli; and we analyze the associations between infants' neural entrainment measures and the concurrent level of development. 25 8-month-old infants were included in the study. Their EEG signals were recorded while they passively listened to non-speech and speech rhythmic stimuli modulated at different rates. In addition, Bayley Scales were administered to all infants to assess their cognitive, language, and social-emotional development. Neural entrainment to the incoming rhythms was measured in the form of peaks emerging from the EEG spectrum at frequencies corresponding to the rhythm envelope. Analyses of the EEG spectrum revealed clear responses above the noise floor at frequencies corresponding to the rhythm envelope, suggesting that – similarly to adults – infants at 8 months of age were capable of entraining to the incoming complex auditory rhythms. Infants' measures of neural entrainment were associated with concurrent measures of cognitive and social-emotional development.

KEYWORDS

neural entrainment, infants, music, speech, EEG, steady-state evoked potentials (SS-EP)

Introduction

When listening to music, most people spontaneously move or clap following the rhythm. This happens because we are sensitive to periodic metric pulses, referred to as beats. Interestingly, even young children and preverbal infants are sensitive to rhythm and are able to detect the beat in music (Phillips-Silver and Trainor, 2005; Winkler et al., 2009). In the first year of life, infants already display spontaneous engagement and increased rhythmic movements and vocalizations in response to music and rhythmic patterns, although these movements are still not yet temporally modulated by the beat (Zentner and Eerola, 2010; Fujii et al., 2014; Rocha and Mareschal, 2017). These skills typically develop during the second year of life, with children starting to exhibit some tempo flexibility and synchronization (Zentner and Eerola, 2010; Fujii et al., 2014; Yu and Myowa, 2021). Further improvements are seen at older ages (Kirschner and Tomasello, 2009; Yu and Myowa, 2021). In the present study, we investigate neural entrainment to complex musical and speech stimuli in a sample of 8-month-old infants, in order to explore whether infants at this young age are able to synchronize at the neural level with the beat/meter of the stimuli or at least with the perceptual elements detectable in the sound.

Increased evidence suggests that the capacity for behavioral rhythmic sensorimotor synchronization is supported by neural entrainment mechanisms, such as cortical oscillations at specific frequencies. Neural entrainment is defined as the process whereby brain activity, and more specifically neuronal oscillations measured by electroencephalography (EEG), synchronizes with external (exogenous) stimulus rhythms. In the auditory modality, neural entrainment has been mainly investigated in response to two types of dynamic and rhythmic stimuli: speech and music (e.g., Nozaradan et al., 2011, 2012, 2018; Ding and Simon, 2012; Nozaradan, 2014; Tierney and Kraus, 2015; Ding et al., 2016; Zhou et al., 2016; Stupacher et al., 2017; Tal et al., 2017; Lenc et al., 2018; Jin et al., 2020). In these studies, low-frequency (< 6 Hz) neural entrainment has been reliably observed for both physical and abstract properties of the stimuli, such as the rhythms of musical beats and some linguistic constituents. For example, in response to simple musical rhythms Nozaradan et al. (2011) recorded entrained responses both to the existing beat frequency and to an imaginary meter frequency. Similarly, in response to speech stimuli, neural entrainment is typically recorded not only to low-level linguistic units but also to phrasal and sentential units (Ding et al., 2016, 2017). These pieces of evidence suggest that low-frequency neural entrainment may play a role not only in tracking speech and rhythmic sounds but also in parsing their temporal structures and extracting high-level chunks (Jin et al., 2020). They also suggest that neural entrainment might not be exclusively stimulus-driven, but additionally includes the contribution of an internally generated (endogenous) oscillator synchronized to the exogenous stimulation (Tal et al., 2017) and with a functional role in its processing (Obleser and Kayser, 2019).

The use of periodic sequences of stimulation and the recording of Steady-State Evoked Potentials (SS-EPs) has several advantages for the study of early neurocognitive development in infancy, as recently argued by Kabdebon et al. (2022). For example, practical advantages are related to the possibility of using the continuous presentation of several stimuli in a relatively short recording session. Importantly, it offers an objective definition of the targeted responses: based on the frequency of the stimulation, SS-EPs are expected at a specific narrow frequency band (Zhou et al., 2016). Previous studies on newborns and infants applied frequency-tagging paradigms with periodic visual stimulation to investigate face and object processing (e.g., de Heering and Rossion, 2015; Peykarjou et al., 2017; Buiatti et al., 2019), and with rhythmic speech sounds to investigate the tracking of transitional probabilities in the context of artificial grammar learning (Kabdebon et al., 2015; Choi et al., 2020; Fló et al., 2022).

Despite the growing interest in this approach, the synchronization of EEG oscillations with complex auditory exogenous rhythms are still relatively unexplored in infants. Using the same paradigm and methodology proposed by Nozaradan et al. (2011), a recent study has measured infants' neural entrainment to simple rhythmic patterns (i.e., tone sequences), showing SS-EPs at a frequency corresponding to both beat and meter in infants as young as age 7 months (Cirelli et al., 2016). This study additionally provided preliminary evidence that such neural responses can be influenced by infant individual differences and their early musical experiences. To our knowledge, this approach including a frequency-tagging paradigm has never been applied to complex musical/speech stimuli in infancy.

In the present study, we investigated neural entrainment to complex music and speech stimuli in a sample of 8-month-old infants. Two different music rhythm patterns and a nursery rhyme with a regular and rhythmic pattern were selected based on their temporal envelope and their frequency spectrum of acoustic energy. Neural entrainment to the incoming rhythms was measured in the form of SS-EPs at frequencies corresponding to the rhythm envelope. By means of this methodology, the present study has a threefold aim. First and foremost, we aimed to investigate whether infants' neural entrainment, already reported for simple rhythmic stimuli (Cirelli et al., 2016), is also present in response to much more complex stimuli. Second, since these stimuli were never used before, we piloted the experiment on adults: here we provide a developmental analysis to explore potential similarities and differences between infants' and adults' ability to entrain to the stimuli. Third, we analyzed the associations between infants' measures of neural entrainment and the concurrent level of linguistic, cognitive and social-emotional development. This last aim rises from recent theoretical frameworks proposing entrainment of neural oscillations to external rhythmic stimuli as one of the mechanisms underlying speech and language (a)typical development (Ladányi et al., 2020; Fiveash et al., 2021). We expect to see in infants at least some sort of initial capability to entrain to the incoming complex auditory music and speech rhythms, even with some differences with respect to adults (i.e., more robust entrainment in adults than infants). In order to quantify the individual degree of entrainment, for each stimulus, we selected frequencies of interest based on the frequency spectrum of the stimulus sound envelope. As described in detail in **Figure 1**, the selected peaks corresponded to the fastest elements detectable in each stimulus and to more abstract chunks, corresponding to the beat and the meter. Finally, we expect that individual differences in the ability to entrain to these rhythmic features could be related to the overall infant developmental level, and specifically to language skills.

Materials and methods

Participants

Thirty-three families participated in this study. Infants were recruited at 6 months of age via local advertisements as part of a larger longitudinal study. The study was approved by the Medea Institute's Scientific and Ethical Committees and all parents gave their written consent prior to testing. Infants were included if (1) both parents were Italian native speakers, (2) gestational age was ≥35 weeks and birth-weight was ≥2000 grams, (3) first-degree relatives had no certified diagnosis of intellectual deficiency or neurodevelopmental disorders. Data for this study were collected between 7 and 9 months of age. Since data from 8 children were rejected due to insufficient artifact-free trials in the EEG task, the final sample consisted of 25 infants (12 males; mean age = 8.00 months; in days, $M = 240.16$, SD = 19.13, min = 208, max = 279). Sample size was determined based on the previous SS-EP literature in similar populations (Kabdebon et al., 2015; Cirelli et al., 2016; Peykarjou et al., 2017; Choi et al., 2020; Fló et al., 2022). Information about parents' and children's daily exposure to music was collected, in order to better characterize the sample from this point of view. Mothers reported to spend on average 62.4 min per day (SD = 49.8) listening to music together with their child, while fathers reported to do it for 13.9 min on average per day (SD = 16.7). Moreover, 94.1% of the mothers and 62.5% of the fathers reported to be used to moving and "dancing" with their child following the rhythm of the music.

In addition to infants, a sample of 10 adults (8 females, mean age = 25.2 years; SD = 1.2; two left-handed) was recruited to provide a developmental comparison of neural entrainment between adults and infants.

Clinical assessment

The Bayley Scales of Infant and Toddler Development – Third Edition (Bayley, 2006) were used as neuropsychological assessment. For the purpose of this study, the Cognitive, Language, and Social-emotional scales were administered. Considering the age range of our interest, the Cognitive scale mainly investigates sensory-motor development and the ability to explore and manipulate objects; the Language scale investigates receptive communication skills, such as the infants' pre-verbal behavior and verbal comprehension, and expressive skills, such as pre-verbal communication (e.g., babbling and gestures) and early speech. The Social-emotional questionnaire is filled out by the caregiver and investigates the child's interest and responsiveness toward people, objects and sounds and their communicative behavior. Scaled scores ($M = 10$; SD = 3) were calculated based on age.

Electrophysiological recording

Stimuli

The stimuli consisted of three rhythmic patterns: two of them were instrumental music modulated at different rates and one was an Italian nursery rhyme with a regular and rhythmic pattern (mean length was 37.19 s, SD = 0.66). **Figure 1** shows a graphical representation of the three different stimuli that were used.

The first music stimulus, henceforth called "African," was created by extracting two meters of an African 4/4 rhythm (see **Figure 1A**), the average tempo was 104 beats per minute (bpm). Trials of approximately 37 s were created using Praat software package by repeating the extracted meters for eight times.

The second music stimulus, henceforth called "Tabla", was created by extracting four meters of a Tabla composition (see **Figure 1B**), the rhythm was 2/4 and the average tempo was 101 bpm. The extracted segment was repeated 8 times in order to obtain the 37-s-long clip.

The "speech" stimulus was an eight-verse excerpt of an Italian nursery rhyme ("Girotondo di tutto il mondo," Rodari, 1960). The rhyme was recorded by a native Italian female speaker; it was regularly spoken with exaggerated accents following a metronome set on 104 bpm in order to match the tempo of the first musical stimulus (African), see **Figure 1C**. The whole excerpt was repeated two times in order to obtain the 37-s-long clip.

Each 37-s-long trial was repeated 5 times in a pseudo-randomized order (no more than two trials of the same type appeared consecutively).

The temporal envelope of the three rhythm patterns was extracted using the Hilbert function implemented in MATLAB (Mathworks). A Fast Fourier Transform (FFT) was then applied in order to compute the frequency spectrum of acoustic energy.

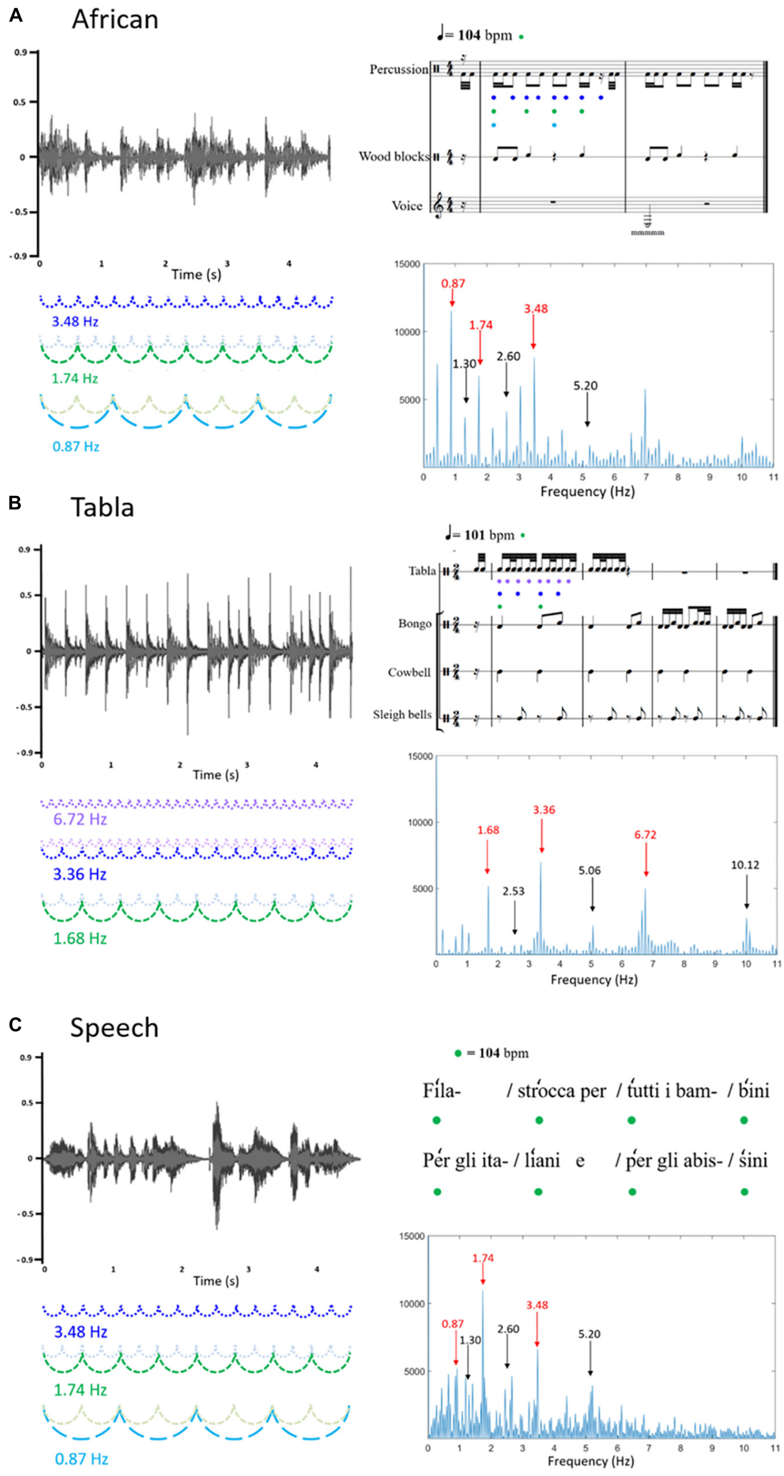


FIGURE 1

Graphical representation of the three sound stimuli: **(A)** African musical stimulus; **(B)** Tabla musical stimulus; **(C)** Speech stimulus. On the left, the three sound waveforms are reported (x-axis: time; y-axis: sound amplitude). On the right, the upper figures represent the rhythmic patterns, depicted in musical notation **(A,B)** or speech metric structure **(C)**; the bottom figures represent the frequency spectrums of the stimulus sound envelopes (*x*-axis: frequency; *y*-axis: magnitude). In the frequency spectrums, peaks at frequencies corresponding to the rhythm envelope are reported in red, whereas peaks at frequencies not corresponding to the rhythm envelope, considered as noise floor, are reported in black. Frequencies corresponding to the rhythm envelope are represented on the left, coupled with the sound waveforms, and on the upper right, coupled with musical notation or speech metric structure. In detail, it can be noted that for the African musical stimulus **(A)** frequency 3.48 Hz

(Continued)

FIGURE 1 (Continued)
(represented in blue) corresponds to the fastest element detectable in the stimulus (matching with the quaver), frequency 1.74 Hz (represented in green) corresponds to the beat (104 bpm, matching with the crotchet), frequency 0.87 Hz (represented in light blue) corresponds to a chunk of notes (matching with the minim); for the Tabla musical stimulus **(B)** frequency 6.72 Hz (represented in lilac) corresponds to the fastest element detectable in the stimulus (matching with the semiquaver), frequency 3.36 Hz (represented in blue) corresponds to a chunk of notes (matching with the quaver), frequency 1.68 Hz (represented in green) corresponds to the beat (101 bpm, matching with the crotchet); the speech stimulus **(C)** was recorded to match with the African musical stimulus, with the metronome used to keep the beat set at 104 bpm (corresponding to frequency 1.74 Hz, represented in green).

As represented in **Figure 1**, the frequency spectrum of acoustic energy varied across stimuli. For the African musical stimulus, frequencies of interest were: 3.48 Hz, corresponding to the fastest element detectable in the stimulus (matching with the quaver); 1.74 Hz, corresponding to the beat (104 bpm, matching with the crotchet); 0.87 Hz, corresponding to a chunk of notes (matching with the minim). For the Tabla musical stimulus, frequencies of interest were: 6.72 Hz, corresponding to the fastest element detectable in the stimulus (matching with the semiquaver); 3.36 Hz, corresponding to a chunk of notes (matching with the quaver); and 1.68 Hz, corresponds to the beat (101 bpm, matching with the crotchet). Finally, the speech stimulus was recorded with the beat set at 104 bpm, corresponding to frequency 1.74 Hz. Peaks at 0.87 Hz and 3.48 Hz were additionally selected as frequencies of interest to match this stimulus with the African musical stimulus.

Procedure

During EEG recording, infants were seated on their caregiver's lap in a sound-attenuated and electrically shielded room and were entertained with silent toys. Both caregiver and the experimenter were given sound-isolating earphones which prevented them from hearing the auditory stimuli. In addition, the experimenter was instructed to make non-rhythmic movements while playing. The whole recording lasted approximately 10 min.

Data acquisition and analysis

Infants' data

Electroencephalography (EEG) was recorded using a 60-channel HydroCel GSN net with an EGI recording system (Electric Geodesic, Inc.). Signals were recorded online with a sampling rate of 250 Hz, vertex as the online reference, and a 0.1–100 Hz online bandpass filter.

After recording, EEG data were exported to a MATLAB (Mathworks) compatible format and processed using EEGLAB (Delorme and Makeig, 2004) and the automated standardized pipeline MADE (Debnath et al., 2020). First, the 13 outermost channels were removed from analysis (Cantiani et al., 2016). Second, continuous EEG data were high-pass filtered at 0.5 Hz and low-pass filtered at 20 Hz. Third, bad channels were identified and removed using the EEGLAB plug-in FASTER (Nolan et al., 2010). As a fourth step, independent component analysis (ICA) was used to identify non-neural artifacts (e.g., ocular artifacts), and generic noise. As described in Debnath et al. (2020), the ICA procedure adopted in the MADE pipeline - including preparing a copy of the dataset by applying a 1 Hz high-pass filter, segmenting data into arbitrary 1-s epochs, removing noisy epochs, running ICA on the copied dataset and then copying ICA weights back to the original dataset - is an improved ICA decomposition that does not sacrifice low-frequency information or data that contain excessive noise. The Adjusted-ADJUST algorithm (Leach et al., 2020) was used to remove artifactual independent components (the average number of removed components was 5.16; SD = 3.13; range: 1–12). After ICA, data were epoched into 9-s segments starting from the onset of each stimulus, thus yielding 20 epochs for each stimulus. According to previous literature (Nozaradan et al., 2011; Cirelli et al., 2016), we removed the first second of each epoch to avoid the auditory evoked-potentials elicited by stimulus onset and because we expect entrainment to occur only after a few cycles of stimulation. Final epochs ranged then from + 1000 to + 9000 ms, with the baseline defined between 900 and 1000 ms. Residual artifacts were identified and treated as follows: epochs containing residual ocular artifacts (amplitude exceeding ±150 μV on frontal channels AF4, F2, Fz, Afz, F1, AF3) were removed, whereas all other non-ocular channels exceeding ±150 μV in a particular epoch were interpolated (within that particular epoch) using a spherical spline interpolation. Epochs with more than 10% of interpolated channels were rejected. Finally, data were re-referenced to the average of all the electrodes. Participants needed to have at least seven artifact-free epochs per stimulus to be included for analysis (excluded participants $N = 8$). The average number of artifact-free epochs per stimulus was: African ($M = 15.56$, SD = 2.68, range: 10–20), Tabla ($M = 14.60$, SD = 3.29, range: 9–20), Speech ($M = 13.92$, SD = 3.94, range: 7–20). A paired t-test revealed a significant difference in the number of artifact-free epochs for African vs. Speech [$t(24) = 2.686$, $p = 0.013$, FDR-adjusted $p = 0.0390$]. The average number of globally interpolated channels was 1.84 (SD = 0.89; range: 0–4).

Artifact-free trials were averaged and FFT was applied using Letswave6 (Mouraux and Iannetti, 2008). In order to remove the unrelated residual background noise, the magnitude of SS-EPs was calculated in relation to the amplitude of the frequency spectrum at surrounding bins. This was managed by extracting z-scores (i.e., the standard deviation relative to the distribution of the reference interval), considering neighboring bins from

–0.15 to –0.09 Hz and +0.09 to +0.15 Hz around each frequency bin, corresponding to –5 to –3 and +3 to +5 bins around each frequency bin. Z-scores were applied instead of other baseline correction procedures (i.e., subtraction) because they are more easily interpretable (i.e., given the value it is possible to determine whether entrainment has occurred or not). As a further step, for each frequency peak of interest determined from the sound stimuli FFT (see section "Stimuli"), SS-EP magnitudes were extracted within a 0.2 Hz band centered on the frequency of interest (for example, for frequency 1.74, SS-EP magnitude was extracted considering the frequency range 1.64–1.84). In order to determine whether entrainment occurred or not at the specific frequencies of interest, we used two approaches. First, we interpreted z-scores as measures that include an estimate of the noise floor at the bins surrounding the frequency of interest (Peykarjou et al., 2017). Second, following Cirelli et al. (2016) we also calculated SS-EP magnitudes for frequencies that we did not expect to be relevant, based on the sound stimuli FFT. To select such frequencies for each sound stimulus, we computed the median between the first two frequency peaks of interest, and then computed the relative harmonics. Amplitudes at these peaks (three for each sound stimulus) are shown in **Figure 1**, and were considered as noise floor amplitudes. We believe that the results obtained applying this second method are even clearer and more interpretable than those obtained by the mere interpretation of z-scores, especially considering that the frequency stimulation with our stimuli cannot be as precise as the frequency stimulation with visual or less complex auditory stimuli.

SS-EP magnitudes computed for each channel were averaged in three clusters corresponding to left (channels AF3, F3, F5, FC5, F7, FT7), midline (channels F2, Fz, Afz, F1) and right (channels AF4, FT8, FC6, F8, F6, F4) frontal areas which were used in the analysis. Channels of interest were defined based on the previous ERP literature suggesting that frontal areas are involved in infants' auditory processing (e.g., Choudhury and Benasich, 2011; Van Zuijen et al., 2012; Cantiani et al., 2016).

Adults' data

The acquisition procedure was kept identical for adult participants, with the only difference concerning the use of 128-channel HydroCel GSN nets. Analytic procedures were also kept identical: the automated standardized pipeline MADE (Debnath et al., 2020) was used, since good performances of the included tools have been reported for adults as well (Leach et al., 2020). The only differences in the analytic procedures included: (1) no channels were removed from analysis *a priori*; (2) residual ocular artifacts were quantified as amplitude exceeding ±150 μV on frontal channels F10, AF8, AF4, FP2, FPZ, Afz, FP1, AF7.

The average number of adults' artifact-free epochs per stimulus was: African (*M* = 18.10, SD = 3.31, range: 9–20), Tabla (*M* = 18.30, SD = 2.87, range: 11–20), Speech (M = 18.30, SD = 3.37, range: 9–20). No differences emerged between stimuli in the number of artifact-free epochs. The average number of globally interpolated channels was 7.10 (SD = 2.88; range: 1–10), whereas the average number of components removed following the ICA procedure was 31; SD = 7.75; range: 19–45.

Steady-state evoked potentials (SS-EP) magnitudes computed for each channel were averaged in three clusters corresponding to left (channels AF3, F3, F5, FC5, F7, FT7), midline (channels F2, Fz, Afz, F1) and right (channels AF4, FT8, FC6, F8, F6, F4) frontal areas which were used in the analysis.

Statistical analyses

Statistical analyses were run separately for infants' and adults' data. For each stimulus, we first ran a repeated measure ANOVA including noise (peaks of interests vs. noise floor), frequency (three levels for each stimulus), and laterality (left, midline, right) as within-subject factors. Greenhouse-Geisser-corrected p-values are reported when appropriate. Second, we ran paired-samples *t*-tests to compare SS-EP magnitudes at each of the frequencies contained in the sound stimulus and the most adjacent frequency not contained in the sound stimulus (noise floor): this follow-up analysis had the goal of identifying at which specific frequency entrainment occurred. Finally, for infants only, associations between SS-EP magnitudes at frequencies of interest significantly different from noise floor and concurrent measures of Cognitive, Language, and Socio-Emotional development were assessed using Pearson's correlations. False discovery rate (FDR; Benjamini and Yekutieli, 2001) was applied to correct for multiple comparisons in the paired-samples *t*-tests (three comparisons for each stimulus) and Pearson's correlations (twelve comparisons).

Results

Infants' data

Group-level average SS-EPs for the three sound stimuli are shown in **Figure 2**. Descriptive statistics are reported in **Table 1**. SS-EP magnitudes significantly different from background noise are highlighted for both corrected ($\alpha < 0.005$) and uncorrected ($\alpha < 0.05$) thresholds.

The ANOVA carried out on the African musical stimulus revealed a main effect of noise, $F(1,24) = 21.681$, $p < 0.001$, $\eta^2 = 0.475$, a main effect of frequency, $F(2,48) = 11.071$, $p < 0.001$, $\eta^2 = 0.316$, and a significant interaction frequency × noise $F(2,48) = 4.112$, $p = 0.035$, $\eta^2 = 0.146$. Overall, as shown in **Figure 2A**, amplitudes relative to peaks of interest were higher than amplitudes relative to noise floor, and amplitudes were higher at higher frequencies than lower frequencies. Since we did not find any significant interactions including laterality, frequency, and noise, paired-samples

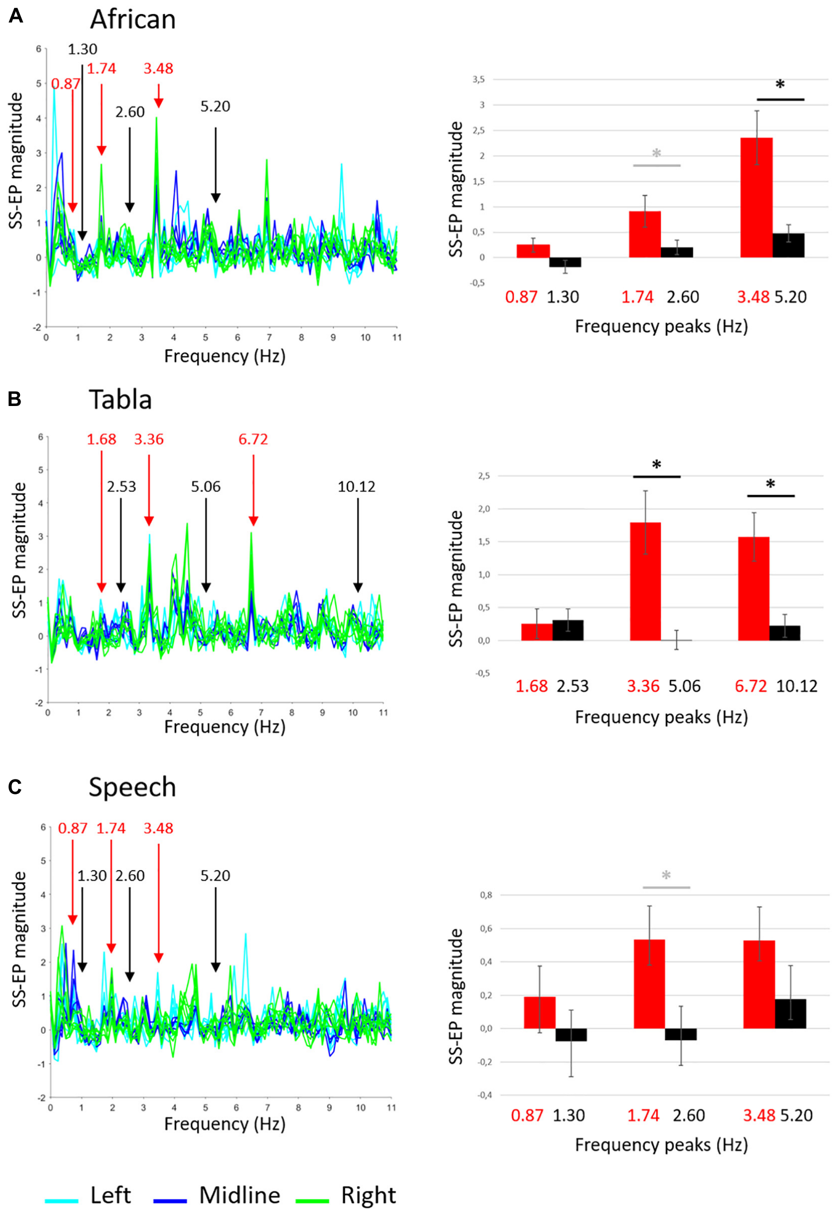


FIGURE 2
Group-level average SS-EPs for the three sound stimuli: **(A)** African musical stimulus; **(B)** Tabla musical stimulus; **(C)** Speech stimulus. On the left, butterfly plot of SS-EPs obtained applying the Signal-to-Noise correction (*x*-axis: frequency; *y*-axis: SS-EP magnitude). All EEG channels included in the clusters are plotted (Left: light blue lines, Midline: blue lines, Right: green lines). Peaks at frequencies corresponding to the rhythm envelope are highlighted in red, whereas peaks at frequencies not corresponding to the rhythm envelope, considered as noise floor, are highlighted in black. On the right, bar graphs (error bars indicate SEM) show the averaged SS-EP magnitude. Significant paired *t*-tests are indicated by black asterisks, whereas differences not surviving False discovery rate (FDR) correction are indicated by gray asterisks.

t-tests were carried out to compare SS-EP magnitudes at each frequency of interest vs. noise floor averaging the three electrode clusters. Paired *t*-tests revealed that the difference between amplitudes in peaks of interests vs. noise floor was significant at the 3.48 Hz peak, $t(24) = 3.610$, $p = 0.0030$ (FDR-adjusted *p*-value), whereas the difference at the 1.74 Hz peak did not survive correction [$t(24) = 2.148$, $p = 0.042$, FDR-adjusted $p = 0.0630$]. These results suggest that infants reliably entrain to the fastest element detectable in the African stimulus (the quaver).

The ANOVA carried out the Tabla musical stimulus revealed a main effect of noise, $F(1,24) = 17.533$, $p < 0.001$, $\eta^2 = 0.422$, and a significant interaction of frequency × noise, $F(2,48) = 6.240$, $p = 0.010$, $\eta^2 = 0.206$. As shown in Figure 2B, paired *t*-tests averaging the three electrode clusters reveal that the difference between amplitudes in peaks of interests vs.

TABLE 1 Descriptive statistics: Mean (Standard Deviation) on SS-EP magnitudes for frequencies of interest vs. noise floor separated for sound stimuli and cluster of channels.

	Peaks on the frequencies of interest				Noise floor amplitudes			
		Left	Midline	Right		Left	Midline	Right
	Hz	M (SD)	M (SD)	M (SD)	Hz	M (SD)	M (SD)	M (SD)
African	0.87	0.47 (1.24)	0.27 (1.18)	−0.12 (0.56)	1.30	−0.24 (0.56)	−0.09 (1.48)	−0.21 (0.72)
	1.74	0.90 (2.27)	0.76 (2.06)	1.07 (1.96)	2.60	0.24 (1.18)	−0.82 (0.81)	0.45 (1.38)
	3.48	**2.03** (2.25)	**1.96** (3.23)	**3.07*** (4.57)	5.20	0.44 (1.29)	0.26 (1.49)	0.73 (1.09)
Tabla	1.68	0.36 (1.69)	0.19 (1.68)	0.21 (0.91)	2.53	0.22 (1.08)	0.34 (1.46)	0.37 (1.12)
	3.36	**1.80** (2.81)	1.23 (2.02)	**2.34** (3.41)	5.06	0.15 (1.65)	−0.23 (0.87)	0.24 (0.85)
	6.72	**1.64** (2.36)	1.20 (2.25)	**1.80** (2.79)	10.12	0.17 (1.19)	0.35 (1.33)	0.14 (0.87)
Speech	0.87	−0.18 (0.67)	0.58 (1.74)	0.17 (1.31)	1.30	−0.28 (0.47)	0.17 (1.85)	−0.11 (1.49)
	1.74	0.89 (1.98)	0.44 (1.32)	0.26 (1.11)	2.60	0.20 (1.21)	−0.09 (1.18)	−0.31 (0.55)
	3.48	0.69 (2.38)	0.36 (1.02)	0.53 (1.00)	5.20	0.38 (0.85)	0.01 (1.28)	0.15 (0.85)

SS-EP magnitudes significantly different from background noise are highlighted for both corrected ($\alpha < 0.005$, bold and *) and uncorrected ($\alpha < 0.05$, bold only) thresholds.

noise floor was significant at the 3.36 Hz peak, $t(24) = 3.356$, $p = 0.0039$ (FDR-adjusted p-value) and at the 6.72 Hz peak, $t(24) = 3.843$, $p = 0.0030$ (FDR-adjusted p-value). These results suggest that infants reliably entrain to the fastest elements detectable in the tabla stimulus (the semiquaver and the quaver).

Finally, the ANOVA run on the Speech stimulus revealed only a main effect of noise $F(1,24) = 6.478$, $p = 0.018$, $\eta^2 = 0.213$ and a significant interaction frequency × laterality $F(4,96) = 4.612$, $p = 0.006$, $\eta^2 = 0.139$. Overall, as shown in Figure 2C, amplitudes relative to peaks of interest were higher than amplitude relative to noise floor. The interaction with laterality showed that at lower frequencies (e.g., 0.87 and 1.30 Hz) SS-EP magnitudes were higher at midline than at left, $t(24) = -2.814$, $p = 0.010$, whereas at higher frequency (e.g., 1.74 and 2.60 Hz) SS-EP magnitudes were higher at left than at right, $t(24) = 2.414$, $p = 0.024$. Since laterality affected SS-EP at frequencies of interest and at background noise similarly, paired t-tests were run averaging the three electrode clusters. They only reveal a difference between amplitudes in peaks of interest vs. noise floor at the 1.74 Hz peak, not surviving correction [$t(24) = 2.342$, $p = 0.028$, FDR-adjusted $p = 0.0835$]. These results suggest that infants show an overall effect of entrainment to the Speech stimulus, but they failed to reliably entrain to any specific frequencies.

Adults' data

Descriptive statistics for SS-EP magnitudes in all the three topographical clusters included in the analyses are reported in Table 2. SS-EP magnitudes significantly different from background noise are highlighted for both corrected ($\alpha < 0.005$) and uncorrected ($\alpha < 0.05$) thresholds.

The ANOVA run on the African musical stimulus reveal a main effect of noise, $F(1,9) = 7.419$, $p = 0.023$, $\eta^2 = 0.452$, and a significant interaction frequency × noise $F(2,18) = 4.252$, $p = 0.031$, $\eta^2 = 0.321$. Overall, as shown in Figure 3A, amplitudes relative to peaks of interest were higher than amplitude relative to noise floor. Paired t-tests reveal that the difference between amplitudes in peaks of interests vs. noise floor was significant at the 1.74 Hz peak, $t(9) = 3.101$, $p = 0.0381$ (FDR adjusted p-value). These results suggest that adults reliably entrain to the beat of the African stimulus (the crotchet).

The ANOVA run on the Tabla musical stimulus reveal no significant main effects or interactions. As shown in Figure 3B, paired t-tests reveal a significant difference between amplitudes in peaks of interests vs. noise floor at the 6.72 Hz peak, $t(9) = 4.950$, $p = 0.0024$ (FDR adjusted p-value). These results suggest that adults reliably entrain to the fastest element detectable in the tabla stimulus (the semiquaver).

The ANOVA run on the Speech stimulus reveal a main effect of noise, $F(1,9) = 7.491$, $p = 0.023$, $\eta^2 = 0.454$. Overall, as shown in Figure 3C, amplitudes relative to peaks of interest were higher than amplitude relative to noise floor. Paired t-tests only reveal a difference between amplitudes in peaks of interests vs. noise floor at the 3.48 Hz peak, $t(9) = 2.333$, $p = 0.045$, not surviving FDR correction (FDR adjusted $p = 0.1095$). These results suggest that – similarly to infants – adults show an overall effect of entrainment to the Speech stimulus, but they failed to reliably entrain to any specific frequencies.

Association with cognitive, language, and socio-emotional development

Associations between SS-EP magnitudes at frequencies of interest and concurrent measures of Cognitive, Language

TABLE 2 Descriptive statistics for the adults' data: Mean (Standard Deviation) on SS-EP magnitudes for frequencies of interest vs. noise floor separated for sound stimuli and cluster of channels.

	Peaks on the frequencies of interest				Noise floor amplitudes			
		Left	Midline	Right		Left	Midline	Right
	Hz	M (SD)	M (SD)	M (SD)	Hz	M (SD)	M (SD)	M (SD)
African	0.87	–0.37 (0.42)	-0.28 (0.58)	0.02 (1.09)	1.30	0.03 (0.73)	–0.33 (0.74)	–0.35 (0.40)
	1.74	**2.23** (3.73)	**2.88*** (3.70)	**2.78*** (6.20)	2.60	-0.74 (0.67)	-0.40 (1.46)	-0.58 (0.29)
	3.48	0.42 (0.83)	**2.17** (3.83)	**2.91*** (5.33)	5.20	0.23 (0.99)	–0.04 (0.97)	0.28 (1.17)
Tabla	1.68	0.32 (1.24)	0.31 (0.94)	-0.11 (0.64)	2.53	0.35 (1.44)	0.73 (1.86)	–0.07 (1.16)
	3.36	1.36 (2.23)	**3.90*** (7.07)	**2.92*** (5.53)	5.06	**1.87** (2.21)	**1.69** (1.98)	0.59 (1.14)
	6.72	1.16 (2.04)	**2.43** (1.19)	**2.80*** (2.03)	10.12	–0.13 (0.99)	–0.42 (0.81)	–0.51 (0.71)
Speech	0.87	0.43 (1.59)	–0.11 (0.74)	0.73 (2.22)	1.30	0.13 (0.79)	–0.08 (0.84)	0.07 (0.64)
	1.74	0.97 (1.92)	**1.65** (3.17)	0.87 (1.73)	2.60	–0.07 (0.56)	-0.31 (0.69)	–0.14 (0.72)
	3.48	0.59 (1.70)	**2.37** (3.13)	**1.82** (1.75)	5.20	0.02 (0.77)	0.29 (0.91)	–0.22 (0.74)

SS-EP magnitudes significantly different from background noise are highlighted for both corrected ($\alpha < 0.005$, bold and *) and uncorrected ($\alpha < 0.05$, bold only) thresholds.

(combined measure including both expressive and receptive skills), and Socio-Emotional development were assessed using Pearson's correlations. For each stimulus, we entered in the correlations SS-EP magnitudes at frequencies of interest that differ most from noise floor (i.e., 3.48 Hz for the African musical stimulus, 3.36 and 6.72 Hz for the Tabla musical stimulus, and 1.74 Hz for the Speech stimulus). Since basically no differences emerged concerning laterality, SS-EP magnitudes were averaged by topographical region.

Steady-state evoked potentials (SS-EP) magnitude for the Tabla musical stimulus at 3.36 Hz correlated with the Cognitive score, $r(23) = 0.534$, $p = 0.0417$ (FDR-adjusted p-value). Children with a greater SS-EP magnitude were characterized by higher cognitive scores. Additionally, SS-EP magnitude for the Speech stimulus at 1.74 Hz correlated with the Social-emotional score, $r(20) = 0.604$, $p = 0.0417$ (FDR-adjusted p-value). Children with a greater SS-EP magnitude were characterized by higher behaviors such as ease of calming, social responsiveness, and imitation play. Scatterplots are shown in **Figure 4**.

Discussion

The main aim of this study was to investigate low-frequency neural entrainment in response to complex rhythmic stimuli in a sample of 8-month-old infants. Replicating the methodology used by previous studies with simple beat stimuli (Nozaradan et al., 2011, 2012; Cirelli et al., 2016), we successfully showed some degree of neural entrainment in young infants even for much more complex rhythmic stimuli. As already mentioned, one of the advantages of using SS-EP recording is that it offers an objective definition of what to expect (Kabdebon et al., 2022). Here, based on the frequency of the presented auditory stimuli, we expected SS-EPs at very specific narrow frequency bands. Interestingly, we have shown that for all the stimuli our sample of infants, as a group, presented overall higher SS-EP magnitude for frequencies of interest (selected based on the stimulus envelope) than for "noise floor" frequencies. These results are in line with recent works computing cortical/neural tracking to complex and naturalistic continuous stimuli, such as sung nursery rhymes or infant/adult directed speech (Kalashnikova et al., 2018; Attaheri et al., 2022a,b; Menn et al., 2022). These studies provided evidence for low frequency cortical tracking of the stimulus envelope, already at 4 months of age.

For the two music stimuli we found robust entrainment at specific frequencies, whereas for the speech stimulus we only found an overall effect against "noise floor." It should be noted that in our study the music stimuli, although complex, were much more rhythmic than the speech stimulus. As described in section "Stimuli," the nursery rhyme was pronounced following the beat of a metronome (hearable in the stimulus) with exaggerated accents. However, despite our efforts to make the speech excerpt as rhythmic as possible, the rhythmic entrainment to the physical properties of this kind of stimulus was not robust, probably because even young infants tend to process the speech stimuli differently from non-speech rhythmic stimuli. Similarly, studies on behavioral rhythmic sensorimotor synchronization show that infants aged between 6 and 16 months engage significantly more with rhythmic movements to music and non-speech rhythmically regular sounds than to speech (Zentner and Eerola, 2010). Overall, we found no hemispheric differences in the magnitude of SS-EPs, with the only exception being the speech stimulus, for which we found slightly more left-lateralized SS-EPs at higher frequencies. This left lateralization restricted to the processing of speech stimuli is not unexpected (Dehaene-Lambertz, 2017), and we interpret this finding as further evidence of different neural mechanisms underlying speech and non-speech processing.

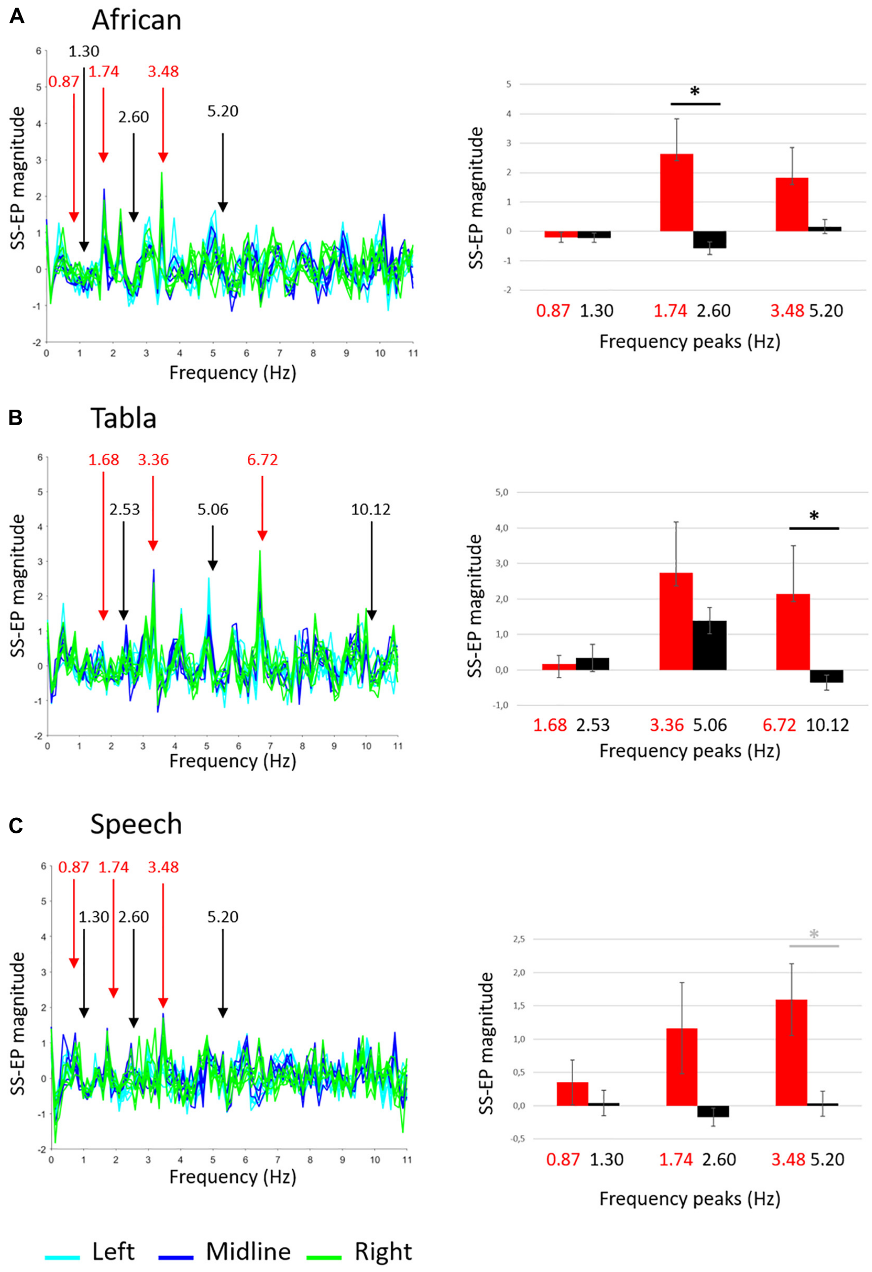


FIGURE 3
Adults' group-level average SS-EPs for the three sound stimuli: **(A)** African musical stimulus; **(B)** Tabla musical stimulus; **(C)** Speech stimulus. On the left, butterfly plot of SS-EPs obtained applying the Signal-to-Noise correction (*x*-axis: frequency; *y*-axis: SS-EP magnitude). All EEG channels included in the clusters are plotted (Left: light blue lines, Midline: blue lines, Right: green lines). Peaks at frequencies corresponding to the rhythm envelope are highlighted in red, whereas peaks at frequencies not corresponding to the rhythm envelope, considered as noise floor, are highlighted in black. On the right, bar graphs (error bars indicate SEM) show the averaged SS-EP magnitude. Significant paired *t*-tests are indicated by black asterisks, whereas differences not surviving False discovery rate (FDR) correction are indicated by gray asterisks.

It should be noted that we retained a smaller number of artifact-free epochs for the Speech stimulus with respect to the African musical stimulus in infants. We cannot exclude that the smaller number of trials might explain the absence of a robust entrainment at specific frequencies for this stimulus. However, since we found the same results in adults, where no between-stimulus difference in the number of accepted trials emerged, we can rule out this possibility.

When comparing infants' neural entrainment with that of a small sample of adults using the identical experimental stimuli and analytic methods, we found very similar results in terms of SS-EP magnitudes and frequencies to which neural entrainment

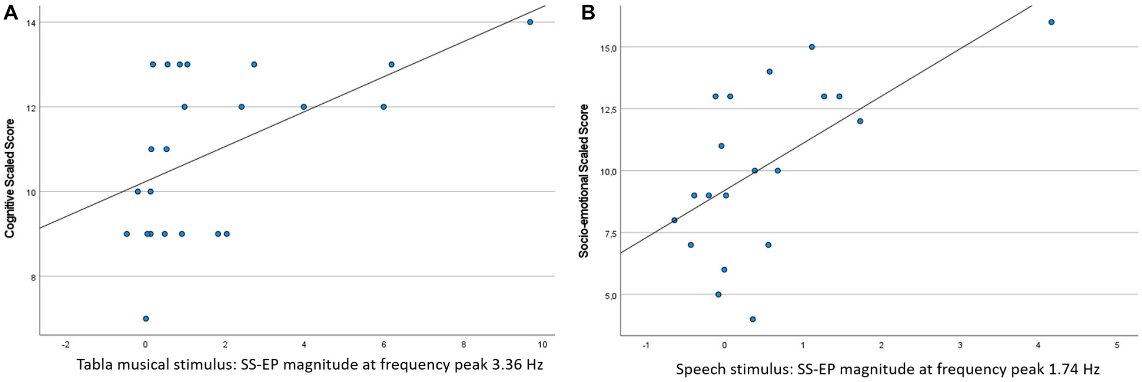


FIGURE 4
Scatterplots of the significant correlations between SS-EP magnitude in the neural entrainment task and concurrent standardized measures of development: **(A)** SS-EP magnitude in the Tabla musical stimulus (at frequency peak 3.36 Hz) is plotted against the cognitive scaled score; **(B)** SS-EP magnitude in the Speech stimulus (at frequency peak 1.74 Hz) is plotted against the socio-emotional scaled score.

occurred, in line with recent neural tracking evidence revealing overall "more developmental similarities than developmental differences" between infants and adults in the entrainment to sung nursery rhymes (Attaheri et al., 2022b). For the African musical stimulus, we found robust neural entrainment for both infants and adults, although at different frequencies: whereas infants entrained at 3.48 Hz, corresponding to the faster perceived elements at the rhythmic level (i.e., the quaver, as shown in **Figure 1A**), adults entrained at 1.74 Hz, corresponding to the beat (i.e., a less perceptual and more «abstract» feature of the stimulus). For the Tabla musical stimulus, we found robust neural entrainment for both infants and adults at the 6.72 Hz frequency, corresponding to the faster perceived elements at the rhythmic level (i.e., the semiquaver, as shown in **Figure 1B**). Interestingly, for this stimulus, we additionally found that children entrained at frequency 3.36 Hz, corresponding to the quaver. It should be noted that for this sound stimulus this more "abstract" level was also somehow detectable at the perceptual level (i.e., given by the combination of cowbell and sleigh bells). For the Speech stimulus, we found that adults, similarly to infants, did not show robust entrainment for any frequencies (nor for the perceptual beat of the metronome nor for more «abstract» beats). Overall, these results suggest that both adults and infants exhibit neural synchronization and tempo flexibility. This is partially contrasting with evidence reporting that, behaviorally, these skills do not emerge until the second year of age (Zentner and Eerola, 2010; Fujii et al., 2014; Yu and Myowa, 2021) and suggests that neural responses might be more sensitive than behavioral measures at a younger age.

Our last aim concerned the association between individual differences in infants' measures of neural entrainment and the concurrent level of linguistic, cognitive, and social-emotional development. Contrary to our expectations, supported by theoretical frameworks proposing entrainment of neural oscillations to external rhythmic stimuli as one of the factors underlying speech and language development (Ladányi et al., 2020; Fiveash et al., 2021), we did not find any association with language skills. It should be noted, however, that here we only considered concurrent and very early language skills. Further studies should investigate such an association in a longitudinal perspective. Interestingly, associations were indeed found between neural entrainment to non-speech rhythmic stimulus – and more specifically between the ability to synchronize at the more «abstract» level (i.e., frequency 3.36 Hz) in the Tabla stimulus – and overall cognitive development. Additionally, we found that responses to nursery rhymes spoken by an adult female in infant-directed speech were associated to social-emotional development. Although not conclusive and in need of replication in larger samples, these correlations provide some preliminary evidence on the role of neural entrainment in development and support its functional interpretation.

Overall, the present study is a starting point for further investigation on the role of rhythm perception/synchronization in infancy (with more sensitive measures than behavioral ones) on later speech and language (a)typical development (Ladányi et al., 2020; Fiveash et al., 2021). In future research, the methodology proposed here could be coupled with novel analytic approaches including neural/cortical tracking performed in the time domain and thus might add convergent pieces of information. Both of these approaches seem appropriate to be used with infants at familial risk for Developmental Language Disorder and Dyslexia (e.g., Cantiani et al., 2016, 2019), since recent theories suggest that individual differences in this phenomenon could be one factor leading to atypical development trajectories of language acquisition found in these disorders (e.g., Goswami, 2011; Molinaro et al., 2016; Di Liberto et al., 2018). Furthermore, such measures seem appropriate for the investigation of the effect of early music/rhythmic training in infancy (e.g., Zhao and Kuhl, 2016; Dondena et al., 2021) in typical and atypical populations.

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 studies involving human participants were reviewed and approved by Ethical Committee of the Scientific Institute IRCCS Eugenio Medea. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

CC, VR, MM, and CP designed the study. CC and CD run the experiment and collected the data. CC, CD, and CP analyzed the data. CC, VR, and CP interpreted the results. CC drafted the manuscript. All authors edited and revised the manuscript.

Funding

This work was supported by the Italian Ministry of Health (Ricerca Corrente).

Acknowledgments

We wish to thank Dr. Renata del Giudice for her help in stimuli design and Livia Auguadro for providing her musical expertise. Additionally, we wish to express our gratitude to Prof. Laura Cirelli who very kindly replied to our questions about data processing and analytic procedures. Special thanks go to all infants and their parents who participated in this study, as well as to all the adult volunteers.

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

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., et al. (2022a). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *Neuroimage* 247:118698. doi: 10.1016/J.NEUROIMAGE.2021.118698

Attaheri, A., Panayiotou, D., Phillips, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., et al. (2022b). Cortical tracking of sung speech in adults vs infants: A developmental analysis. *Front. Neurosci.* 16:842447. doi: 10.3389/FNINS.2022.842447

Bayley, N. (2006). *Bayley scales of infant and toddler development*. San Antonio, TX: PsychCorp, Pearson.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., et al. (2019). Cortical route for facelike pattern processing in human newborns. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4625–4630. doi: 10.1073/PNAS.1812419116

Cantiani, C., Ortiz-Mantilla, S., Riva, V., Piazza, C., Bettoni, R., Musacchia, G., et al. (2019). Reduced left-lateralized pattern of event-related EEG oscillations in infants at familial risk for language and learning impairment. *Neuroimage Clin.* 22:101778. doi: 10.1016/j.nicl.2019.101778

Cantiani, C., Riva, V., Piazza, C., Bettoni, R., Molteni, M., Choudhury, N., et al. (2016). Auditory discrimination predicts linguistic outcome in Italian infants with and without familial risk for language learning impairment. *Dev. Cogn. Neurosci.* 20, 23–34. doi: 10.1016/j.dcn.2016.03.002

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., and Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychol. Sci.* 31, 1161–1173. doi: 10.1177/0956797620933237

Choudhury, N., and Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clin. Neurophysiol.* 122, 320–338. doi: 10.1016/J.CLINPH.2010.05.035

Cirelli, L. K., Spinelli, C., Nozaradan, S., and Trainor, L. J. (2016). Measuring neural entrainment to beat and meter in infants: Effects of music background. *Front. Neurosci.* 10:229. doi: 10.3389/FNINS.2016.00229/BIBTEX

de Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *Elife* 4, 1–14. doi: 10.7554/ELIFE.06564

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., and Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology* 57:e13580. doi: 10.1111/PSYP.13580

Dehaene-Lambertz, G. (2017). The human infant brain: A neural architecture able to learn language. *Psychon. Bull. Rev.* 24, 48–55. doi: 10.3758/S13423-016-1156-9/FIGURES/1

Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Di Liberto, G. M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., and Lalor, E. C. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *Neuroimage* 175, 70–79. doi: 10.1016/J.NEUROIMAGE.2018.03.072

Ding, N., and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., and Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front. Hum. Neurosci.* 11:481. doi: 10.3389/FNHUM.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186

Dondena, C., Riva, V., Molteni, M., Musacchia, G., and Cantiani, C. (2021). Impact of early rhythmic training on language acquisition and electrophysiological functioning underlying auditory processing: Feasibility and preliminary findings in typically developing infants. *Brain Sci.* 11:1546. doi: 10.3390/BRAINSCI11111546

Fiveash, A., Bedoin, N., Gordon, R. L., and Tillmann, B. (2021). Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology* 35, 771–791. doi: 10.1037/NEU0000766

Fló, A., Benjamin, L., Palu, M., and Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Sci. Rep.* 12, 1–13. doi: 10.1038/s41598-022-08411-w

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., and Taga, G. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS One* 9:e97680. doi: 10.1371/JOURNAL.PONE.0097680

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10. doi: 10.1016/J.TICS.2010.10.001

Jin, P., Lu, Y., and Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *Elife* 9:e55613. doi: 10.7554/ELIFE.55613

Kabdebon, C., Fló, A., de Heering, A., and Aslin, R. (2022). The power of rhythms: How steady-state evoked responses reveal early neurocognitive development. *Neuroimage* 254:119150. doi: 10.1016/J.NEUROIMAGE.2022.119150

Kabdebon, C., Pena, M., Buiatti, M., and Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain Lang.* 148, 25–36. doi: 10.1016/J.BANDL.2015.03.005

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., and Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-32150-6

Kirschner, S., and Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/J.JECP.2008.07.005

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., and Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdiscip. Rev. Cogn. Sci.* 11:e1528. doi: 10.1002/WCS.1528

Leach, S. C., Morales, S., Bowers, M. E., Buzzell, G. A., Debnath, R., Beall, D., et al. (2020). Adjusting ADJUST: Optimizing the ADJUST algorithm for pediatric data using geodesic nets. *Psychophysiology* 57:e13566. doi: 10.1111/PSYP.13566

Lenc, T., Keller, P. E., Varlet, M., and Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8221–8226. doi: 10.1073/PNAS.1801421115

Menn, K., Ward, E., Braukmann, R., and van den Boomen, C. (2022). Neural tracking in infancy predicts language development in children with and without family history of autism. *Neurobiol. Lang.* 3, 495–514.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., and Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Hum. Brain Mapp.* 37, 2767–2783. doi: 10.1002/HBM.23206

Mouraux, A., and Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magn. Reson. Imaging* 26, 1041–1054. doi: 10.1016/J.MRI.2008.01.011

Nolan, H., Whelan, R., and Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *J. Neurosci. Methods* 192, 152–162. doi: 10.1016/J.JNEUMETH.2010.07.015

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130393. doi: 10.1098/RSTB.2013.0393

Nozaradan, S., Peretz, I., and Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581. doi: 10.1523/JNEUROSCI.3203-12.2012

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240. doi: 10.1523/JNEUROSCI.0411-11.2011

Nozaradan, S., Schönwiesner, M., Keller, P. E., Lenc, T., and Lehmann, A. (2018). Neural bases of rhythmic entrainment in humans: Critical transformation between cortical and lower-level representations of auditory rhythm. *Eur. J. Neurosci.* 47, 321–332. doi: 10.1111/EJN.13826

Obleser, J., and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends Cogn. Sci.* 23, 913–926. doi: 10.1016/J.TICS.2019.08.004

Peykarjou, S., Hoehl, S., Pauen, S., and Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* 7, 1–12. doi: 10.1038/s41598-017-12760-2

Phillips-Silver, J., and Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science* 308, 1430–1430. doi: 10.1126/SCIENCE.1110922

Rocha, S., and Mareschal, D. (2017). Getting into the groove: The development of tempo-flexibility between 10 and 18 months of age. *Infancy* 22, 540–551. doi: 10.1111/INFA.12169

Rodari, G. (1960). "Girotondo di tutto il mondo," in *Filastrocche in cielo e in terra*, ed. Einaudi.

Stupacher, J., Wood, G., and Witte, M. (2017). Neural entrainment to polyrhythms: A comparison of musicians and non-musicians. *Front. Neurosci.* 11:208. doi: 10.3389/FNINS.2017.00208/BIBTEX

Tal, I., Large, E. W., Rabinovitch, E., Wei, Y., Schroeder, C. E., Poeppel, D., et al. (2017). Neural entrainment to the beat: The "missing-pulse" phenomenon. *J. Neurosci.* 37, 6331–6341. doi: 10.1523/JNEUROSCI.2500-16.2017

Tierney, A., and Kraus, N. (2015). Neural entrainment to the rhythmic structure of music. *J. Cogn. Neurosci.* 27, 400–408. doi: 10.1162/JOCN_A_00704

Van Zuijen, T. L., Plakas, A., Maassen, B. A. M., Been, P., Maurits, N. M., Krikhaar, E., et al. (2012). Temporal auditory processing at 17 months of age is associated with preliterate language comprehension and later word reading fluency: An ERP study. *Neurosci. Lett.* 528, 31–35. doi: 10.1016/J.NEULET.2012.08.058

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2468–2471. doi: 10.1073/PNAS.0809035106

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: A study of 18- to 42-month-old children. *Infancy* 26, 635–646. doi: 10.1111/INFA.12403

Zentner, M., and Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773. doi: 10.1073/PNAS.1000121107

Zhao, T. C., and Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5212–5217. doi: 10.1073/PNAS.1603984113

Zhou, H., Melloni, L., Poeppel, D., and Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics. *Front. Hum. Neurosci.* 10:274. doi: 10.3389/FNHUM.2016.00274/BIBTEX

frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 08 November 2022
DOI 10.3389/fpsyg.2022.1015611

Frequency tagging with infants: The visual oddball paradigm

Stefanie Peykarjou*

Department of Psychology, Heidelberg University, Heidelberg, Germany

Combining frequency tagging with electroencephalography (EEG) provides excellent opportunities for developmental research and is increasingly employed as a powerful tool in cognitive neuroscience within the last decade. In particular, the visual oddball paradigm has been employed to elucidate face and object categorization and intermodal influences on visual perception. Still, EEG research with infants poses special challenges that require consideration and adaptations of analyses. These challenges include limits to attentional capacity, variation in looking times, and presence of artefacts in the EEG signal. Moreover, potential differences between age-groups must be carefully evaluated. This manuscript evaluates challenges theoretically and empirically by (1) a systematic review of frequency tagging studies employing the oddball paradigm and (2) combining and re-analyzing data from seven-month-old infants (*N*=124, 59 females) collected in a categorization task with artifical, unfamiliar stimuli. Specifically, different criteria for sequence retention and selection of harmonics, the influence of bins considered for baseline correction and the relation between fast periodic visual stimulation (FPVS) responses and looking time are analyzed. Overall, evidence indicates that analysis decisions should be tailored based on age-group to optimally capture the observed signal. Recommendations for infant frequency tagging studies are developed to aid researchers in selecting appropriate stimulation and analysis strategies in future work.

KEYWORDS

frequency tagging, fast periodic visual stimulation, categorization, analysis strategies, visual processing, infants

EDITED BY
Vivian Ciaramitaro,
University of Massachusetts Boston,
United States

REVIEWED BY
Wanze Xie,
Peking University,
China
Peter Gerhardstein,
Binghamton University,
United States
Olga V. Sysoeva,
Institute of Higher Nervous Activity and
Neurophysiology (RAS), Russia

*CORRESPONDENCE
Stefanie Peykarjou
stefanie.peykarjou@psychologie.
uni-heidelberg.de

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 09 August 2022
ACCEPTED 18 October 2022
PUBLISHED 08 November 2022

CITATION
Peykarjou S (2022) Frequency tagging with
infants: The visual oddball paradigm.
Front. Psychol. 13:1015611.
doi: 10.3389/fpsyg.2022.1015611

Introduction

Stimulating the brain rhythmically and measuring the resulting rhythmic changes of electrical activity on the scalp has been successfully employed in vision research for many decades (Adrian and Matthews, 1934; Adrian, 1944; Walter et al., 1946; Dawson, 1954). More recently, rhythmic stimulation has been applied in cognitive research and provided evidence for very fast face and object categorization in infants and adults (Liu-Shuang et al., 2014; de Heering and Rossion, 2015; Rossion et al., 2015; Peykarjou et al., 2017; Kabdebon et al., 2022). Due to its objectivity and high signal-to-noise ratio, this approach holds great potential for developmental cognitive neuroscience. In particular, the so-called fast periodic visual stimulation (FPVS) oddball paradigm (Heinrich et al., 2009; Liu-Shuang et al., 2014) embeds categorical changes at a slower frequency rate into a fast periodic stream of stimuli and thus allows for dissociating the rate of categorization from low-level changes within a single stimulation stream. This paper summarizes studies utilizing the approach, addresses its methodological strengths and challenges and systematically examines analysis decisions on response patterns to the visual FPVS oddball paradigm in infants.

In the past, inducing rhythmic responses through visual stimulation has been described in the frameworks of steady-state visual evoked potentials (SSVEPs; Regan, 1966, Norcia et al., 2015), fast periodic visual stimulation responses (FPVS; Liu-Shuang et al., 2014, Rossion et al., 2020), and frequency tagging (Tononi et al., 1998). Here, the term frequency tagging will be adopted as it captures the idea of associating a cognitive process with a frequency tag, and thus stresses an important advantage of the approach, the *a priori* defined dependent measure. Frequency tagging can be understood as the process of stimulating in a given frequency and analyzing EEG responses at the corresponding frequencies. To this purpose, the continuous EEG signal recorded in the time-domain is converted to the frequency domain *via* a fast Fourier Transformation (FFT). Thus, changes in amplitude over time are converted and represented as amplitude and phase across frequencies, allowing the researcher to track response strengths at stimulated frequencies.

The frequency tagging approach has recently been employed to measure face individuation and object categorization with great validity, objectivity, reliability and sensitivity (Stothart et al., 2017; Rossion et al., 2020; Peykarjou et al., under review[1]). It is objective in the sense that the response of interest is defined *a priori* by the input frequency f (fundamental frequency), and analysis are confined to the input frequency and its harmonics (nf). Satisfactory reliabilities of adult frequency-tagging responses in face individuation have recently been established (Dzhelyova et al., 2019; Stacchi et al., 2019), but reliability certainly depends on the specific stimuli and design as well as processing proficiency of participants. It is conceivable that more variable responses in less capable participants will lead to decreased reliability (see also Stacchi et al., 2019). There also is first evidence that frequency tagging responses are correlated with performance, e.g., decreased face individuation responses in prosopagnosia (Liu-Shuang et al., 2016). In addition, previous work has demonstrated high sensitivity of frequency tagging, requiring only a few minutes of stimulation to reliably differentiate responses from noise (Dzhelyova et al., 2019). Finally, flickering images are highly attractive for infants, thus eliciting prolonged looking and alertness, further enhancing the usability of this approach in development.

Thus, frequency tagging holds great potential for developmental cognitive neuroscience, and has been successfully employed in vision and development over decades. Since the 1970, it was leveraged to delineate the development of spectral sensitivity, acuity, contrast sensitivity and binocular interaction (Dobson, 1976; Sokol, 1978; Atkinson et al., 1979; Petrig et al., 1981). During the last years, frequency tagging has received surging interest and been applied in face and object categorization, intermodal perception, and attention (De Heering and Rossion, 2015; Peykarjou et al., 2017; Barry-Anwar et al., 2018; Christodoulou et al., 2018; Leleu et al., 2020; Rekow et al., 2021). The present study systematically evaluates design and analysis decisions to develop recommendations for future applications of this method.

First, methodological differences between infant and adult studies will be described, focusing on particulars of frequency tagging (Hoehl and Wahl, 2012, for more general discussions of infant EEG studies, see DeHaan, 2007). This part is supplemented by a systematic review of existing studies following the FPVS oddball approach. In the empirical part, the effects of methodological decisions will be assessed.

Methodological characteristics of frequency tagging employed with infants

A great advantage of EEG frequency tagging studies is that the same stimulation and dependent measures can be employed across age. As the frequency of interest is defined and embedded into the stimulation *a priori* by the researcher, responses at this precise frequency can be analyzed regardless of age-group or other sample characteristics. While one might adapt the speed of stimulation and thus the fundamental frequency for different samples, past research has employed the same speed for infants and adults and found that infants are capable of processing images at presentation times employed with adults, typically less than 170 ms per image. Anyhow, studies determining how infants respond to stimulation at different fundamental frequencies are lacking (see, for example, Gentile and Rossion, 2014; Retter et al., 2020, for evidence in adults) and would be very helpful for furthering our understanding of differences and commonalities in rhythmic responses throughout development.

While the general stimulation can thus be applied irrespective of age, some specifics require adaptation for infant participants. For example, infants' attentional capacity is relatively short, which requires adapting the stimulation in terms of duration of flicker sequences, the number of sequences and, thus, the number of conditions that can be tested within-subjects. Moreover, the rather high amount of artefacts typically present in infant EEG requires careful handling. To provide an empirical overview of methodological decisions made in FPVS oddball studies with infants, a focused systematic review was conducted based on the search terms "infant" and "frequency tagging," "SSVEPs" or "fast periodic visual stimulation" (see Figure 1).

The Psycinfo search was performed on March 23rd 2022 and supplemented with records identified by personal communication. All studies employing a visual FPVS oddball paradigm with infant participants (< 4 months of age) were included, yielding $N = 11$ records (see Table 1 for an overview of all included studies). All information was double-coded by the author and student assistants. The following information

---

[1] Peykarjou, S., Hoehl, S. and Pauen, S. (under review). The Development of Visual Categorization Based on High-Level Cues.

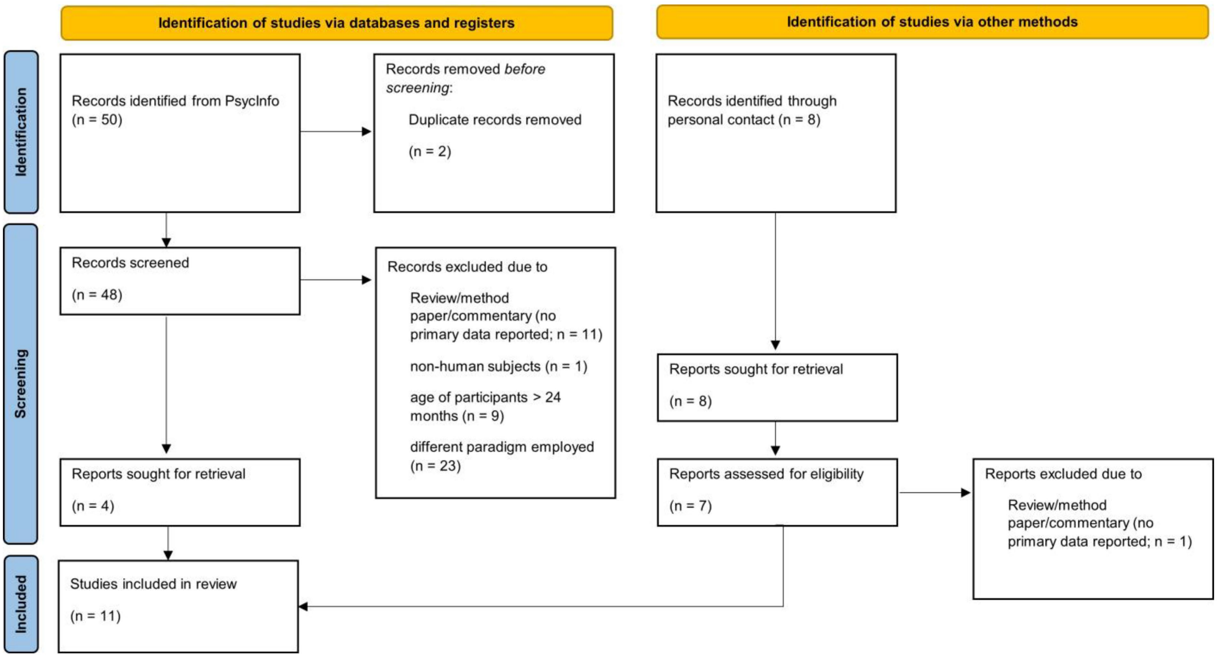


FIGURE 1
Flow diagram of study selection process for systematic review (Page et al., 2021).

TABLE 1 Design of studies included in the systematic review.

First Author/Year of Publication	Journal	*N/M* age (mo)	Software/Hardware	No within – subject – conditions	Sequence duration / total presentation duration (s)
Baccolo et al., in preparation (see footnote 3)	submitted	29/6	24 inch, 60 Hz monitor/Matlab	2	15/120
Barry-Anwar et al., 2018	Neuropsychologia	23/619/9	Not specified/Matlab	2	10/200
Bertels, et al., 2020	Scientific Reports	39/8	800 × 600 monito, 60 Hz/Windows, Matlab, Psychotoolbox	2	20/251
De Heering and Rossion, 2015	eLIFE	30/5	not specified/Windows, Matlab, Psychtoolbox	1	20/160
Leleu et al., 2020	Developmental Psychology	18/4	24 inch, 60 Hz monitor/custom Java software	2	31.17/264.95
Peykarjou et al., 2017	Scientific Reports	41/9	60 Hz monitor/Windows, Matlab	2	20/212.1
Peykarjou et al., 2022	Cortex	39/5	60 Hz monitor/Windows, custom Java software	2	20/113.2
Peykarjou et al., under revision (see footnote 1)	under revision	44/4 40/756/11	60 Hz monitor/Windows, Matlab and custom Java software	1	20/166
Poncet et al., 2022	Frontiers in Neuroscience	18/3.5 18/7	24 inch, 60 Hz monitor/not specified	3	31.17/311.7
Rekow et al., 2020	Cognitive Development	18/4	24 inch, 60 Hz monitor/custom Java software	2	31.167/311.7
Rekow et al., 2021	Proceedings of the National Academy of Sciences	20/4	24 inch, 60 Hz monitor/not specified	2	31.17/280.53

was extracted from included articles, separately for all groups tested: mean participant age, sample size, software and hardware employed, number of within-subject conditions, duration of sequences, number of sequences presented and kept, criteria for excluding sequences, methods employed for artifact detection, treatment of artefacts, number of harmonics analyzed, criteria for selection of harmonics, number of bins for baseline-correction, and response strength of significant responses (see Tables 1, 2).

Due to infants' limited attentional capacity, the duration of frequency tagging sequences is reduced compared to studies with older participants. The literature review (see Table 1) indicates that previous studies employed sequence durations of 15–30 s with infants, whereas 60 s are typical for adults (e.g., Liu-Shuang

TABLE 2 Analysis decisions of studies included in the systematic review.

First Author/ Year of Publication	*M* no sequences pres./incl.	Criterion for sequence exclusion	Artifact detection/ criterion for harmonics selection	Handling of artefacts	Bins for baseline-correction	Response strength (of significant categorization)	Harmonics considered (categorization)
Baccolo et al., in preparation (see footnote 3)	8/7.65	looking at screen for less than majority of sequence, SNR < 2 at base frequency or first harmonic on 70, 75 or 83 (=O1, O2, or Oz)	visual inspection/ Bonferroni-corrected Z-scores of grand-averages, averaged across conditions	interpolation, exclusion	10 (excl. Adjacent bins)	SNR 2.34	3rd to 4th harmonic summed
Barry-Anwar et al., 2018	20/11.70 20/9.79	looking at screen for <50%, artefacts/noise >50% of time	visual inspection/no harmonic responses evident	exclusion or values replaced with zeros	10 (excl. Adjacent bins)	SNR 1.05–1.27	no harmonic responses considered
Bertels, et al., 2020	12.55/8.41	looking at screen for less than majority of sequence, SNR < 2 at base frequency on O1, O2, and Oz	visual inspection/ based on literature	interpolation, exclusion, normalization of Fourier coefficients	10 (excl. Adjacent bins)	SNR 1.84–4.17	first 4 harmonics averaged
de Heering and Rossion, 2015	8/4.75	looking at screen for less than majority of sequence, SNR < 2 at base frequency on O1, O2, or Oz	visual inspection/ X	interpolation, exclusion	20 (excluding adjacent bins)	SNR 0.99–2.56	harmonics 1 and 2 described
Leleu et al., 2020	8.5/7.72	crying, > = 2 Z-scores >1.64 or 1 Z-score > 2.3 at base frequency and 1st harmonic on Oz, POz, O1, or O2	Artifact Blocking algorithm, visual inspection/Z-scores of grand-averages, averaged across conditions	Artifact Blocking algorithm, interpolation, exclusion	10	Z 1.91–4.79	only 1st harmonic significant
Peykarjou et al., 2017	10.60/9.25	significant response at base frequency, looking at screen for <50% of time	visual inspection/ Bonferroni-corrected Z-scores of grand-averages, averaged across conditions	interpolation, exclusion	10 (excl. Adjacent bins)	SNR 1.08–1.59	1st to 11th harmonic summed
Peykarjou et al., 2022	5.66/2.08	> 3 channels requiring interpolation, too many artifacts, SNR < 2 at base frequency and 1st harmonic on O1, O2, and Oz	visual inspection/ Bonferroni-corrected Z-scores of grand-averages, averaged across conditions	interpolation, exclusion	10 (excl. Adjacent bins)	BCA-43-1.82	1st to 9th harmonic summed
Peykarjou et al., under revision (see footnote 1)	6.68/3.59 8.46/5.27 9.28/5.14	> 3 channels requiring interpolation, too many artifacts, SNR < 2 at base frequency and 1st harmonic on O1, O2, and Oz	visual inspection/ Bonferroni-corrected Z-scores of grand-averages, averaged across conditions	interpolation, exclusion	20 (excl. Adjacent bins, and two extreme bins)	BCA 0.08–0.53	harmonics summed; 4 mo: 3rd harmonic, 7 ms 3rd - 4th harmonic, 11 mo 2nd - 4th harmonic
Poncet et al., 2022	<=14/10.11 < =14/10.56	> = 2 Z-scores >1.64 or 1 Z-score > 2.3 at base frequency and 1st harmonic on Oz, POz, O1, or O2; global noise-corrected amplitude > < 2 SD of mean of all sequences, < 2 sequences left after preprocessing	Artifact Blocking algorithm, visual inspection/Z-scores of grand-averages	Artifact Blocking algorithm, interpolation, exclusion	10	SNR 2.05–4.34	condition 1: only stimulation frequency significant, condition 2: 1st to 2nd harmonic summed
Rekow et al., 2020	10/9.20	> = 2 Z-scores >1.64 or 1 Z-score > 2.3 at base frequency and 1st harmonic on Oz, POz, O1, or O2; global noise-corrected amplitude > < 2 SD of mean of all sequences, only one sequence left after preprocessing	Artifact Blocking algorithm, visual inspection/Z-scores of grand-averages	Artifact Blocking algorithm, interpolation, exclusion	10	Z 2.44	1st to 3rd harmonic summed
Rekow et al., 2021	9/8.50	> = 2 Z-scores >1.64 or 1 Z-score > 2.3 at base frequency and 1st harmonic on Oz, POz, O1, or O2; global noise-corrected amplitude > < 2 SD of mean of all sequences	Artifact Blocking algorithm, visual inspection/Z-scores of grand-averages	Artifact Blocking algorithm, interpolation, exclusion	10	Z 1.76–3.13	only 1st harmonic significant

(Continued)

TABLE 2 (Continued)

First Author/ Year of Publication	*M* no sequences pres./incl.	Criterion for sequence exclusion	Artifact detection/ criterion for harmonics selection	Handling of artefacts	Bins for baseline-correction	Response strength (of significant categorization)	Harmonics considered (categorization)
	SNR 1 channel >1.5	SNR 1 channel >2	SNR 2 channels >2	SNR 3 channels >2	Z-score 1 channel >2.33	Z-score 2 channels >1.64	Z-score 3 channels >1.64
M (SD) number of remaining sequences	5.76 (2.94)	5.18 (2.90)	4.39 (2.82)	2.96 (2.43)	5.02 (2.80)	4.75 (2.79)	3.54 (2.60)
M (SD) number of excluded sequences	0.06 (0.25)	0.64 (0.90)	1.43 (1.52)	2.86 (2.24)	0.80 (1.05)	1.07 (1.23)	2.28 (1.78)
N remaining participants	124	123	119	107	123	123	113
SNR of GA categorization response	1.32	1.32	1.37	1.40	1.35	1.37	1.33
SNR of GA base response	3.12	3.32	3.40	3.40	3.27	3.34	3.14
Correlation of 6 Hz response with looking time		0.39 ($N = 546$)	0.34 ($N = 418$)	0.29 ($N = 292$)	0.37 ($N = 529$)	0.35 ($N = 498$)	0.34 ($N = 366$)
BF of correlation 6 Hz + looking time		$7.23e^{17}$	$1.48e^{10}$	23,553	$4.67e^{15}$	$5.73e^{12}$	$2.13e^{8}$

Comparison of exclusion rates, grand-average responses and relations with looking times across different criteria for sequence retention.

et al., 2014; Stothart et al., 2017; Stacchi et al., 2019). This adaptation allows researchers to provide more breaks, soothe infant participants if necessary, and redirect attention to the screen. However, by adapting the duration of sequences, the resulting frequency resolution is also changed as both are inversely related *via* the FFT (1/duration(s)). This leads to a lower frequency resolution when using shorter stimulation times with infants. Moreover, the duration of sequences is potentially related to signal strength at the tagged frequency for two reasons (e.g., Dzhelyova et al., 2019): First, increasing stimulation time may support perception of embedded rhythms, thus strengthening responses, and second, a higher frequency resolution may enhance the signal-to-noise ratio by confining the stimulated response to a narrow frequency bin relative to the broad-band EEG noise.

In addition to reducing stimulation time, infants' attentional span also limits the number of conditions that can be tested within-subjects. Based on the systematic review, infants (0–11 months) on average watched $M = 213.06$ ($SD = 89.51$) seconds of stimulation, arranged into 1–3 conditions. Most common were sequence durations of 15–30 s, so that an overall average of $M = 9.91$ ($SD = 3.60$) sequences were presented. Considering that some sequences will need to be excluded during preprocessing (see below), in most cases, running two to three conditions within-subjects is feasible.

Another caveat posed by infants' attentional limits is variable looking time. Infants cannot be instructed to attend to the screen and will thus engage and disengage in a variable pattern during frequency tagging sequences. In contrast, most adult studies employ an orthogonal task to control attention and keep it constant (Liu-Shuang et al., 2014; Stothart et al., 2017; Stacchi et al., 2019). When participants look away from the screen, this may reduce the response strength as, on the one hand, they do not observe the stimulation and, on the other hand, they likely produce motor artefacts. Faced with this problem, all infant studies define a minimal threshold for inclusion of sequences, mostly based on responses to the general visual stimulation, sometimes considering looking times and/or data quality criteria (see Table 1 for an overview of inclusion criteria). Determining a criterion for retention of sequences is critical. On the one hand, sequences that were not attended well should be excluded, as categorization responses cannot be measured validly. On the other hand, as much data as possible should be retained to remain representative of the original dataset.

As elaborately discussed elsewhere (DeHaan, 2007; Hoehl and Wahl, 2012), infant EEG is characterized by a high amount of artefacts. These can be due to movement, limited preparation time during capping (leading to higher impedances), sweating or biological processes (e.g., heartbeat or respiration). Artefacts impact frequency tagging responses differently than other dependent variables, such as event-related-potentials (ERPs), as they are distributed broadly but not necessarily evenly in the frequency spectrum. When the infant blinks, she will do so in a partly random fashion, thus contributing to a range of frequencies. This means that artefacts may distort amplitudes across ranges of frequencies rather than at specific frequencies.[2] Analyses of responses spread across the frequency spectrum, such as ERPs, are heavily influenced by these artefacts. In contrast, frequency tagging is based on analyzing specific frequencies of interest, and while artefacts may influence amplitude at each specific frequency, they will exhibit a similar influence on surrounding frequency bins and can thus be controlled by baseline correction.

The artefacts distributed throughout the spectrum increase standard deviations (SDs) of amplitudes across frequencies. This increased SD of whole frequency ranges may diminish the strength of the baseline corrected signal compared to cleaner data collected with older participants. Therefore, artefacts will enhance variability of the signal across bins, likely diminishing the statistical significance of responses at frequencies of interest. However, they will not influence the tagged frequency specifically, providing one major strength of this approach.

Dealing with artefacts is a central aspect of EEG analysis. Similar to other EEG approaches, in the context of frequency tagging, artefacts can be identified visually or by algorithms. As can be seen in Table 2, previous studies treated artifacts most commonly by excluding data, interpolating channels, and applying an artifact blocking algorithm. Semi-automizing the detection and handling of artifacts seems an important venture for the near future to facilitate and objectify analysis.

To determine the strength of the response at the tagged frequency, amplitude at this frequency is compared to and corrected for amplitude at the surrounding bins (baseline correction). The systematic review indicates that 10 bins (5 on each side, excluding the immediately adjacent bins), corresponding to approx. + − 0.35 Hz, are employed most frequently. The number of bins taken into account is reduced relative to typical studies with adults (considering 20 bins, Liu-Shuang et al., 2014; Stacchi et al., 2019) due to the shorter duration of sequences and, in turn, lower frequency resolution. Thus, by reducing the number of bins, a similar frequency range is taken into account as in longer sequences.

Tagged responses are analyzed at the stimulation frequency (f) and its harmonics, multiple integers of the stimulation frequency (nf). Due to non-linear properties of the stimulation and brain processes, responses can be spread across the stimulation frequency and harmonics, and thus the signal needs to be aggregated to provide a balanced comparison of conditions (Norcia et al., 2015; Retter et al., 2021). The systematic review reveals that recent studies statistically evaluate the presence of harmonics and proceed by summing significant harmonics, consistent with current recommendations (Retter et al., 2021).

Previously, it has been demonstrated that the pattern of dominant harmonics (i.e., the harmonic in which the highest amplitude is elicited) is relatively stable in adult participants over the course of 2 months (Dzhelyova et al., 2019). Very little is known regarding potentially systematic relations of harmonics with age, and, more generally, the number of studies comparing responses across age are still limited. In studies on face categorization (De Heering and Rossion, 2015; Rossion et al., 2015), the number of harmonics was reduced in 4-to-6-month-old infants compared to adults. In infants, only the first harmonic (i.e., the stimulation frequency) reached significance, whereas harmonics 1 to 14 were significant in adults. In a study on categorization of animals and furniture items, the number of significant harmonics increased from one to three harmonics during infancy (four to 11 months), compared to significant harmonics 1 to 22 in adults (Peykarjou et al., under review see footnote 1). In a similar vein, a study on facial trustworthiness discrimination observed harmonics 2 to 7 to be significant in adults, whereas only two harmonics reached significance in six-month-old infants (Baccolo et al., in preparation see footnote[3]). Thus, though more systematic work is needed, it appears that responses are confined to fewer harmonics during infancy than in adulthood.

In general, whether and how many harmonics are elicited varies much between tasks and samples, and it is currently still unclear which factors drive these differences. Importantly, as harmonics may differ between conditions and samples, they need to be considered during analysis to provide an unbiased estimation of responses elicited. If harmonics are present, they can unequivocally be attributed to the stimulation, as no enhanced harmonic responses are elicited when there is no stimulation frequency f.[4] Some frequency tagging papers have passed over the analysis of harmonics (e.g., Barry-Anwar et al., 2018; Bekhtereva et al., 2018), which may lead to biased results. There may be cases where the first harmonic does not reach significance (e.g., Peykarjou et al., 2022), so focusing exclusively on the first harmonic would lead to the erroneous conclusion that no frequency tagging response was observed. On the other hand, even if the fundamental frequency is present, when conditions vary regarding the recruitment of harmonics, employing only the first harmonic would bias comparisons of conditions. Thus, inspection and balanced inclusion of harmonics is recommended, for example by averaging across conditions and extracting significant harmonics based on this overall grand-average.

Another decision that is potentially influenced by age (or other sample characteristics) is selecting regions of interest (ROIs).

---

[2] Note that the only case of noise at one precise frequency bin concerns electrical line noise at 50 or 60Hz depending on the region of recording. This precise frequency should therefore be excluded from analysis.

[3] Baccolo, E., Peykarjou, S., Quadrelli, E., Conte, S. and Macchi Cassia, V. (in preparation). Neural discrimination of facial cues to trustworthiness in adults and six-month-old infants as revealed by Fast Periodic Visual Stimulation.

[4] Of course, this assumes that adequate corrections for multiple tests across harmonics have been adopted.

Based on the existing evidence, it seems that general visual activation associated with the flicker stimulation can be recorded in infants and adults likewise over the medial occipital cortex, predominantly on electrode Oz and spreading to O1/O2 (e.g., Leleu et al., 2020; Rossion et al., 2020; Pauen and Peykarjou et al., under revision [5]). This is independent of the type of visual stimuli employed, consistent with the assumption that this base response reflects general visual processing. Selecting appropriate ROIs for the specific cognitive processes involved in frequency tagging studies in an age-fair way is much more challenging. Drawing on the limited number of studies in which the same paradigm was conducted with infants and adults (Baccolo et al., in preparation see footnote 3; Peykarjou et al., under review see footnote 1; de Heering and Rossion, 2015; Rossion et al., 2015), it seems that categorization responses are often confined to smaller ROIs in infants than adults. Adult data may provide a reasonable starting point for determining the spatial layout of frequency tagging responses, but the response may be restricted to fewer electrodes early in development.

A final challenge in developmental frequency tagging studies pertains to the reliability of responses. First reports on reliability in frequency tagging with adults are highly promising, indicating that base and cognitive responses can be measured with high retest-reliability in adults (Dzhelyova et al., 2019). As low reliabilities may pose a challenge in ERP studies (Cassidy et al., 2012; Munsters et al., 2019), this finding is very promising and may provide a basis for individual diagnosis in frequency tagging. So far, there are no reports on reliability of frequency tagging in infants, but due to the limitations discussed already, it can be expected to be lower than in adults. Among other factors, reliability increased linearly with increasing stimulation time in adults (Dzhelyova et al., 2019), so the shorter stimulation time with infant samples may limit reliability.

To sum up, employing frequency tagging with infants holds great potential for cognitive research, but is associated with challenges regarding stimulation and analysis. The stimulation needs to be adjusted to accommodate infants' attentional limits by decreasing duration of sequences and limiting the number of conditions. During analysis, artefacts need to be handled with care, and decisions on inclusion of harmonics and ROIs need to be made in an age-fair way.

In the following, the influence of analysis decisions on response strengths will be explored empirically, drawing upon a large dataset of categorization by seven-month-old infants in a simple oddball categorization task. The empirical part will focus on comparing different criteria for retaining sequences, evaluating the relation between looking time and EEG responses, the influence of the frequency range for baseline correction and the number of sequences per participants, as well as the influence of considering different numbers of harmonics.

Analyses of infant frequency tagging data

Stimuli and design

Infants were presented with sequences of artificial stimuli flickering at 6 Hz (6 items/s). At every fifth position, the type of stimulus changed, corresponding to 1.2 Hz ($6/5 = 1.2$, see Figure 2). Two types of stimuli were employed, red-orange/curvy shapes and blue-green/straight-edged shapes, with 10 individual exemplars of each type varying in size, color and number of pedals and parts (Pahnke, 2007; Ropeter and Pauen, 2013). Global luminance contrast did not differ across categories ($p > 0.05$). Pixel size of images was 545 (width) x 542 (height), corresponding to approximately 8 × 8 degrees of visual angle. Participants were compiled across two studies, so the maximum of sequences varied between 8 (Experiment 2) and 16 (Experiment 1).

Sequences were initiated manually when participants looked attentively at the screen and showed an artifact-free EEG signal. Each sequence started with a uniform grey background (random length, 5–6 s) while a short doorbell-sound was played to attract attention if necessary. Each sequence lasted 24 s, consisting of a fade-in of 2 s (contrast of images against the background increased gradually from 0 to 100%), a stimulation sequence of 20 s, and a fade-out of 2 s. Fade-in and fade-out phases served to avoid surprise reactions and blinks. Images changed size (+/− 5%) and were presented using a sinusoidal contrast-modulation function. Testing ended when participants became too inattentive or fussy.

Sample

A sample of $N = 124$ (mean age = 7 months, 14 days, SD = 9 days, 59 females) participants were aggregated across two studies. On average, participants participated in $M = 6.91$ sequences (SD = 2.60, range 1–14), of which $M = 1.09$ (SD = 1.43, range 0–6) were excluded due to bad data (see specifics in Recordings and Analyses). A data base of $M = 5.82$ sequences (SD = 2.94, range 1–12) was thus employed for all analyses. All data combined here was based on the same stimulation, a categorization task with artificial categories as described in the section on stimuli & design. The first study evaluated the impact of familiarization on categorization responses. Categorization in the frequency tagging paradigm was compared following three different familiarization conditions: an ERP familiarization with stimuli from the base category (10 different exemplars presented 5 times for 1 s each); familiarization to the base category with subsequent presentation of a stimulus from the contrasting category (10 different exemplars presented 5 times for 1 s each, one stimulus from contrasting category presented for 15 s), and no familiarization. A previous study indicated that categorization

---

[5] Pauen, S. and Peykarjou, S. (under revision). Infants' Fast Neural Categorization of Artificial Objects: The impact of stimulus- and task-characteristics Developmental Psychology.

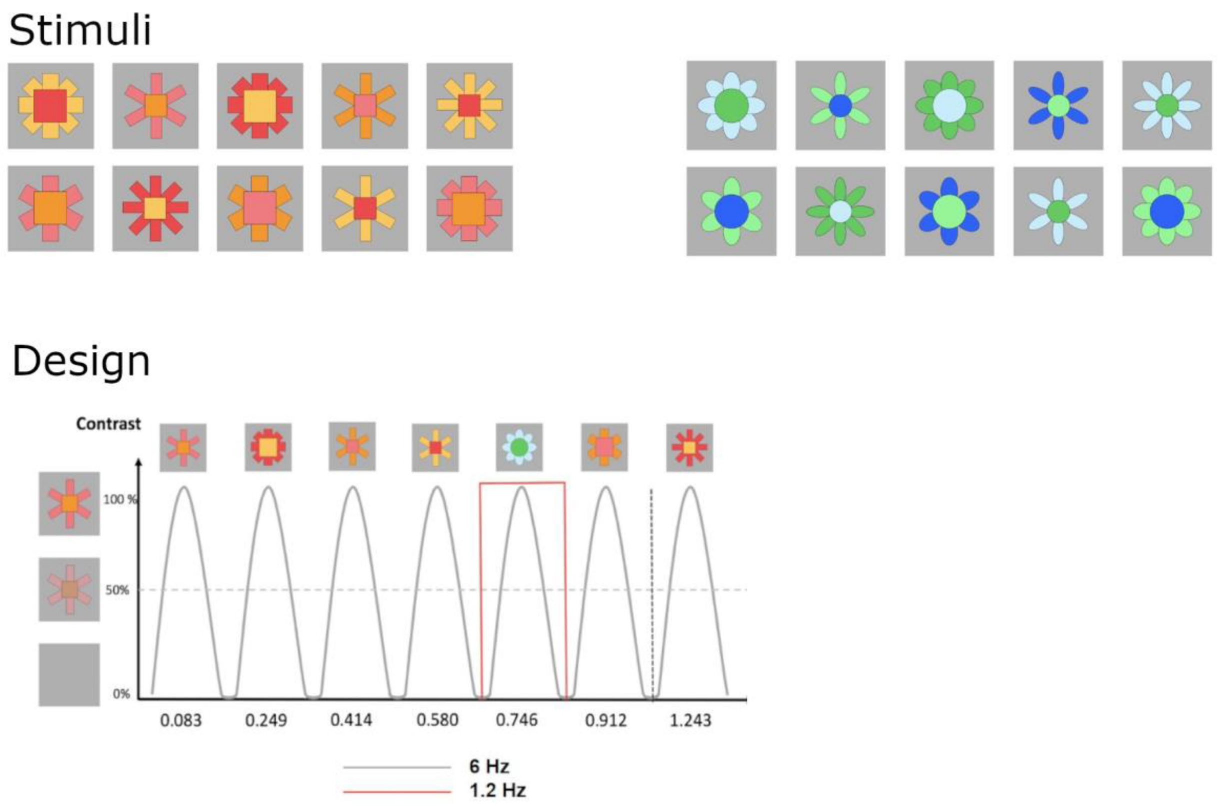


FIGURE 2
Schematic illustration of the stimuli and experimental paradigms. Ten different angular red-orange and round blue-green stimuli were employed, with contrasting categories differing in the shape of all parts (angular vs. round outlines) and colour spectrum. Phase-scrambled control images were created from original stimuli. In 20-s-sequences, images were presented by sinusoidal contrast modulation at a rate of 6cycles per second=6Hz (1cycle ≈ 170ms). Angular red-orange stimuli (A) were presented as standards, with round blue-green stimuli presented as deviants at every 5th cycle (B; 6/5Hz=1.2Hz).

did not differ based on familiarization condition (Experiment 2 in Pauen and Peykarjou, under revision see footnote 5), so data from participants was collapsed here across familiarization conditions. In the second study, subjects participated twice (T1, T2) in the same categorization task with an interval of approximately 2 weeks (data partly reported in Exp. 1, Pauen and Peykarjou, under revision see footnote 5). Only data from T1 are included into the aggregated sample. 41 additional infants were tested but excluded due to excessive crying (Experiment 1 $N=2$, Experiment 2 $N=2$), fussiness (Experiment 1 $N=7$, Experiment 2 $N=7$), too many artefacts (Experiment 2 $N=18$), or technical problems (Experiment 1 $N=2$, Experiment 2 $N=3$). In accordance with the terms provided by the local ethics committee of Heidelberg University that approved the general procedure, written informed consent was obtainted from caretakers.

Procedure

In a dimly lit and quiet room, infants sat on the lap of their caregivers with a computer monitor in front of them (looking distance of approx. 80 cm). Parents were asked not to interact with their infant during data collection. Infant behavior was recorded on video, and brain responses were registered using a BrainProducts actiCap (Gilching, Germany) with 32 active Ag-AgCl electrodes arranged according to the 10-10-system and a right mastoid reference.

Recordings and analyses

Sampling rate was set at 250 Hz and the EEG signal was amplified using BrainAmp. Impedances were considered acceptable if <20 kΩ. EEG processing steps was performed using Matlab 2012b (The Mathworks) and Letswave.[6] Data was band-pass filtered at 0.1–100 Hz using a 4th order Butterworth filter and segmented for each sequence. Noisy channels were identified and pooled from surrounding channels (for a maximum of three channels) and a common average reference computation was applied to all channels.

---

[6] https://nocions.github.io/letswave6/

Preprocessed data segments were cropped to an integer number of 6 Hz cycles beginning 5 s after onset of the sequence (after fade-in) until 20 s (120 cycles, 5,000 time bins in total = 20 s). Sequences were averaged per participant before a Fast Fourier Transform (FFT) was applied to these averaged segments to extract amplitude spectra for all channels (square root of sum of squares of the real and imaginary parts divided by the number of data points). By averaging prior to the FFT, the contribution of responses not in phase (predominantly noise) is reduced (Rossion et al., 2020). Frequency analysis yielded spectra with a high frequency resolution of 0.05 Hz (1/20s).

The utility of different criteria for retention of sequences was compared by evaluating the number of sequences and participants included by criterion, the relation of each criterion with looking time data and the resulting pattern of base and categorization data. Amplitude at the base frequency (6 Hz) and its harmonics represents the brain's response to the general visual stimulation and has previously been used as an index of whether the infant paid attention to the screen (de Heering and Rossion, 2015; Peykarjou et al., 2017). However, as the base frequency is also a multiple integer of the categorization frequency (5*1.2 Hz), both general stimulation and categorization processes may contribute to it. Base rate criteria can be based on SNR or Z-scores and are generally evaluated on channels O1, O2, Oz at 6 and, in several studies, 12 Hz. Here, the following base rate criteria will be evaluated: On O1, O2 or Oz, for 6 or 12 Hz, at least 1 SNR > 1.5; 1, 2 or 3 SNR > 2; 1 Zscore > 2.33; 2 or 3 Z-scores > 1.64. All following steps employ the base rate criterion selected based on this comparison. EEG amplitude at the frequency of $F/5 = 1.2$ Hz and its harmonics (i.e., $2F/5 = 2.4$ Hz, $3F/5 = 3.6$ Hz…) served as an index for categorization. In accord with prior work and recommendations (e.g., Peykarjou et al., 2017; Retter et al., 2021), responses were summed across significant harmonics. The range of harmonics for quantification of responses was defined by extracting all harmonics above noise level at a $p < 0.05$threshold. This threshold was Bonferroni-corrected for the number of tests run, in this case three independent tests for electrodes O1, O2 and Oz. Harmonics 1–19 reached significance and were summed for the categorization response, and harmonics 1–5 for the base response.

Looking time was coded offline. Twenty percent of videos were independently double-coded by two student assistants to estimate reliability, which reached high levels, Cronbach's alpha = 0.96. Based on these analyses, a base rate criterion was selected and employed for all subsequent analyses.

To measure the magnitude of activity at pre-defined bins of interest, baseline corrected amplitudes were computed by subtracting the average amplitude of surrounding bins. SNRs were computed by dividing the signal by the amplitude at the neighboring frequency bins and used to display response patterns. Similarly, Z-scores were calculated by subtracting the average baseline-corrected amplitude of surrounding bins from each bin and dividing this by the standard deviation (SD) of the surrounding bins. The number of bins employed for statistical analysis was compared for 10 and 20 bins (removing the 2 bins with highest and lowest amplitude), corresponding to a frequency range of $+-0.35$ to $+-0.70$ Hz (excluding the immediately adjacent bins). Based on these analyses, the number of bins included in all further analyses was selected.

As a next step, the number of sequences included was varied systematically from 1 to 9 to evaluate how base and categorization response strengths relate to the number of sequences. To explore the impact of considering harmonics on response strengths, the number of harmonics was then increased systematically from 1 to 29 for the categorization response (always excluding harmonics corresponding to the base frequency, i.e., 5, 10, 15, 20, 25), and from 1 to 10 for the base response. The resulting data pattern was compared to the analysis based on harmonics selected according to Z-scores. Based on prior work (Peykarjou et al., 2017) and visual inspection of grand averages, all analyses were focused on visual responses recorded at O1, O2, and Oz. For all comparisons of different approaches with approximately similar numbers of participants, Friedman tests, a non-parametric alternative for repeated-measures ANOVAs, were employed to test for statistical significance.

Results

Evaluating different criteria for retention of sequences

Sequence inclusion criteria were evaluated first by comparing the numbers of sequences and participants kept by criterion, and in a second step by examining the resulting pattern of base and categorization data and calculating the relation of responses with looking time data. An overview of the results can be found in Table 2.

Across all criteria for retaining sequences, the same harmonics were summed. Inspection of Bonferroni-corrected Z-scores revealed that a significant base response was obtained across harmonics 1 through 5 regardless of sequence retention criterion, while the 6th harmonic failed to reach significance. A significant categorization response was observed across harmonics 1–19, with the following exceptions: In the analysis on 3 channels with an SNR > 2 and 1 channel with Z > 2.33, the 18th harmonic failed to reach significance, and for 3 channels Z > 1.64, the 2nd harmonic failed to reach significance. Vice versa, the 21st harmonic reached significance only when considering 2 channels Z > 1.64. For consistency, the dominant pattern of significant harmonics 1–19 was included.

As revealed by Table 2, rates of sequence exclusion ranged from.06 to 2.86. On average, criteria which were focused on only one channel excluded less than one sequence per participant. Likewise, while participant exclusion rates ranged from 0 to 13.7%, no more than one participant was excluded when criteria focused on only one channel. In contrast, when 2 or 3 channels were considered, more than 1 sequence and up to 13.7% of participants were excluded.

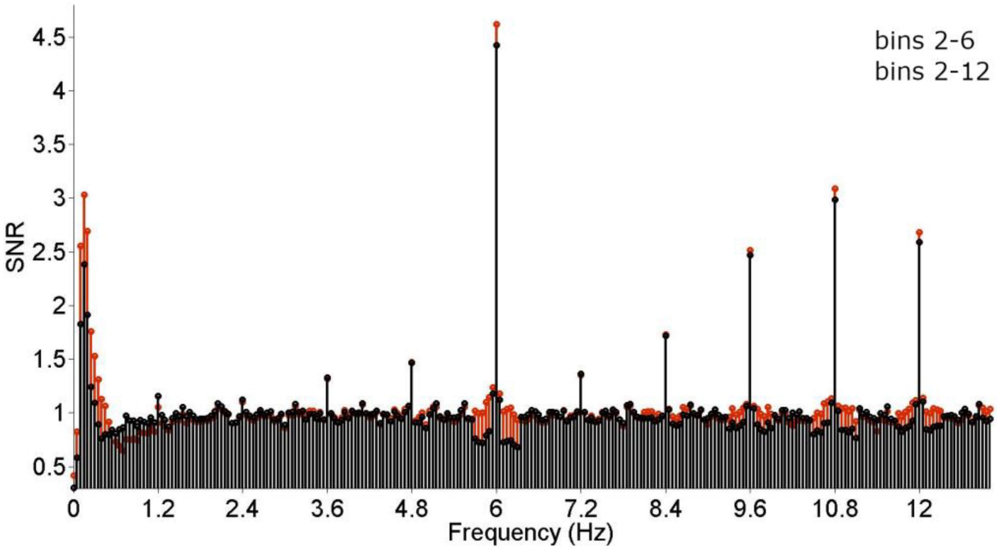


FIGURE 3
Grand-average frequency-spectrum responses. Base (6+12Hz) and categorization (1.2, 2.4, 3.6... 10.8Hz) responses are clearly visible and highly significant. The effect of the number of bins employed for baseline correction was evaluated by comparing signal-to-noise-ratios (SNRs) was compared across 10 (2–6) and 20 (2–12) bins. While employing a larger number of bins did not seem to affect responses at higher frequencies (from 6Hz onwards) detrimentally, response strength was stronger when employing fewer bins at lower frequencies (below 6Hz). Crucially, the effect seemed strongest at 1.2Hz, corresponding to the stimulation frequency.

Signal-to-noise-ratios (SNRs) of resulting response patterns varied only slightly across criteria. For the categorization response, variation of average SNRs was no more than 4%, and less than 5% for the base response. Regardless of which criterion was employed, a substantial but non-exhaustive correlation (r =0.29–0.39) between the base rate response and looking times emerged.

Overall, sequence retention criteria varied regarding participant retention, but negligibly regarding response patterns. Thus, a more inclusive criterion seems desirable considering parsimony of data collection and representativeness of the obtained sample. Therefore, in accord with the majority of prior research as indicated by the systematic literature review (Table 1), SNR > 2 on at least one channel at 6 or 12 Hz was selected. All further analyses are based on sequences included according to this criterion.

Comparing the number of bins employed for baseline correction

Next, the number of bins employed during baseline correction was compared. In adults watching 60-s-sequences, typically 20 bins are employed (Stothart et al., 2017; Rossion et al., 2020), whereas in infant studies, mostly 10 bins are used (see Table 1). These two criteria were evaluated by comparing the resulting grand-average SNRs and the results pattern of individual participants. As evident from Figure 3, while the use of 20 bins results in visually higher baseline-corrected amplitudes at higher harmonics of the categorization frequency, it decreases the response at 1.2 Hz. This is most likely due to the steep amplitude curve at the very low spectrum, which unproportionally skews the baseline correction at 1.2 Hz. Statistical tests including all significant harmonics (i.e., 1–19 for the categorization response, 1–5 for the base response) did not indicate an effect of the number of bins considered, $p < 0.05$.However, a hypothesis-driven analysis on 1.2 Hz only revealed that the response at this frequency is indeed reduced when employing 10 bins, $X^2(123, N = 123) = 25.609$, $p < 0.001$. Considering that measuring at the stimulation frequency 1.2 Hz is essential to this paradigm, employing 10 bins is recommended for studies using shorter presentation times (<= 30 s).

Systematically increasing the number of included sequences 1 to 9

The number of sequences included was increased systematically from 1 to 9. Naturally, the number of participants contributing to the analysis decreased with increasing numbers of required sequences. Thus, analysis was stopped at $i_{\text{sequences}} = 9$ as the number of participants providing larger numbers of sequences was too small ($N < 10$).

Regardless of the number of sequences considered, all analyses revealed significant responses, Zs > 11. Numerically, the base response was smaller when averaging across <4 sequences compared to higher numbers (Figure 4). As evident from Figure 5, the categorization response was numerically lower when averaging across 2–3 sequences. Importantly, responses were comparable when taking into account all children providing at least one sequence and higher sequence numbers. This descriptive analysis was not supplemented by inferential statistics as the number of subjects contributing to the different conditions varies substantially, violating assumptions of statistical tests.

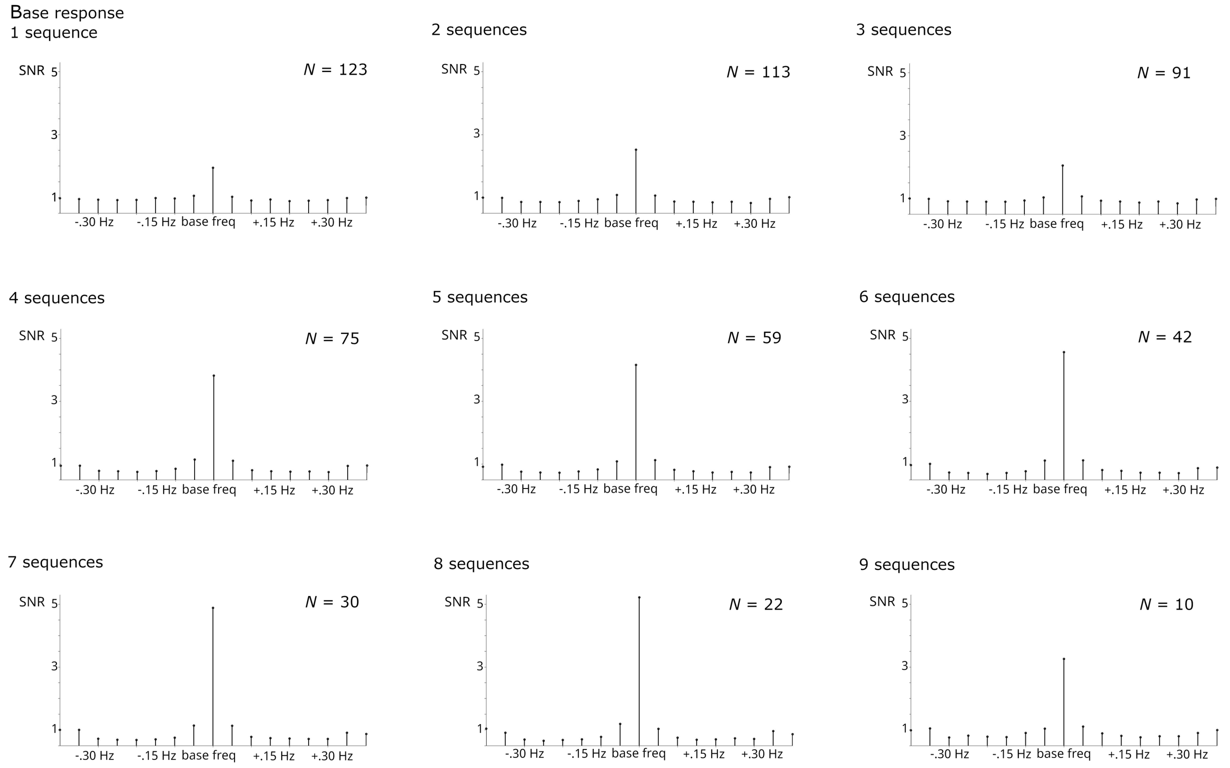


FIGURE 4
SNR of grand-average base responses summed across harmonics 1–5 and compared across numbers of sequences averaged. The respective *N* contributing to each analysis is specified per graph. Highly significant base responses were obtained in all analyses.

Considering the problem of different sample sizes and the trade-off between numbers of sequences and participants who provide sufficient data to be included. In a control analysis with 20 participants who contributed 8 usable sequences, it was evident that increasing numbers of sequences per participant increase both base and categorization responses (Figures 6, 7). This was supported by statistical analyses which showed that the number of sequences included has a significant effect on response strength for the categorization response, $X^2(7, N = 20) = 17.951$, $p < 0.05$. The effect on the base response strength was only marginal, $X^2(7, N = 20) = 12.651$, $p = 0.081$. Bonferroni-corrected pairwise comparisons did not reach significance.

Systematically increasing the number of harmonics considered

To evaluate the effect of including harmonics into response measurement, the number of harmonics was systematically increased from 1 to 10 for the base response (corresponding to frequencies 6, 12… 60) and from 1 to 29 for the categorization response (corresponding to frequencies 1.2, 2.4… 34.8, always excluding harmonics corresponding to the base frequency). In addition, a Z-score criterion was used to select the number of harmonics summed as in prior work (base: 1–5, categorization: 1–19).

As evident from Figure 8, lower base harmonics (1–3) contributed strongly to the summed responses, whereas including higher harmonics (particularly 7–10) descriptively reduced the SNR at the stimulated frequencies. However, estimation of SNR at the surrounding bins employed for baseline correction was also reduced by summing, so overall, including harmonics contributed to a cleaner estimation of base responses. In this case, the number of harmonics for summing obtained by the Z-score criterion, 1–5, provides an optimal balance between high responses at the stimulated frequencies and evenly distributed responses across the surrounding bins. Friedman tests confirmed that the number of harmonics influenced base response strength, $X^2(9, N = 123) = 182.889$, $p < 0.001$. Post-hoc tests indicated that baseline-corrected amplitudes were significantly smaller when considering only the first harmonic (i.e., 6 Hz) compared to analyses including at least 3 harmonics, all ps < 0.05. No other comparison reached significance.

For the categorization response, including harmonics generally led to an increase in SNR at the stimulated frequencies. As evident from Figure 9, while the categorization response was visible from the 1st harmonic, signal strength increased numerically until summation of harmonics 1–9 and stabilized from that point onward. Signal strength at the surrounding bins was not much influenced beyond harmonics 1–6. Here, including harmonics (at least up to the 9th) seems desirable for optimal signal estimation. The analysis based on the Z-score criterion (harmonics 1–19) does not seem to have an advantage in terms of

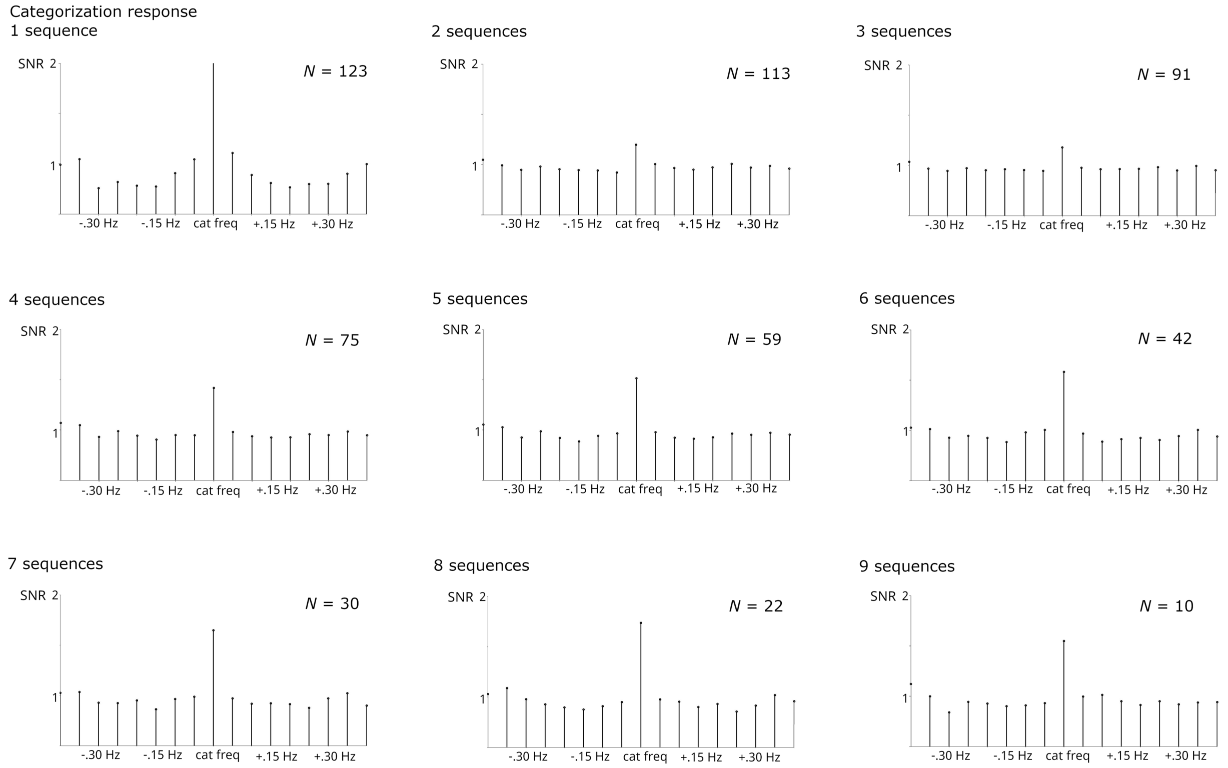


FIGURE 5
SNR of grand-average categorization responses summed across harmonics 1–19 and compared across numbers of sequences averaged. The respective *N* contributing to each analysis is specified per graph. Highly significant categorization responses were obtained in all analyses.

signal at stimulated frequencies or surrounding bins compared to related harmonic ranges (e.g., 1 through 11–1 through 29). For the categorization response, Friedman tests also confirmed that the number of harmonics influenced response strengths, $X^2(9, N = 123) = 1295.21$, $p < 0.001$. Post-hoc tests indicated that baseline-corrected amplitudes were significantly smaller when considering only the first two harmonics (i.e., 1.2 + 2.4 Hz) compared to analyses including at least 3 harmonics, all $ps < 0.05$. Moreover, analyses including at least 16 harmonics yielded higher amplitudes than those including up to 11 harmonics, all $ps < 0.05$.

Discussion

Based on a systematic literature review and re-analyses of data on categorization in seven-month-old infants, this manuscript evaluates stimulation and analysis decisions of infant frequency tagging studies employing the FPVS oddball paradigm. In the following, current standards and future recommendations will be summarized.

A major challenge in developmental neuroscience is the limited attentional capacity of participants. Prior infant frequency studies following the FPVS oddball paradigm have faced this challenge by reducing sequence duration to 15–30 s. The longer the sequence, the more stimulation can be delivered, and the higher is the spectral resolution of resulting data. Thus, maximizing sequence duration is highly desirable. Given the apparent feasibility of 30-s-sequences, this duration is recommended for future studies.

Relatedly, the number of sequences that can be presented is limited by infants' attentional capacities. The systematic literature review indicates that on average, about 9 sequences can be presented, so limiting the number of conditions to 2–3 within-subject factors seems reasonable. However, this number will need to be adjusted to the nature of the stimulation and the age of participants, with more engaging presentations (e.g., concurrent visual and auditory stimulation) and older participants being likely to provide more usable data.

Analyzing the effect of the number of sequences included may aid in determining minimum requirements for participant inclusion. Generally, this decision follows a trade-off: The more sequences are set as a requirement for participant retention, the more data is lost due to exclusion of participants. Indeed, increasing the number of required sequences from 1 to 9 led to exclusion of 113 participants. Most important is probably the effect of excluding all participants providing less than two usable sequences. This led to exclusion of 10 participants (8%). However, while the amount of data loss may be considered passable, excluding those participants decreased the categorization response numerically, reflecting the loss of meaningful data. One sequence in the FPVS oddball paradigm displayed at 6 Hz contains, depending on the precise duration, between 90 and 180 images

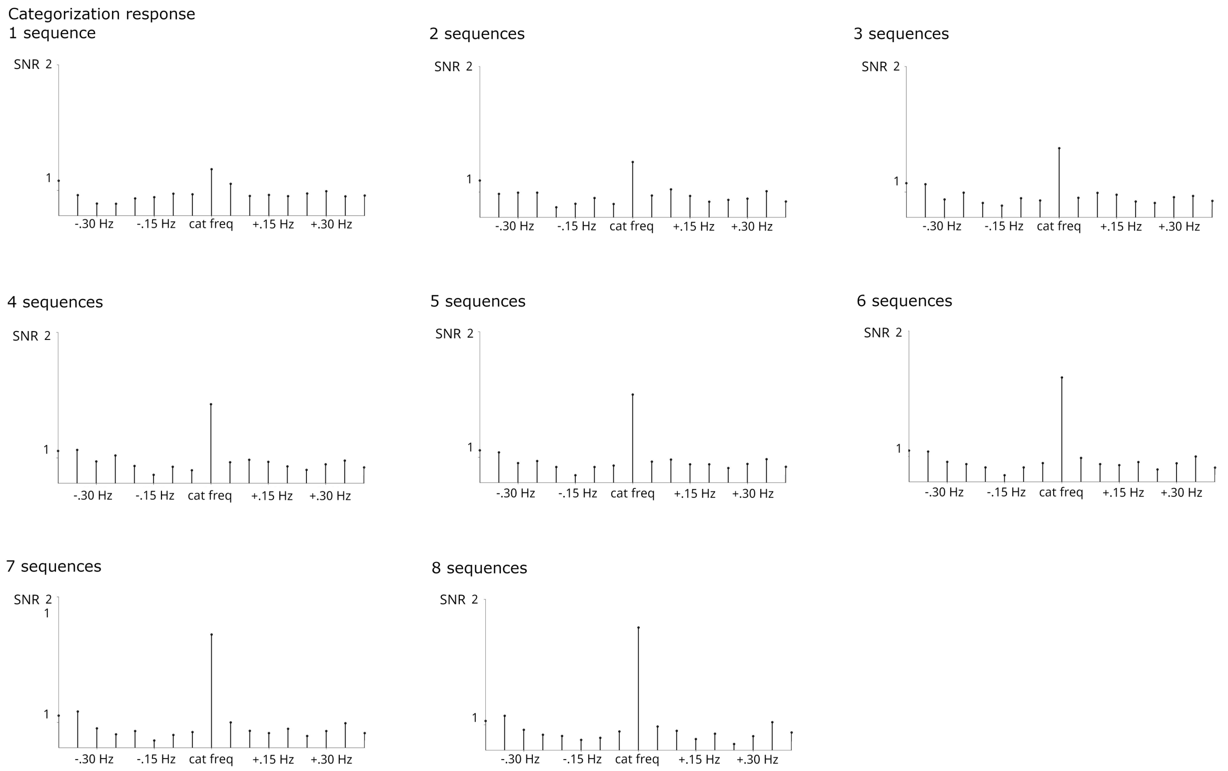


FIGURE 7
SNR of grand-average categorization responses summed across harmonics 1–19 and compared across numbers of sequences averaged for a subsample of *N*=20 who provided usable data for at least 8 sequences. The respective *N* contributing to each analysis is specified per graph. Significant categorization responses were obtained in all analyses. Averaging across larger numbers of sequences increased response strengths.

(75–150 base stimuli, 15–30 stimuli from the contrasting category). In addition, generalization within and discrimination across categories is inherent within each sequence and reflected by the categorization response. This argues in favor of including all participants contributing at least one usable sequence to the analysis (or each condition).

Of all presented sequences, several might need to be excluded based on data quality and participant attention. The literature review indicated that data quality was mostly assessed by visual inspection, and data quality issues were treated mainly by exclusion of sequences or channel interpolation. It seems desirable to supplement visual inspection through more automated artifact detection approaches, to increase preprocessing accuracy, objectivity, and efficiency. Certainly this will be an important avenue for future studies.

Sequences with sufficient data quality still need to be screened for attention towards the presentation, as infant attention naturally fluctuates throughout the session. Different approaches have been employed for this step, based on looking times or base rate responses during a given sequence. Comparing such approaches in a re-analysis of a large data-set comprising $N = 124$ seven-month-olds reveals that sequence and participant retention rates are influenced by the criteria employed, whereas the resulting data pattern and the relations of base responses with looking time do not vary much. To maximize data retention and in accord with most prior studies, use of an SNR (>2) or Z-score (>2.33) criterion based on only one channel (O1, O2, or Oz) can thus be recommended.

Due to the shorter sequence duration employed with developmental populations, the frequency resolution of resulting data is decreased. This should be taken into consideration when estimating the signal at each bin and may be especially important in the low frequency spectrum, which is typically characterized by a larger amount of noise and distortions due to the FFT. Thus, employing a large range of bins for noise estimation, e.g., to determine Z-scores or SNRs, may disproportionately affect responses in the low spectrum, such as 1.2 Hz. This was reflected in the re-analysis reported here, which showed a decrease in response strength at lower frequencies (<6 Hz), but an increase of responses at higher frequencies (>6 Hz) when employing 20 bins for baseline correction. It is therefore recommended to employ a reduced frequency range (e.g., $\sim + -0.35$ Hz, corresponding to 10 bins) when stimulating for up to 30 s.

Two final decisions that should be made *a priori* regard the selection of harmonics for summing and electrodes for analysis. Due to the limited number of studies comparing infant responses in the same paradigm to that of older participants, recommendations are based only on anecdotal evidence at this point. It seems likely that responses can

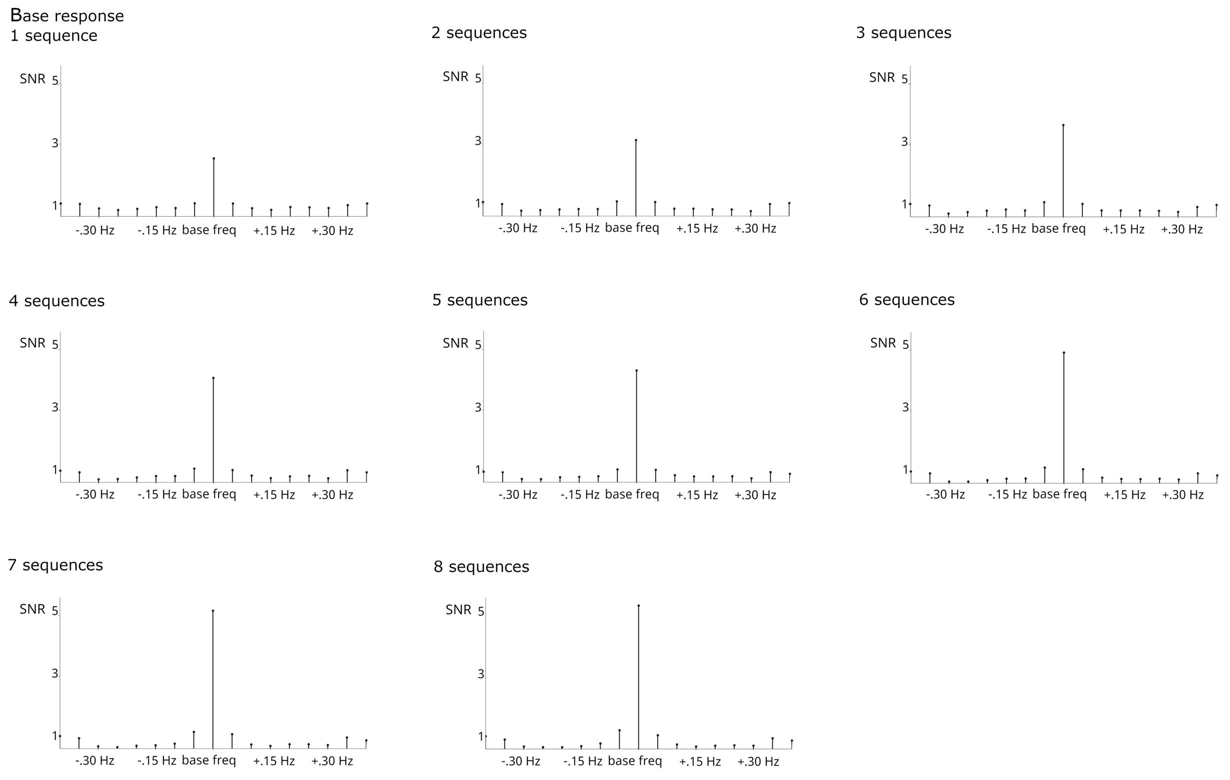


FIGURE 6
SNR of grand-average base responses summed across harmonics 1–5 and compared across numbers of sequences averaged for a subsample of N=20 who provided usable data for at least 8 sequences. Highly significant base responses were obtained in all analyses, with a tendency for increasing amplitude with number of sequences averaged.

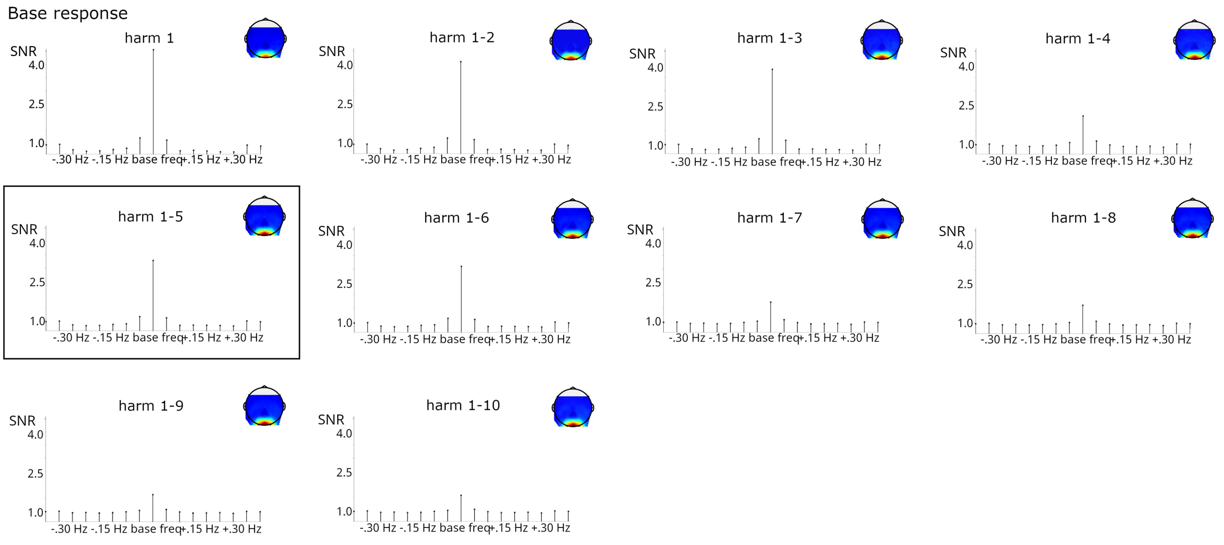


FIGURE 8
SNR of base grand-average responses summed across increasing numbers of harmonics (considering only the 1st or up to harmonics 1–10). Highly significant responses were obtained in all analyses. The analysis on harmonics 1–5, marked by a black box, was indicated by the Z-score criterion.

be confined to fewer harmonics and electrodes early in life than later on, so simply basing these decisions on adult data does not seem appropriate. Therefore, similar, *a priori* defined criteria should be employed to select electrodes and harmonics across age-groups. For harmonics, a selection based on Bonferroni-corrected significance levels of

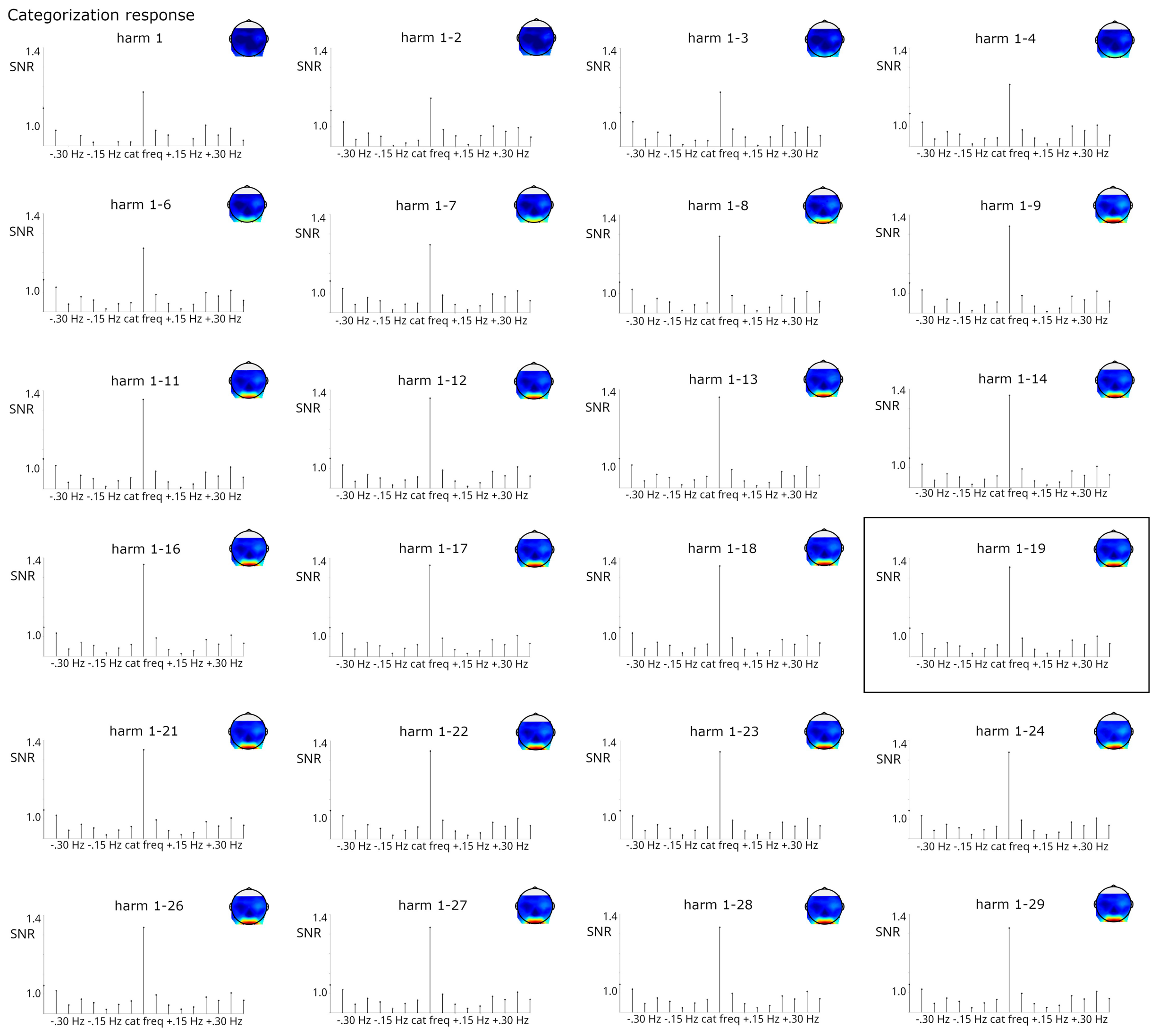


FIGURE 9
SNR of categorization grand-average responses summed across increasing numbers of harmonics (considering only the 1st or up to harmonics 1–29, always excluding harmonics corresponding to the frequency). Highly significant responses were obtained in all analyses. The analysis on harmonics 1–19, marked by a black box, was indicated by the Z-score criterion.

harmonics can be recommended (Retter et al., 2021). This selection seemed optimal in the analysis of base responses. In the analysis of categorization responses, while summing across several harmonics clearly had advantages in terms of signal strength at stimulated and surrounding frequencies, none of the analyses could be clearly favored, as the responses remained relatively stable from summing harmonics 1–11 onward. In such a case, considering Z-scores may provide an objective criterion for deciding which harmonics to include.

For electrodes, decisions are generally based more on theoretical accounts and prior studies but should be open to adjustments based on age-group. In any case, transparent and *a priori* selection is advisable, for example through preregistering analyses.

Limitations

Given the particularities of the approach, the systematic review was limited to studies employing the FPVS oddball paradigm with infants, so this overview and the recommendations derived are likewise limited. However, they may be useful for other developmental frequency tagging studies following related stimulation paradigms.

Moreover, the data re-analysis was based on only one age-group (seven-month-olds), and a categorization paradigm with artificial, unfamiliar stimuli that should not evoke much prior experience. Accordingly, categorization responses were elicited primarily in the occipital region. In prior studies using more familiar categories, responses were sometimes also recorded in the occipital region (Peykarjou et al., under review see footnote 1; Barry-Anwar et al., 2018), but in other cases at more parietal and/or temporal regions (de Heering and Rossion, 2015; Leleu et al., 2020). Future work should certainly test and generalize analysis recommendations to categorization tasks involving other regions of interest. Moreover, the categorization response elicited in the present study is much stronger than in most studies using familiar, natural categories (Baccolo et al., in preparation see footnote 3; De Heering and Rossion, 2015; Barry-Anwar et al., 2018; Leleu et al., 2020; Rekow et al., 2021). Upon investigating a subtler response, it seems likely that analysis decisions will carry even more weight than for the robust response observed in the current paradigm.

While these limitations may of course restrict generalizability, most analysis decisions evaluated here should in principle be independent of the specific age-group and stimuli presented. The increased noise level in the low frequency spectrum will be evident across infant samples of all research questions, arguing in favor of generally employing 10 bins for baseline correction. While sequence and participant retention rates will vary with age-group and stimulation, the general pattern of being more inclusive (and thus representative) when employing a base-rate criterion directed at only one occipital channel will likely be independent of these aspects.

In contrast, of course the response patterns resulting from different sequence inclusion criteria may vary. On the one hand, there is no indication to assume that being less inclusive (i.e., requiring base rate responses on more channels, or excluding participants providing less than one usable sequence) will lead to enhanced data cleanness. On the other hand, keeping more data will increase the representativeness of analyses for the tested sample. The observation that the number of sequences included did, overall, not change base or categorization data patterns much likely reflects the overall amount of data analyzed. It seems probable that the effects are less driven by the number of sequences *per se* but rather the overall amount of data (i.e., participants * sequences), which again argues for an inclusive criterion. However, his question should also be addressed in future studies, as it is likely not independent from study-specific factors.

Open questions and avenues

As we are beginning to systematically develop and evaluate developmental frequency tagging, some central aspects remain to be addressed in future work. First, as already mentioned in the Introduction, frequency sweeps testing the optimal flicker speed with infant participants are lacking. These would be highly desirable not only to optimize the stimulation of developmental frequency tagging, but also to further our understanding of the development of basic neural processes. For instance, certain frequency bands have been associated with particular cognitive functions (e.g., theta range: ~ 4 Hz in infants; learning; e.g., Begus et al., 2015, alpha range: ~ 6 Hz in infants; attention; e.g., Friese et al., 2013; Gulbinaite et al., 2019). Labeling frequency bands in a consistent way across development is challenging, and more basic research is needed to ascertain the relations between cognitive functions and given frequencies at different ages. Moreover, a systematic analysis of the assumption that stimulating at certain frequencies supports particular cognitive processes would require sampling a large range of frequencies. Inherently, stimulating at a given frequency also means presenting stimuli for a certain period of time, and changing presentation times may in itself alter cognitive processes. Therefore, systematically varying the speed of stimulation and associating it with response patterns and cognitive functions seems an important avenue for future work.

Moreover, to be able to apply frequency tagging more broadly, reliability of responses needs to be evaluated. The promising reliabilities obtained in adult face individuation studies (Dzhelyova et al., 2019; Stacchi et al., 2019) give rise to the hope that frequency tagging may provide a window into cognitive processes that is suited even for differential approaches and longitudinal research. Both within-and between-sessions reliability of base and categorization responses should be examined to ascertain the utility of the approach. While relatively long inter-session intervals (2–6 months) have been employed with adults, developmental progress makes it unlikely that responses will remain stable over such time-windows. However, frequency tagging seems to have the potential to provide reliable estimates of categorization, at least on the time-scale of a few weeks.

More broadly speaking, the paradigm holds potential for investigating cognitive processes beyond categorization. For example, it might be useful for investigating rule learning, the process of extracting and generalizing repetition-based, abstract patterns to new elements (Marcus et al., 1999). Moreover, it could be used concurrently (and even orthogonally) in other tasks to track attention, when a cognitive task is presented in the context of an object or background flickering at a given frequency. The relation between frequency tagging responses and attention has been broadly established in studies with adult participants (Morgan et al., 1996; Chen et al., 2003; Toffanin et al., 2009). Similarly, the relation of infant frequency tagging responses to overt and covert visual attention has been established previously (Christodoulou et al., 2018), but the approach has so far not been used as a control mechanism during an independent task. The dissociated and objective measure of the base rate response could become a valuable tool in developmental cognitive neuroscience.

Conclusion

Frequency tagging is a tool increasingly utilized in developmental cognitive neuroscience and has previously been employed to demonstrate fast and high-level face and object categorization in the first year of life (De Heering and Rossion, 2015; Kabdebon et al., 2022; Peykarjou et al., under review See footnote 1). Objectivity and a high signal-to-noise ratio are inherent to this approach, and its reliability and validity have been demonstrated in adult research. This paper has provided a systematic review of studies following the FPVS oddball paradigm and a re-analysis of a dataset comprising 124 seven-month-old infants to develop recommendations for future work following this approach. Future work should set a particular focus on automizing and objectifying artifact detection, assessing the reliability of infant frequency tagging responses, and systematically evaluating the development of infants' response patterns across a range of frequencies. Together, these endeavors will serve to enhance the usefulness of frequency tagging for research on cognitive development, in basic functions such as categorization and beyond.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethikkommission Fakultät für Verhaltens- and Empirische Kulturwissenschaften, Universität Heidelberg. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

SP conceptualized the manuscript, performed analyses, and wrote the manuscript.

Acknowledgments

I am grateful to all participating families and student research assistants involved in data collection and analyses, in particular Sabrina Junk, Mara Lehr, Cecile Palouzie, and Sara Vragolic. This work was supported by a grant from the Deutsche Forschungsgesellschaft (DFG, grant number PE 3125/1-1). For the publication fee, financial support was provided by the DFG within the funding programme "Open Access Publikationskosten" and by Heidelberg University.

Conflict of interest

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

Publisher's note

References

Adrian, E. D. (1944). Brain rhythms. *Nature* 153, 360–362. doi: 10.1038/153360a0

Adrian, E. D., and Matthews, B. H. (1934). The interpretation of potential waves in the cortex. *J. Physiol.* 81, 440–471. doi: 10.1113/jphysiol.1934.sp003147

Atkinson, J., Braddick, O., and French, J. (1979). Contrast sensitivity of the human neonate measured by the visual evoked potential. *Invest. Ophthalmol. Vis. Sci.* 18, 210–213.

Barry-Anwar, R., Hadley, H., Conte, S., Keil, A., and Scott, L. S. (2018). The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia* 108, 25–31. doi: 10.1016/j.neuropsychologia.2017.11.019

Begus, K., Southgate, V., and Gliga, T. (2015). Neural mechanisms of infant learning: differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biol. Lett.* 11:20150041. doi: 10.1098/rsbl.2015.0041

Bekhtereva, V., Pritschmann, R., Keil, A., and Müller, M. M. (2018). The neural signature of extracting emotional content from rapid visual streams at multiple presentation rates: a cross-laboratory study. *Psychophysiology* 55:E13222. doi: 10.1111/psyp.13222

Bertels, J., Bourguignon, M., De Heering, A., Chetail, F., De TIÈGE, X., Cleeremans, A., and Destrebecqz, A. (2020). Snakes elicit specific neural responses in the human infant brain. *Sci. Rep* 10: 7443

Cassidy, S. M., Robertson, I. H., and O'connell, R. G. (2012). Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology* 49, 659–664. doi: 10.1111/j.1469-8986.2011.01349.x

Chen, Y., Seth, A. K., Gally, J. A., and Edelman, G. M. (2003). The power of human brain magnetoencephalographic signals can be modulated up or down by changes in an attentive visual task. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3501–3506. doi: 10.1073/pnas.0337630100

Christodoulou, J., Leland, D. S., and Moore, D. S. (2018). Overt and covert attention in infants revealed using steady-state visually evoked potentials. *Dev. Psychol.* 54, 803–815. doi: 10.1037/dev0000486

Dawson, G. D. (1954). A summation technique for the detection of small evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* 6, 65–84. doi: 10.1016/0013-4694(54)90007-3

De Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *elife* 4:E06564. doi: 10.7554/eLife.06564

DeHaan, M. (2007). *Infant Eeg and event-related potentials*, Hove, England: Psychology Press.

Dobson, V. (1976). Spectral sensitivity of the 2-month infant as measured by the visually evoked cortical potential. *Vis. Res.* 16, 367–374. doi: 10.1016/0042-6989(76)90198-X

Dzhelyova, M., Jacques, C., Dormal, G., Michel, C., Schiltz, C., and Rossion, B. (2019). High test-retest reliability of a neural index of rapid automatic discrimination of unfamiliar individual faces. *Vis. Cogn.* 27, 127–141. doi: 10.1080/13506285.2019.1616639

Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., and Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded Eeg. *NeuroImage* 66, 642–647. doi: 10.1016/j.neuroimage.2012.11.002

Gentile, F., and Rossion, B. (2014). Temporal frequency tuning of cortical face-sensitive areas for individual face perception. *NeuroImage* 90, 256–265. doi: 10.1016/j.neuroimage.2013.11.053

Gulbinaite, R., Roozendaal, D. H. M., and Vanrullen, R. (2019). Attention differentially modulates the amplitude of resonance frequencies in the visual cortex. *NeuroImage* 203:116146. doi: 10.1016/j.neuroimage.2019.116146

Heinrich, S. P., Mell, D., and Bach, M. (2009). Frequency-domain analysis of fast oddball responses to visual stimuli: a feasibility study. *Int. J. Psychophysiol.* 73, 287–293. doi: 10.1016/j.ijpsycho.2009.04.011

Hoehl, S., and Wahl, S. (2012). Recording infant Erp data for cognitive research. *Dev. Neuropsychol.* 37, 187–209. doi: 10.1080/87565641.2011.627958

Kabdebon, C., Fló, A., De Heering, A., and Aslin, R. (2022). The power of rhythms: how steady-state evoked responses reveal early neurocognitive development. *NeuroImage* 254:119150. doi: 10.1016/j.neuroimage.2022.119150

Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., et al. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Dev. Sci.* 23:E12877. doi: 10.1111/desc.12877

Liu-Shuang, J., Norcia, A. M., and Rossion, B. (2014). An objective index of individual face discrimination in the right Occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* 52, 57–72. doi: 10.1016/j.neuropsychologia.2013.10.022

Liu-Shuang, J., Torfs, K., and Rossion, B. (2016). An objective electrophysiological marker of face individualisation impairment in acquired prosopagnosia with fast periodic visual stimulation. *Neuropsychologia* 83, 100–113. doi: 10.1016/j.neuropsychologia.2015.08.023

Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science* 283, 77–80. doi: 10.1126/science.283.5398.77

Morgan, S. T., Hansen, J. C., and Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4770–4774. doi: 10.1073/pnas.93.10.4770

Munsters, N. M., Van Ravenswaaij, H., Van Den Boomen, C., and Kemner, C. (2019). Test-retest reliability of infant event related potentials evoked by faces. *Neuropsychologia* 126, 20–26. doi: 10.1016/j.neuropsychologia.2017.03.030

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.6.4

Page, M. J., Mckenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The Prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:N71. doi: 10.1136/bmj.n71

Pahnke, J. (2007). *Visuelle Habituation Und Dishabituation Als Maße Kognitiver Fähigkeiten Im Säuglingsalter.* Heidelberg, Germany: Heidelberg University.

Petrig, B., Julesz, B., Kropfl, W., Baumgartner, G., and Anliker, M. (1981). Development of stereopsis and cortical binocularity in human infants: electrophysiological evidence. *Science* 213, 1402–1405. doi: 10.1126/science.7268443

Peykarjou, S., Hoehl, S., Pauen, S., and Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* 7:12526. doi: 10.1038/s41598-017-12760-2

Peykarjou, S., Langeloh, M., Baccolo, E., Rossion, B., and Pauen, S. (2022). Superior neural individuation of Mother's than Stranger's faces by five months of age. *Cortex* 155, 264–276. doi: 10.1016/j.cortex.2022.07.011

Poncet, F., Leleu, A., Rekow, D., Damon, F., Dzhelyova, M., Schaal, B., et al. (2022). A neural marker of rapid discrimination of facial expression in 3.5-and 7-month-old infants. *Front. Neurosci.* 16:901013. doi: 10.3389/fnins.2022.901013

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr. Clin. Neurophysiol.* 20, 238–248. doi: 10.1016/0013-4694(66)90088-5

Rekow, D., Baudouin, J. Y., Poncet, F., Damon, F., Durand, K., Schaal, B., et al. (2021). Odor-driven face-like categorization in the human infant brain. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2014979118. doi: 10.1073/pnas.2014979118

Rekow, D., Leleu, A., Poncet, F., Damon, F., Rossion, B., Durand, K., et al. (2020). Categorization of objects and faces in the infant brain and its sensitivity to maternal odor: further evidence for the role of intersensory congruency in perceptual development. *Cogn. Dev.* 55:100930. doi: 10.1016/j.cogdev.2020.100930

Retter, T. L., Jiang, F., Webster, M. A., and Rossion, B. (2020). All-or-none face categorization in the human brain. *NeuroImage* 213:116685. doi: 10.1016/j.neuroimage.2020.116685

Retter, T. L., Rossion, B., and Schiltz, C. (2021). Harmonic amplitude summation for frequency-tagging analysis. *J. Cogn. Neurosci.* 33, 2372–2393. doi: 10.1162/jocn_a_01763

Ropeter, A., and Pauen, S. (2013). Relating 7-month-olds visuo-spatial working memory to other basic mental skills assessed with two different versions of the habituation–dishabituation paradigm. *Infancy* 18, 578–603. doi: 10.1111/j.1532-7078.2012.00133.x

Rossion, B., Retter, T. L., and Liu-Shuang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *Eur. J. Neurosci.* 52, 4283–4344. doi: 10.1111/ejn.14865

Rossion, B., Torfs, K., Jacques, C., and Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *J. Vis.* 15:15.1.18. doi: 10.1167/15.1.18

Sokol, S. (1978). Measurement of infant visual acuity from pattern reversal evoked potentials. *Vis. Res.* 18, 33–39. doi: 10.1016/0042-6989(78)90074-3

Stacchi, L., Liu-Shuang, J., Ramon, M., and Caldara, R. (2019). Reliability of individual differences in neural face identity discrimination. *NeuroImage* 189, 468–475. doi: 10.1016/j.neuroimage.2019.01.023

Stothart, G., Quadflieg, S., and Milton, A. (2017). A fast and implicit measure of semantic categorisation using steady state visual evoked potentials. *Neuropsychologia* 102, 11–18. doi: 10.1016/j.neuropsychologia.2017.05.025

Toffanin, P., De Jong, R., Johnson, A., and Martens, S. (2009). Using frequency tagging to quantify attentional deployment in a visual divided attention task. *Int. J. Psychophysiol.* 72, 289–298. doi: 10.1016/j.ijpsycho.2009.01.006

Tononi, G., Srinivasan, R., Russell, D. P., and Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged Neuromagnetic responses. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3198–3203. doi: 10.1073/pnas.95.6.3198

Walter, W. G., Dovey, V. J., and Shipton, H. (1946). Analysis of the electrical response of the human cortex to photic stimulation. *Nature* 158, 540–541. doi: 10.1038/158540a0

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fpsyg.2022.959534

OPEN ACCESS

EDITED BY
Aliette Lochy,
University of Luxembourg,
Luxembourg

REVIEWED BY
Marie-Pier Godin,
Université du Québec à Montréal,
Canada
Verónica Martínez,
University of Oviedo,
Spain

*CORRESPONDENCE
Aline Frey
aline.frey@univ-amu.fr

SPECIALTY SECTION
This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 01 June 2022
ACCEPTED 17 November 2022
PUBLISHED 08 December 2022

CITATION
Frey A, Lessard A, Carchon I, Provasi J and Pulido L (2022) Rhythmic training, literacy, and graphomotor skills in kindergarteners. *Front. Psychol.* 13:959534.
doi: 10.3389/fpsyg.2022.959534

Rhythmic training, literacy, and graphomotor skills in kindergarteners

Aline Frey[1]*, Andrée Lessard[2,3,4,5], Isabelle Carchon[6], Joëlle Provasi[6] and Loïc Pulido[7,8]

[1]Laboratoire de Neurosciences Cognitives, UMR 7291, CNRS – INSPE de l'Université d'Aix-Marseille, Marseille, France, [2]Département des sciences de l'éducation, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, [3]Centre de recherche interuniversitaire sur la formation et la profession enseignante (CRIFPE), Montréal, QC, Canada, [4]Consortium régional de recherche en éducation (CRRE), Saguenay, QC, Canada, [5]Observatoire interdisciplinaire de création et de recherche en musique, Université Laval (OICRM-ULaval), Québec, QC, Canada, [6]Laboratoire CHArt, Cognitions Humaine et ARTificielle, EPHE - PSL, École Pratique des Hautes Études - Paris Sciences Lettres, Campus Condorcet, Aubervilliers, France, [7]Consortium Régional de Recherche en Éducation (CRRE) et département des sciences de l'éducation, Université du Québec à Chicoutimi, Saguenay, QC, Canada, [8]Centre de Recherche et d'Intervention sur la Réussite Scolaire (CRIRES), Université Laval, Québec, QC, Canada

Introduction: The aim of this manuscript is twofold: first, to investigate the relationship between rhythmic, phonological and graphomotor skills in kindergarten children; and second, to evaluate the possible impact of rhythmic training on the two other skills.

Methods: To that end, we selected a sample of 78 children in Québec. Forty-two children received rhythmic training (experimental group) and 34 arts training (active control group) during the same period (10 weeks). Before and after training, children in both groups were assessed for general skills (forward and backward memory span, vocabulary, non-verbal ability), rhythmic skills (synchronization and discrimination tasks), literacy skills (phonological skills - syllable counting, syllable deletion, rhyme discrimination – and invented spelling skills) and graphomotor skills (legibility of letter writing, quality of copying of geometric shapes).

Results: Results showed correlations between the child's rhythmic and literacy skills, as well as between rhythm synchronization and pen pressure. In addition, rhythmic training showed improvement in rhythmic abilities, but this did not transfer to literacy or graphomotor development (apart from a significant increase in the duration of pauses in both groups at post-test, with a larger improvement for the rhythm group).

Discussion: These results are discussed in terms of duration and intensity of learning, and they highlight the possible benefits of informal rhythm practices in the classroom.

KEYWORDS

rhythm, literacy, spontaneous motor tempo, graphomotor abilities, kindergarten

Introduction

Rhythm, and especially the ability to be synchronized with one's environment, is crucial for the sensory-motor, cognitive, emotional, and social development of young children

(Feldman, 2007). Thus, children must be able to perceive the rhythms around them, produce rhythmic activities, and modify the rhythm of their activities in order to be in synchrony with the rhythm of their environment. The dynamic systems theory model (Jones, 1976; Jones and Boltz, 1989; Drake et al., 2000) explains how the subject's selective attention works. Rhythmic sequences are broken down into different hierarchical levels of temporal regularity in relation to a preferred period of processing which is specific to each individual and acts as their reference period. The subject's selective attention will spontaneously focus on events that occur during a period close to their reference period. On average, this reference period is 600 ms long in adults (Drake and Botte, 1993) and 400 ms long for children aged between 2 and 14 years (Provasi and Bobin-Bègue, 2003; Provasi et al., 2014). For children, this duration is accompanied by a time window of plus or minus 20% of its speed, meaning that they start to differentiate between two tempi once one is 20% faster or slower than the other. This 20% difference serves as the differentiation threshold between two tempi. It is thus more difficult for children to synchronize to a tempo of 600 ms +20% than to a tempo of 400 ms +20% (Provasi and Bobin-Bègue, 2003; Provasi et al., 2014). Similarly, in production, it is easier for children to synchronize to a 400 ms tempo than to a 600 ms tempo.

Furthermore, the ability to perceive and produce rhythmic patterns promotes fundamental human capacities that go beyond the strict rhythmic or musical field, including the coordination of action, and cognitive and language processing, specifically oral language skills (Tierney and Kraus, 2013). Thus, prosodic-level features such as speech rhythm and stress facilitate syllabic segmentation of the acoustic signal (Choi et al., 2005). Developmentally, speech rhythm is one of the first cues used by infants to segment the speech stream into words and word parts (Ramus et al., 2000) and parents naturally use emphatic stress and exaggerated rhythmic patterns when interacting with their infants.

Along those lines, a growing number of studies have examined the potential importance of rhythm perception and/or production in older children, between 2 and 5 years of age, and its relationship to reading skills. Working with American kindergarteners, Moritz et al. (2013) showed that performance in rhythm pattern copying marginally correlated to the ability to isolate initial phonemes, segment syllables, delete word and syllable parts from compound and multi-syllabic words, and segment sentences into words. In a beat synchronization task, Carr et al. (2014) demonstrated that American children between the ages of 3 and 4 who were unable to synchronize with an external auditory beat had poorer pre-literacy skills, specifically in phonological processing. More recently, Lundetræ and Thomson (2018) studied to what extent Norwegian first graders' production of rhythm predicts reading and spelling difficulties. Significant group differences in emergent literacy skills and rhythmic tapping were found between children below and above the national threshold in word reading and spelling. More specifically, they were interested in whether students' performance at the rhythm task at school entry served as a predictor of poor in word-reading and spelling abilities at the end of grade one. They found significant group differences in children's ability to tap in time to an externally delivered beat, measured at school entry, when groups were defined upon whether children went on to score above or below the 20th percentile threshold in national assessment tests at the end of grade one. Taken together, these results demonstrate that rhythm production, perception and/or synchronization, and language skills are potentially interrelated and suggest the importance of temporal sensitivity for language learning in preschoolers. More specifically, it appears that rhythmic skills support speed naming of letters and phonological awareness (David et al., 2007), which are considered to be robust predictors of reading ability (Wolf and Bowers, 1999). Phonological awareness (PA) corresponds to the ability to manipulate speech sounds, phonemes, and syllables, and facilitates the decoding and the analysis of words during spelling (Bus and Van Ijzendoorn, 1999). This ability involves auditory processes for fine analysis of the temporal structure of sound patterns, which provides, for example, statistical clues to word boundaries (Culter, 2012). Phoneme distinctions may involve very fine temporal cues, such as distinguishing a /b/from a/p/ (Greenberg et al., 2006). Thus, as speech relies on fine temporal and rhythmic patterns, a body of evidence shows that rhythm mechanisms support PA. Rhythm skills could help children break down language chains into sublexical units, such as syllables, thus facilitating awareness of the phonological units of language (Goswami, 2011; Moritz et al., 2013).

Although much research has focused on the links between rhythmic and language abilities, much less has been devoted to the study of potential relationships between these rhythmic skills and handwriting skills. And yet, similarly to language and other complex human actions, handwriting is not just a succession of isolated acts, but rather an organized process, in which the time and the space of each motor unit (i.e., strokes, letters, and words) are contextually interdependent on one another within a larger unit (Van Galen and Teulings, 1983). Specifically, handwriting respects two principles based on rhythmic organization: the phenomena of homothety and isochrony (Lashley, 1951). Briefly, homotheticity states that the ratio between the durations of the motor events composing a writing movement remains unchanged even if the way in which the word is written is different (i.e., bigger, smaller, faster, or slower), and thus despite changes in overall duration. Isochrony, on the other hand, refers to the proportional relationship between the speed of execution of the movement and the length of its trajectory. In other words, depending on the size of the written output, the duration of execution of a movement remains stable; thus, writing speed adjusts proportionally to the size of the output. The typical example is a signature, which is always written at the same speed regardless of its size. In a study conducted on some 300 primary school children, Pagliarini et al. (2017) showed that these two principles of the rhythmic organization of handwriting are already present in children's first handwriting productions, suggesting "that an internal representation of the rhythm of handwriting is available before the age in which handwriting is performed automatically" (p.1). These results suggest that, despite being a cultural acquisition, handwriting appears to be shaped by more general constraints on the temporal planning of movements and highlight the relationship with the rhythmic dimension that is present before handwriting movements become automatized (Evertz and Primus, 2017). Furthermore, Ben-Pazi et al. (2007) showed that rhythmic performance in a tapping task correlates with children's handwriting quality. In this context, Véron-Delor et al. (2017) tested the hypothesis that rhythm can serve as an external cue that guides movement, through the natural and spontaneous tendency to synchronize with it (Bangert et al., 2006; Zatorre et al., 2007; Schaefer, 2014). In a child with severe writing difficulties, they show that playing rhythmic music in the background during a writing task significantly improved the duration, speed, and fluency of movement. However, apart from these few studies, and although it is known that there is a dynamic interaction between perception and action, both clearly involved in handwriting, the rhythmic dimension of writing has largely been neglected.

All of the above results are based on correlation analyses indicating links between rhythmic and literacy skills (language and/or handwriting skills). However, as correlation is not causality (Schellenberg, 2004; Chobert et al., 2014), one way to determine whether rhythm competence could mediate literacy skills is to conduct longitudinal studies that develop rhythmic skills and assess whether this development is accompanied by the development of literacy skills. Typically, in these kinds of studies, one group of children is trained in rhythmic skills, while the other group does something else, and the effect of these different practices on language and/or writing is measured before and after the training. Many studies have shown that musical training (i.e., instrumental practice, singing, pitch training) significantly improves language skills (e.g., pre-attentive perception, Frey et al., 2019; fluency, Zuk et al., 2014; novel word-learning and semantic associations, Dittinger et al., 2017) in children, showing beneficial transferrable effects of musical training. Of course, musical activities are usually based on rhythm, but training children solely on rhythmic aspects is less common. In Verney (2013)found that training in rhythmic activities led to increased PA, especially rhyme and syllable awareness. More recently, and in order to better understand the respective involvement of rhythm and pitch in the beneficial effects of musical learning on PA, Patscheke et al. (2018) randomly assigned preschoolers aged between 4 and 6 years old to a rhythm, pitch, or sport (control group) training group. The first two groups were trained three times a week during 20-min sessions over a period of 16 weeks. Results showed that only the pitch program improved PA abilities, even if the rhythm group also showed an increase which did not reach significance (when compared to the sports control group). Furthermore, rhythmic skills (as well as pitch perception) were not measured, which makes it impossible to know if the training worked well, and to know what the lack of improvement in language is related to.

Finally, this research also has educational and pedagogical implications. Indeed, most school-based learning is based on oral and written language skills, which must therefore be acquired as early as possible in education. Despite this, various surveys (e.g., PISA, OECD, 2019; PIRLS, Martin et al., 2017) have reported for several years that some children have early and persistent difficulties in processing language information and more specifically in identifying words all over the words, and notably for French speakers. These difficulties continue throughout schooling, showing that on average, 20% of adolescents have relatively significant reading difficulties in France for example. Thus, despite advances in scientific knowledge about learning psychology and pedagogy, teachers are still quite helpless to remedy students' language difficulties (Frey and Sappey-Marinier, 2018). Faced with this observation, new alternatives appear indispensable, and an additional objective of this research is to propose a simple and playful approach for indirectly improving language and handwriting skills through rhythmic training.

Thus, the aim of this manuscript is twofold: first, to investigate the relationship between rhythmic, phonological, and graphomotor skills in 6-year-old children, and second, to evaluate the possible impact of rhythmic training on the other two skills. We hope to confirm correlations between rhythm and phonological skills and to observe understudied relationships between rhythm and handwriting skills. To achieve these objectives, we conducted a study on 78 children in Quebec. Rhythm skills were evaluated in both discrimination and synchronization tasks (Provasi et al., 2014). Phonological skills were classically assessed through syllable counting and deletion, rhyme discrimination, and initial phoneme identification tasks. We also set up an invented spelling task (Pulido and Morin, 2018), which is a way for children to explore the written code that involves variety of early literacy skills, such as alphabetic knowledge and phonological awareness in a more ecological way (Ouellette and Sénéchal, 2008). Finally, we examined the graphomotor component of handwriting (initiation and implementation of motor programs and neuromuscular execution) by evaluating the legibility of letter writing and their graphomotor characteristics, as well as the quality of the copying of geometrical shapes, on a digital tablet. These different skills were evaluated in pre- and post-tests, before and after 10 weeks of training, either in rhythm or in arts (active control group). We hypothesized that children trained in rhythm would not only improve their rhythmic skills, but also, through a transfer of learning effect, improve their literacy skills (phonological awareness and writing).

Materials and methods

Participants

Seventy-eight kindergartners from six classes across two schools in a middle town in Quebec, were involved in the experiment. Children whose teachers reported a learning disability were excluded from the analyses (2). Three classes were randomly assigned to the rhythm training ($n = 42$; 22 girls, mean age = 5 years 11 months, $SD = 113$ days), and the other three followed the arts training (*n* = 34; 17 girls, mean age = 5 years 10 months, *SD* = 129 days).

In Quebec, from the first year to 4 or 5 years of age, children mainly attend daycare centers where they are guided through playful activities to support their overall development (Ministère de la Famille, 2019). At age five, children can enter kindergarten, where they are offered informal or play-based activities. It aims to foster emerging literacy skills and children are expected to have phonological awareness skills and to be able to invent spelling by the end of kindergarten. None of them could read or write during the study.

Measures

Control measures

In order to ensure the homogeneity of the two groups prior to trainings on skills that could influence the development of the skills we have focused on in this research (specifically phonological awareness and invented spelling), we carried out three control tasks during pre-testing.

To control for the vocabulary ability of children, we used the Expressive vocabulary task from CELF-CDN-F (Wiig et al., 2009). The task consists of naming 27 pictures for a total score of 54 (2 points for each spontaneous response and 1 point for a response given after a cue). The reported interitem reliability of this task is 0.90.

Raven's Standard color Progressive Matrices (Raven et al., 1998) were used to control for the non-verbal ability of children. In this task, children are shown a picture with a partial mask, and have to decide which of six pictures completes the pattern. There are 36 items, and each correct is scored 1 point. The reported interitem reliability of this task is 0.76.

Finally, to control for working memory ability, we used both the forward and backward digit span from the Wechsler Intelligence scale for children (WISC IV, Wechsler, 2005), in which children have to repeat, either in the original order or backward, series of orally presented numbers (8 maximum in the forward test, 7 maximum in the backward test, 1 point for a correct response, total score out of 15).

Main measures

Rhythm skills

Children were given both discrimination and synchronization tasks (Provasi et al., 2014).

The two first discrimination tasks differed in the inter-stimulus interval (ISI) used in the rhythmic sequence of the reference period, that is 400 ms and 600 ms (400 ms is close to the reference period for children, and 600 ms is the reference period for adults). Children heard a first sequence of 15 beeps, followed by a second sequence of 15 more beeps 1,500 ms later. The children had to say orally whether the two sequences heard were the same or not.

Each pair of sequences contained a target sequence (with an ISI of 400 ms or 600 ms respectively), and the second sequence presented with the same ISI as the target sequence (i.e., 400 or 600 ms) for 4 trials, or 10, 20, 30 and 40% faster than the target sequence for 8 trials (2 of each). The position of the target sequence in each pair, the order of the 12 trials, and the order of presentation of these 2 discrimination tasks (400 or 600 ms) were counterbalanced between the children. Since children's reference period is close to 400 ms, it is more difficult for them to discriminate two tempi around 600 ms than two tempi around 400 ms. As the threshold of discrimination of tempi is situated around 20%, the most difficult discrimination task is therefore that at 600 ms −20%, and the score from this task was considered the relevant discrimination score.

The motor rhythm task was divided into 3 successive phases: a synchronized tapping phase with an auditory tempo and two spontaneous motor tempo (SMT) phases, one just before and the other just after the synchronization phase. During the two phases of SMT, the children had to type on the spacebar of a computer keyboard, as regularly as possible, for a total of 30 successive strikes. During 11 trials of the synchronization phase, the children had to press the spacebar on a computer keyboard at the same time as a beep. Eleven trials differed in the duration of the ISI, either equal to their SMT (0%), or 10, 20, 30, or 40% faster or slower than their SMT, and two trials had an ISI of 400 and of 600 ms. The order of the 11 ISI trials was balanced between subjects. Synchronization to 400 ms ISI and 600 ms ISI were always the two last trials of synchronization phase. For these tasks, as in Provasi et al. (2014), we calculated the mean and standard deviation of Inter-Tap Interval (ITI) for the pre-, the post-synchronization phases, and for the synchronization phase. Because the ISI of the synchronization phase was dependent on the child's SMT, the data were normalized by dividing the mean ITI by the mean ITI obtained during the pre-synchronization SMT. In addition, we calculated an adjusted value of the Rayleigh score as explained in Appendix 1 by Provasi et al. (2014). A higher adjusted Rayleigh score indicates a larger variability in rhythmic movement. The lower the value of the corrected Rayleigh, the more the taps are adjusted to the rhythmic stimulation. Since the children's SMT is close to 400 ms, it is more difficult for them to synchronize at a 600 ms ISI than at a 400 ms ISI. A Rayleigh score corrected to 600 ms is thus the most relevant measure of synchronization when examining a rhythmic learning effect.

Literacy skills

Literacy skills of the participants were assessed through phonological awareness and invented spelling tasks. Four tasks of the phonological awareness subtest of CELF-CDN-F (Wiig et al., 2009) were used, consisting, respectively, in (i) counting the number of syllables in six words (syllable counting/6); (ii) repeating a word after removing a target syllable (syllable deletion/8); (iii) saying if two target words rhyme (rhyme discrimination/8) (iv); listening to words and identifying their first phoneme (phoneme identification/8). We chose these four tasks

TABLE 1 Criterium used to assess the legibility of the participant's productions in graphomotor tasks.

Task	Criteria
Copy of geometrical shapes	*Square*: 4 secant segments of equal length, opposite segments parallel; tolerance of a 2 mm opening in the shape
	Circle: closed shape with no angles
	Triangle: 3 secant segments. Tolerance of a 2 mm opening
Cycloids	*Clockwise cycloid*: three buckles made clockwise
	Counterclockwise: three buckles made counterclockwise
Letters and first name	For each letter, criteria of the Evaluation Tool of Children's Handwriting (ETCH, Amundson, 1995)

among the 11 proposed as Pulido and Morin (2018) did because they are close to the phonological mechanisms involved in word writing. The reported interitem reliability of this test is.96.

In the invented spelling task (Pulido and Morin, 2018), children were told to do their best to try to write four bisyllabic words with a simple spelling (in French: "ami," "piano," "citron," "café"), with the letters they know. For each sound written with an appropriate grapheme, they earned one point, for a total maximum score of 17.

Graphomotor skills

To assess graphomotor skills, children were instructed to perform the following four tasks:

- Copy three geometrical shapes: a square, a circle, and a triangle.
- Copy three cycloids clockwise and three cycloids counterclockwise.
- Copy six letters presented in lower script: a, m, u, l, p, and e. These letters were chosen because of their variety in terms of forms.

For this copy task, a still model was presented to the children which disappeared when the pen touched the screen.

- Write of their first name.[1]

For each production, a legibility score was given following the agreement of two trained coders: 1 for a legible production and 0 for an illegible production, following the criterion presented in Table 1.

These tasks were realized by writing or drawing on a digital tablet (Cintiq Creative Pen Display) connected to a Mac mini using Boot Camp. The Eye and Pen software (Alamargot et al., 2006) was used to present stimuli when necessary, and to record the graphomotor activity.

We thus collected different information about the kinematics of handwriting: (i) mean speed of the pen during production (in cm per second); (ii) mean pressure; (iii) number of pauses during production (threshold for the pause was set to 20 ms, considering the frequency of the digital tablet, *cf.* Alamargot and Morin (2015) for detailed explanations); and (iv) mean duration of pauses (in ms). For these four variables, we calculated an average value which concerns the drawing tasks (geometric and cycloid shapes) and the writing tasks (writing of isolated letters and writing of the first name).

Procedure

Each child was tested one time before the trainings and one time after. The tasks were performed face to face with one experimenter, in a quiet room, at school. The order of the tasks was: forward and backward digit, phonological awareness, Raven's progressive matrices, expressive vocabulary (these four tasks were performed only in pre-testing), and rhythm and writing tasks. The whole experiment lasted about 30 min.

Trainings

Trainings were delivered to the children by their respective teachers, 5 min per day, 5 days per week, over a 10-week period, starting in February. This format was chosen to fit easily into students' daily routines so as to minimize disruption and to encourage children's involvement in the proposed activities (Harper and O'Brien, 2015). Although it is possible to observe improvements in cognitive abilities after music training in some studies, the frequency and duration of this training did not always influence the results among these studies (Sala and Gobet, 2020). However, in reading and writing, short, daily trainings are considered more effective than long, spaced-out interventions (Chapleau et al., 2020). Also, the effects of musical practice (neuroanatomical or functional changes) are observed when the practice is done a minimum of three times a week (Paquet, 2017).

Before the start of training, teachers received a half-day of instruction in which the interventions were explained to them, the tasks were demonstrated, and they practiced each task until they were completed correctly. During the 10-week training period, supervising researchers contacted all teachers once a week to ensure that there were no issues and to answer any questions. It was also ensured that both trainings involved the same level of child engagement. Finally, teachers were required to complete an intervention logbook, which ensured that the interventions were carried out accurately and involved the same level of child engagement in each condition.

Rhythm training

A typical rhythmic training session consisted of four different tasks (not necessarily in a fixed order), for a total of 5 min each day. The first task, "rhythmic vitamins," consisted of repeating short rhythmic sequences using the fingers, hands, thighs, and feet (duration: 1 min). To do this, one-minute videos of pre-recorded sequences were projected to a screen in the classroom, visible to all children.

---

[1] All these unpublished copying tasks were developed in the framework of the ANR-FRQSC NUMEC project (leaders: M.-F. Morin and D. Alamargot).

These pre-recorded sequences were produced by a researcher in the team, who holds a level II certification in the Orff Schulwerk approach (Orff Canada), using a progression adapted to the rhythmic abilities of young children. Teachers had 60 video clips that increased in difficulty over time. The video to be shown each day was pre-planned, for example "Week 1, Day 4," but teachers could choose to redo the videos already used if they were not mastered by the whole class or most students. For example, one of the "easy" rhythmic patterns proposed at the beginning of the training was composed of 5 pulses (long, long, short, short, and long) and used only one body part, while a more complex pattern could be composed of 9 faster sounds with two alternating body parts (e.g., fingers and thighs), and could be proposed when children became more comfortable over the weeks.

The second task, "rhythmic vitamins on tape," also involved repeating short rhythmic sequences on different body parts, but this time synchronizing with metric instrumental music. A series of 60 one-minute video clips of increasing difficulty were also provided to the teachers, and projected to the children each morning, using different musical styles (e.g., jazz, classical, and world music), different time signatures (e.g., 3/4, 2/4, 4/4, and 6/8), and different tempi (slow, medium, and fast). The addition of soundtracks allowed the children to synchronize to an external tempo while reproducing rhythms.

The third task, also one-minute in duration, involved moving in rhythm in the classroom while listening to a piece of music. Different instrumental soundtracks of various tempos, atmospheres, and styles were proposed for this musical synchronization activity.

Finally, the fourth task, which lasted 2 min, consisted of a game of moving to the beat. The task could either involve walking to the tempo of instrumental music for 2 min (with different instructions, such as imitating an animal), or inventing a movement to accompany the rhythm of the music. The task could also consist of following the tempo played by the teacher on the tambourine, who played regular fast, medium, or slow beats, without a musical soundtrack. These activities allowed for students to synchronize their body movements to a given metric. Materials for this rhythm training included an activity description guide, videos, and audio tapes.

Art training

For the duration of the project, the teachers had 60 digitized visual artworks (such as paintings and sculptures) according to a different theme each week (e.g., monochrome, portraits, landscapes, book illustrations, animals, graffiti, etc.). A typical art training session included five tasks for a total of 5 min each day. First, the teacher showed a reproduction of a visual artwork to children. Then, the first task was to produce, altogether, a human copy of the artwork in question. To that end, each child had to point the part of the work they wanted to reproduce, then position themself so as to represent this element in space. Each child took their turn to collectively reproduce the entire work (e.g., one child could make the beak of the eagle, the second its wings).

The second task was a static mime. The teacher asked the children, placed face to face, to choose a character or element of the artwork. They had to take its shape with their body and keep this pose for a few seconds without moving. The third task was to keep the posture chosen in the previous task and act out an emotion they felt while observing the item they chose, or an emotion they thought the object or character might feel (e.g., anger, joy, and sadness), without moving.

For the fourth task, they were asked to move through the space while keeping the emotion and character (or object) they had chosen.

Finally, the fifth task was an "imaginary object" game. The teacher mimed an object related to the artwork, which the children had to guess by miming different actions related to this object (e.g., pretending to shovel snow, make a snowball or a snowman, the object to be guessed being "snow"). Then, the children were placed in a circle, and the teacher passed the "object" to a child, who, in turn, would "mime" a movement with the object (e.g., making an angel in the snow), and then pass it to the child next to them.

If time allowed, a sixth task was proposed, where the children worked in pairs. One of them had to "draw" an excerpt from the artwork on the back of the other (with their finger). The peer then had to guess which part it was and point it out on the interactive whiteboard.

Materials for this art training included a guide to describe activities, 60 visual artworks with their description, and many suggestions of elements or emotions to exploit for each work, which did not limit the teacher and/or children in their choices. In the end, the visual arts condition was made up of activities of the same duration as the rhythm routines, so as to maintain equivalent lengths of intervention.

Data analysis

To better understand the relationships between rhythmic, graphomotor, and literacy skills prior to the training, we first pooled the pretest data independently of the training group (since the trainings had not yet begun at the time of the pretests), and performed Pearson's correlations analysis between all the measures, taken two by two (bivariate correlations, with appropriate Bonferroni correction, Curtin and Schulz, 1998). Next, to ensure the homogeneity of the rhythm and art groups before trainings, Student's *t*-tests were performed between the two groups (rhythm vs. art) on the three control measures (Vocabulary, Progressive matrices, and Span tasks). Finally, to evaluate training effects, we carried out ANOVA with moment (pre vs. post-test) as the within-factor, and groups (rhythm vs. art) as the between-factors. When they were significant, interactions were decomposed with Tukey HSD *post-hoc*, and when relevant, effect sizes were also reported: Cohen's d (Cohen, 1988) for *t*-tests and eta-squared (η^2) for ANOVA. These analyses were realized using the JASP Team (2022) software.

TABLE 2 Mean (*SD*) of the 3 control measures (vocabulary, progressive matrices, and span tasks) at pre-test, for the 2 groups (rhythm and art).

Control tasks	Rhythm training pre-test	Art training pre-test	Student's *t*-test	*p*-value
Vocabulary/54	26.88 (5.05)	26.65 (2.06)	$t(74) = -0.24$	$p = 0.81$
Progressive matrices/36	18.62 (4.29)	19.09 (4.09)	$t(74) = 0.47$	$p = 0.64$
Span tasks/15	6.91 (1.55)	6.59 (1.78)	$t(74) = 0.19$	$p = 0.85$

TABLE 3 Mean (*SD*) of the rhythm measures at pre and post-test and results from the ANOVA (*F*, *p* values and size effect) for the interaction between pre versus post-test moment and groups (rhythm vs. art).

Rhythm	Rhythm training pre-test	Rhythm training post-test	Art training pre-test	Art training post-test	*F*	*p*-value	η^2
Discrimination score	1.48 (0.63)	1.38 (0.80)	1.26 (0.74)	1.18 (0.72)	$F_{(1,74)} = 0.07$	0.93	0.00004
Synchronization score	475.86 (86.70)	428.36 (90.85)	445.07 (92.15)	467.77 (94.08)	$F_{(1,74)} = 6.7$	0.012	0.036

Results

Correlations at pre-test

To investigate the relationships between rhythmic, literacy, and graphomotor skills, we first set up a correlation analysis between all the measures collected pre-test, regardless of the group (rhythm and art), for all our participants. The analyses show a significant correlation between the rhythm discrimination score and both syllable counting ($r = 0.240$, $p = 0.036$) and phoneme identification performances ($r = 0.367$, $p = 0.001$). Thus, the better the child performs on the rhythm discrimination task, the better they will perform on the syllable counting task and the phoneme identification task.

Moreover, the rhythm synchronization score selected correlates negatively with the initial phoneme identification score ($r = -0.281$, $p = 0.014$). As a reminder, a low synchronization score (Rayleigh corrected to 600 ms) reflects good rhythm synchronization performances. Thus, the more synchronized the children were to the 600 ms auditory tempo, the higher their performance in initial phoneme identification.

Finally, the rhythm synchronization score correlates negatively with the pressure exerted on the pen during writing ($r = -0.355$, $p = 0.002$) and drawing tasks ($r = -0.303$, $p = 0.008$). The more synchronized the children were to the 600 ms auditory tempo, the more pressure they exerted on the pen during writing and drawing tasks.

Homogeneity of the rhythm and art groups before trainings

Results of Student's *t*-tests (Table 2) showed no significant differences between the two groups on the three control measures (Vocabulary, $t(74) = -0.24$, ns, Progressive Matrices, $t(74) = 0.47$, ns, and Span tasks, $t(74) = 0.19$, ns), indicating that the 2 groups were well homogenized prior to trainings on these skills.

Training effects for rhythm measures

Before observing possible transfer effects from rhythmic learning to other non-rhythmic skills, we verified that the rhythmic interventions influenced rhythmic skills. Descriptive statistics (mean and standard deviation) for the two rhythm measures are presented in Table 3, for the pre-test and the post-test, in function of rhythm training and art training.

Neither the rhythmic nor the visual art training improved the discrimination score, but synchronization score was significantly improved in the rhythm group (*cf.* Figure 1 below, *post-hoc* comparison: in the rhythm group: pre-test = 475.86; post-test = 428.36; $p = 0.05$).

Training effects for literacy measures

Descriptive statistics (mean and standard deviation) of the five literacy measures are presented in Table 4 for the pre-test and the post-test in function of rhythm training and art training.

Neither the rhythmic nor the art training improved the 5 measures of literacy.

Training effects for graphomotricity measures

Descriptive statistics (mean and standard deviation) of the five graphomotricity measures are presented in Table 5, for the pre-test and the post-test in function of rhythm training and art training.

The results were significant only for the duration of pauses in drawing, with a significantly greater duration of pauses in the rhythm group (*cf.* Figure 2 below, *post-hoc* comparison: in the rhythm group: pre-test = 27.18, post-test = 73.61 $p = 0.008$) as compared to the art group.

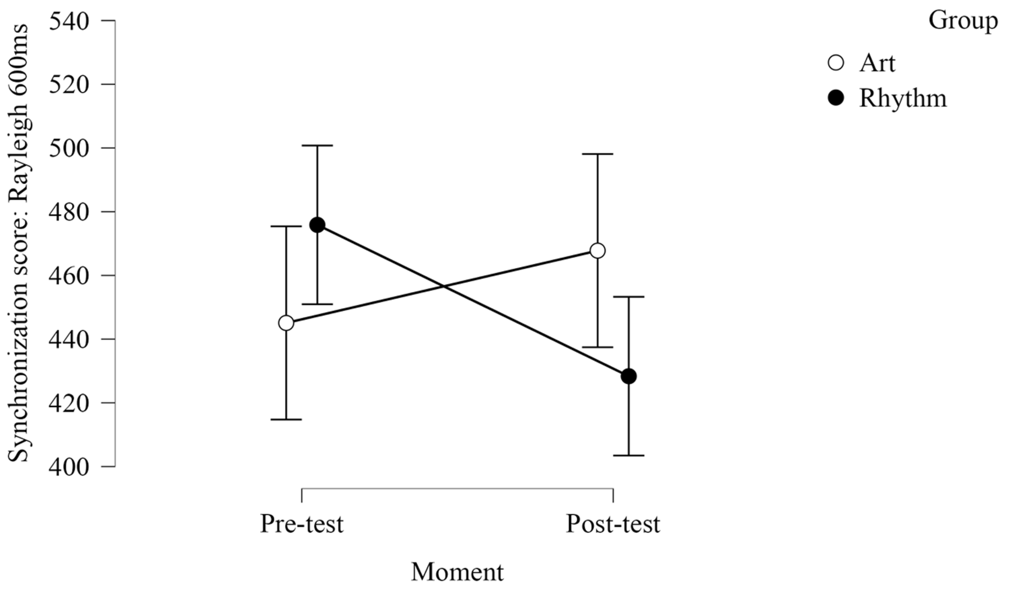


FIGURE 1
Training effect (difference between pre-test and post-test) for rhythm synchronization score (Rayleigh corrected to 600ms) for art and rhythm groups.

TABLE 4 Mean (*SD*) of the literacy measures at pre and post-test, and results from the ANOVA (*F*, *p* values and size effect) for the interaction between pre- versus post-test moment and groups (rhythm vs. art).

Literacy	Rhythm training pre-test	Rhythm training post-test	Art training pre-test	Art training post-test	*F*	*p*-value	η^2
Syllable counting/6	5.70 (0.59)	5.73 (0.54)	5.52 (0.65)	5.64 (0.69)	$F_{(1,74)} = 0.2$	0.61	0.001
Syllable deletion/8	2.31 (1.22)	2.56 (1.51)	2.21 (1.59)	2.12 (1.27)	$F_{(1,74)} = 0.9$	0.34	0.004
Rime discrimination/8	4.91 (1.39)	5.56 (0.96)	5.03 (1.27)	5.56 (0.83)	$F_{(1,74)} = 0.1$	0.70	0.0007
Phoneme identification/8	5.02 (2.43)	6.75 (1.57)	4.18 (2.79)	6.72 (1.76)	$F_{(1,74)} = 2.5$	0.12	0.007
Invented spelling/17	6.53 (3.01)	8.82 (3.75)	5.23 (6.68)	8.63 (3.29)	$F_{(1,74)} = 1.8$	0.18	0.009

TABLE 5 Mean (*SD*) of the graphomotricity measures at pre- and post-test, and results from the ANOVA (*F*, *p* values and size effect) for the interaction between pre- versus post-test moment and groups (rhythm vs. art).

Grapho motricity	Rhythm training pre-test	Rhythm training post-test	Art training pre-test	Art training post-test	*F*	*p*-value	η^2
Legibility of drawing (/6)	2.02 (1.20)	1.33 (0.57)	1.97 (1.11)	1.24 (0.43)	$F_{(1,74)} = 2.35$	0.13	0.011
Speed in drawing (in cm/s)	5.37 (1.71)	5.66 (3.14)	5.77 (1.37)	5.66 (2.77)	$F_{(1,74)} = 3.06$	0.08	0.017
Pressure in drawing	18,352 (4033)	11,697 (4197)	18,311 (3839)	12,624 (4022)	$F_{(1,74)} = 1.28$	0.26	0.002
Nb. of pauses in drawing	19.15 (12.71)	5.49 (2.72)	17.4 (10.03)	6.05 (3.86)	$F_{(1,74)} = 2.12$	0.15	0.012
Duration of pauses in drawing (in ms)	26.88 (4.61)	118.65 (78.44)	26.22 (4.11)	111 (51.84)	$F_{(1,74)} = 4.05$	0.048	0.019

Discussion

The purpose of this research was to investigate possible links between three skill areas: rhythm, literacy, and graphomotricity. To that end, children were evaluated on several tasks measuring those skills, and we looked for correlations between them. We then engaged one-half of students in daily rhythm training, and the other in arts training. We hypothesized that the rhythmic training would improve rhythmic skills, which would allow a transfer effect to phonological and graphomotor skills.

As expected, results showed positive correlations between rhythm discrimination measures and PA, specifically syllable counting and identification of the first phoneme tasks, and between synchronization measures (Rayleigh 600 ms – 20%) and identification of first phoneme. Our results confirm that kindergartners' rhythmic abilities were strongly related to their

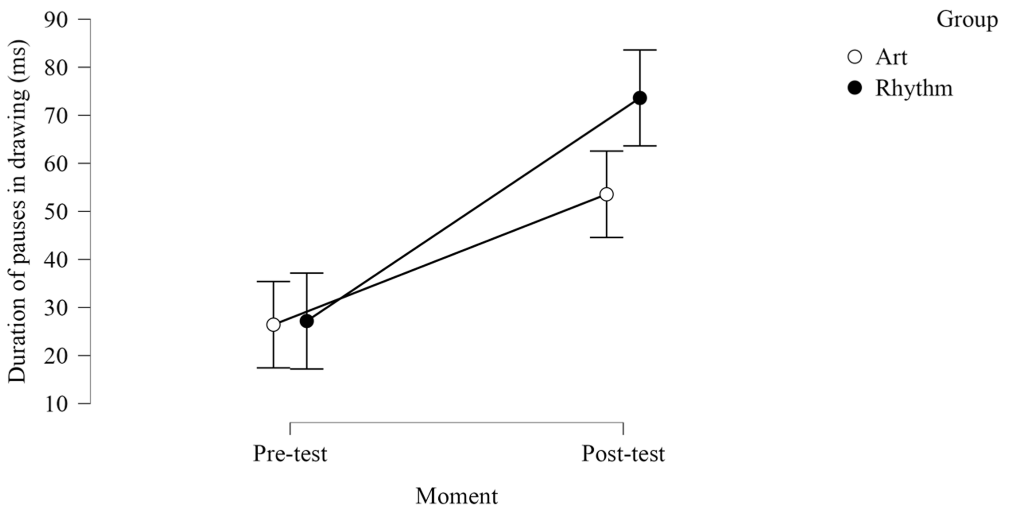


FIGURE 2
Training effect (difference between pre-test and post-test) for duration of pauses in drawing for rhythm and art training.

PA (Moritz et al., 2013). Syllables are a very important element in language segmentation, specifically in French, which is a syllable-timed language (one characterized by the recurrence of a given element at regular intervals vs. stress-timed languages, in which the vowel is accented; Wenk and Wioland, 1982). Good rhythmic skills seem to underlie the ability to break language into syllables, and as previously noted, when children are learning to read, those who can easily segment and blend sounds are able to use this knowledge to read and spell. In contrast, our results did not show any correlation with the syllable deletion task. The scores for this task are very low and this can be explained by the fact that this task also involves working memory capacities, since it consists, in addition to cutting the word into syllables, of maintaining it in memory and removing the syllable “to be deleted” in order to reconstruct the new item by merging the remaining syllables. Some studies have shown that working memory accounts for the variance in some PA subtasks, and in our case, it would appear that 5-6-year-olds are not yet able to perform the task correctly (Oakhill and Kyle, 2000). Moreover, the fact that the rhythm discrimination task is correlated with two of the PA tasks, while the synchronization task is correlated with only one, is consistent with the results indicating that the rhythm perception task could be more related to phonological skills than motor production rhythm skills (Zhang et al., 2017). In line with recent results (Bonacina et al., 2020), showing that the ability to drum to a beat predicts rapid letter-naming abilities, whereas the ability to tap rhythmic patterns predicts PA, our results also highlight that rhythm is a multi-dimensional skill, and that more studies are needed to better specify which rhythmic component (perception, reproduction, synchronization, etc.) underlies which language component (phonological processing, letter processing, etc.), and how this might change with age.

The results of these correlation analyses also show a negative correlation between the 600 ms synchronization score and pen pressure during the writing and drawing tasks. Thus, the more synchronized the children were, the more pressure they exerted on the pen. As indicated in the introduction, some correlational, behavioral, and neuroimaging studies have already suggested relationships between motor skills and rhythmic abilities (Grahn and Brett, 2007; Monier and Droit-Volet, 2019). For example, Monier and Droit-Volet (2019) assessed 57 children aged 5 to 8 years using sensorimotor synchronization and continuation tasks. For both tasks, the stability of production and the ability to produce a new tempo (different from the preferred tempo) were correlated with fine motor skills. However, not many studies have examined at pen pressure, and thus the interpretation of our result is complex. On the one hand, we could have hypothesized a correlation in the opposite direction of the one observed, that is that better rhythmic skills would lead to a decrease in pen pressure. This would build on, for example, the work of Degé et al. (2020), in which 25 preschoolers were randomly assigned to music training (including rhythmic exercises, dance, and familiarization with pitch intervals) and sports training for 20 min, 3 times per week, for 14 weeks. Children in the music group showed better motor inhibition, suggesting that music/rhythmic training may contribute to the development of motor control. On the other hand, we know that at this age, children do not have fully mature graphomotor gestures (Vaivre-Douret et al., 2021) and it is possible that those who perform better on rhythmic aspects hold their pen better, and would thus present a more fluid and more constant tracing, with perhaps more pressure to exert better control on the pen. However, this should also result in a better legibility of the writing, which is not the case in our results.

Our results also show that after only 10 weeks of rhythm training, for 5 min a day each morning, children in the rhythm group improved significantly better on the synchronization task (Rayleigh 600 ms), which is not the case in the arts group. Even if the effect size is small, to our knowledge, this is the first time that improvements of rhythm abilities have been observed following such a short training period, and for training carried out in the school by the respective teachers of each class as a daily routine (even if in reality, the duration of the trainings was 7–8 min rather than 5). Indeed, in the few studies on rhythmic training, it is sometimes surprising to find that rhythmic skills were not measured (neither before nor after learning), but only trained, which limits the scope of the results and the interpretation of possible transfer effects to other skills (e.g., Patscheke et al., 2018). Moreover, trainings are usually performed by external professionals, such as researchers, research assistants, or musicians. While this perhaps guarantees a better implementation of the training program, it is also more expensive and limits the development of this kind of practice. In our case, we requested that the teachers realize the trainings themselves. The fact that the trainings took place in real everyday school situations increases the ecological validity of the research (Laurencelle, 2005). This significant improvement may also be related to the fact that the trainings involved the whole body, and not only the motor response of the hand, as is often the case in other studies. This finding falls within the field of embodied cognition (Barsalou, 2008; Kiefer and Trumpp, 2012) and corroborates the growing research showing the importance of the use of the whole body in learning. In education, musical practice and rhythmic activities are considered "multimodal" because they require the processing of visual, auditory, and motor information (Dormoy, 2019).

Nevertheless, our results do not show any specific improvement in the rhythmic group for either phonological awareness or graphomotor skills, except for an increase in the duration of pauses while drawing in the rhythm group. This result is probably related to the improvement in the Rayleigh score corrected to 600 ms in this rhythm group. Indeed, it can be assumed that the improvement in the prediction process brought about by the rhythmic training will be reflected in the ability to inhibit a motor behavior. The increase in the duration of pauses can be interpreted as a more controlled and better anticipated behavior, allowing for better preparation of the graphomotor gesture (Frischen et al., 2022).

Therefore, with the exception of this last small result, our study does not allow us to form a conclusion about transfer effects between rhythmic skills and other non-rhythmic/general skills. Several explanations can be offered to explain this lack of effect. First, as explained above, most studies showing links between rhythmic skills and other skills use only correlational and/or mediation analysis. In interventional studies involving rhythm, the results are mixed: some of them demonstrated positive impact of rhythmic training on phonological processing and reading (e.g., Huss et al., 2011; Rautenberg, 2015; Ozernov-Palchik et al., 2018), but not all studies have found such relations (Anvari et al., 2002). Thus, Gordon et al. (2015) demonstrated that the association between PA and rhythm in 6-year-olds was no longer significant after non-verbal IQ was partialled out. Patscheke et al. (2018) specifically trained 4 to 6-year-olds on either pitch or rhythm, and their results show that after the training phase, only the pitch program showed a positive effect on PA. Overall, results of longitudinal studies are not so consistent, and probably depend on many factors, such as the duration and content of the training, the tests used, and the age of the children. On this last point, we had chosen kindergarteners, for whom it seems that rhythmic training interventions are the most effective (Hierbert and Taylor 2000; Gordon et al., 2015; Ozernov-Palchik et al., 2018). Moreover, in some of these studies, the interventions involved both non-speech rhythm tasks (e.g., playing a musical instrument, hand clapping) and speech rhythm tasks (e.g., segmenting a word into syllables by hand clapping; Lê et al., 2020), whereas in our study, we only trained motor rhythmic skills. This may also explain in part why we did not observe significant improvement in literacy skills especially due to the rhythm training. Finally, the relationship between rhythm and literacy development has been repeatedly observed in children with learning disabilities (Wolff, 2002; Thomson and Goswami, 2008; Corriveau and Goswami, 2009; Huss et al., 2011; Habib et al., 2016; Lundetræ and Thomson, 2018), but only a handful of studies have been conducted in typically developing children (David et al., 2007; Zhang et al., 2017; Ozernov-Palchik et al., 2018). It is thus well-known that children with developmental dyslexia show an impairment of temporal processing; for example, they show greater variability when asked to tap along a metronome (Thomson and Goswami, 2008) and difficulties reproducing patterned rhythms of tones. Many studies showed a strong link between rhythm skills and phonological and reading abilities in this population, and a possible improvement of the latter skills through rhythmic and/or musical training. The lack of a transfer effect in our case could therefore be explained by a specific influence of rhythm on phonological skills in children with learning disabilities and/or a low level of PA, which was not the case for our study population.

In conclusion, our research shows that teachers can take advantage of a rhythmic learning program and introduce it into their daily routine in order to improve the rhythmic skills of their students. Even if our results did not show any particular improvement in non-rhythmic skills, such as phonological awareness or graphomotricity after this learning, other studies have been able to show such improvements, which may require longer training times or a focus on children struggling with language-based deficits such as developmental dyslexia to be seen. Our research thus indicates avenues for educational practice, in terms of non-verbal (rhythmic) intervention as a way to boost phonological awareness in children who struggle with reading.

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 study was reviewed and approved by UQAC Human Research Ethics Committee (#602.463.05) which follows the framework of Enoncé de politique des trois conseils: éthique de la recherche avec des êtres humains. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

AF, AL, and LP contributed to conception and design of the study. IC, JP, and LP performed the statistical analysis. AF wrote the first draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This research was supported by the program of development and support to research and creation of the University of Quebec in Chicoutimi, as well as by the Mission Recherche at the University Paris-Est Créteil Val-de-Marne.

Acknowledgments

We would like to thank the school's teachers and their students for their active participation in this study. We thank Alban Letanneux (Université Paris-Est Créteil Val-de-Marne) for his precious help in the processing of the rhythmic data.

Conflict of interest

Publisher's note

References

Alamargot, D., Chesnet, D., Dansac, C., and Ros, C. (2006). Eye and pen: a new device for studying reading during writing. *Behav. Res. Methods* 38, 287–299. doi: 10.3758/BF03192780

Alamargot, D., and Morin, M. F. (2015). Does handwriting on a tablet screen affect students' graphomotor execution? A comparison between grades two and nine. *Hum. Mov. Sci.* 44, 32–41. doi: 10.1016/j.humov.2015.08.011

Amundson, S. J. (1995). *Evaluation Tool of Children's Handwriting: ETCH Examiner's Manuel.* Homer, AK: O.T. Kids.

Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* 30, 917–926. doi: 10.1016/j.neuroimage.2005.10.044

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Ben-Pazi, H., Kukke, S., and Sanger, T. D. (2007). Poor penmanship in children correlates with abnormal rhythmic tapping: a broad functional temporal impairment. *J. Child Neurol.* 22, 543–549. doi: 10.1177/0883073807302

Bonacina, S., Krizman, J., White-Schwoch, T., Nicol, T., and Kraus, N. (2020). Distinct rhythmic abilities align with phonological awareness and rapid naming in school-age children. *Cogn. Process.* 21, 575–581. doi: 10.1007/s10339-020-00984-6

Bus, A. G., and Van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *J. Educ. Psychol.* 91:403. doi: 10.1037/0022-0663.91.3.403

Carr, K. W., White-Schwoch, T., Tierney, A. T., Strait, D. L., and Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci.* 111, 14559–14564. doi: 10.1073/pnas.1406219111

Chapleau, N., Godin, M.-P., and Beaupré-Boivin, K. (2020). "Des pratiques reconnues efficaces auprès des lecteurs et scripteurs ayant des difficultés," in *Lecteurs et Scripteurs en Difficulté. Propositions Didactiques et Orthodidactiques.* eds. N. Chapleau and M.-P. Godin (Québec, QC: Presses de l'Université du Québec), 5–22.

Chobert, J., François, C., Velay, J. L., and Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* 24, 956–967. doi: 10.1093/cercor/bhs377

Choi, J. Y., Hasegawa-Johnson, M., and Cole, J. (2005). Finding intonational boundaries using acoustic cues related to the voice source. *J. Acoust. Soc. Am.* 118, 2579–2587. doi: 10.1121/1.2010288

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Corriveau, K. H., and Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex* 45, 119–130. doi: 10.1016/j.cortex.2007.09.008

Culter, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words.* Cambridge, MA: Mit Press.

Curtin, F., and Schulz, P. (1998). Multiple correlations and Bonferroni's correction. *Biol. Psychiatry* 44, 775–777. doi: 10.1016/S0006-3223(98)00043-2

David, D., Wade-Woolley, L., Kirby, J. R., and Smithrim, K. (2007). Rhythm and reading development in school-age children: a longitudinal study. *J. Res. Read.* 30, 169–183. doi: 10.1111/j.1467-9817.2006.00323.x

Degé, F., Patscheke, H., and Schwarzer, G. (2020). The influence of music training on motoric inhibition in German preschool children. *Music. Sci.* 26, 172–184. doi: 10.1177/1029864920938432

Dittinger, E., Chobert, J., Ziegler, J. C., and Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Front. Hum. Neurosci.* 11:233. doi: 10.3389/fnhum.2017.00233

Dormoy, A. (2019). "La musique à l'aide des troubles d'apprentissage," in *Bienfaits de la Musique à l'école. Une Expérience Européenne*. ed. J. A. Rodríguez-Quiles (Potsdam: Universitätsverlag Potsdam), 131–148.

Drake, C., and Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Percept. Psychophys.* 54, 277–286. doi: 10.3758/BF03205262

Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 251–288. doi: 10.1016/s0010-0277(00)00106-2.

Evertz, M., and Primus, B. (2017). Rhythmic constraints in handwriting. A review of Pagliarini et al. (2017). *Written Lang. Lit.* 20, 252–258. doi: 10.1075/wll.00006.eve

Feldman, R. (2007). Parent–infant synchrony: biological foundations and developmental outcomes. *Curr. Dir. Psychol. Sci.* 16, 340–345. doi: 10.1111/j.1467-8721.2007.00532.x

Frey, A., François, C., Chobert, J., Velay, J.-L., Habib, M., and Besson, M. (2019). Music training positively influences the Preattentive perception of voice onset time in children with dyslexia: a longitudinal study. *Brain Sci.* 9:91. doi: 10.3390/brainsci9040091

Frey, A., and Sappey-Marinier, A. (2018). La Musique Comme Vecteur de Développement Langagier: Effet d'un Entraînement Musical sur les Compétences Langagières Chez des Enfants de CE2. *Ressources*. 19, 42–57.

Frischen, U., Degé, F., and Schwarzer, G. (2022). The relation between rhythm processing and cognitive abilities during child development: the role of prediction. *Front. Psychol.* 13:920513. doi: 10.3389/fpsyg.2022.920513

Gordon, R. L., Fehd, H. M., and McCandliss, B. D. (2015). Does music training enhance literacy skills? a meta-analysis. *Front. Psychol.* 6:1777. doi: 10.3389/fpsyg.2015.01777

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10. doi: 10.1016/j.tics.2010.10.001

Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906. doi: 10.1162/jocn.2007.19.5.893

Greenberg, S., Arai, T., and Grant, K. (2006). The role of temporal dynamics in understanding spoken language. *NATO Sci. Series Life Behav. Sci.* 374, 171–190.

Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., and Besson, M. (2016). Music and dyslexia: a new musical training method to improve reading and related disorders. *Front. Psychol.* 7:26. doi: 10.3389/fpsyg.2016.00026

Harper, J., and O'Brien, K. (2015). *Classroom Routines for Real Learning: Daily Management Exercises that Empower and Engage Students.* Markham: Pembroke Publishers.

Hierbert, E., and Taylor, B. (2000). "Begining reading instruction: research on early interventions" in *Handbook of Reading Research*. eds. M. L. Kamil, P. B. Mosenthal, P. D. Pearson and R. Barr. Abington: Routledge.

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689. doi: 10.1016/j.cortex.2010.07.010

JASP Team (2022). JASP (Version 0.16.2) [Computer Software]. Amsterdam: Department of Psychological Methods, University of Amsterdam. Available at: https://jasp-stats.org/

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83, 323–355. doi: 10.1037/0033-295X.83.5.323

Jones, M. R., and Boltz, M. (1989). Dynamic attending and responses to time. *Psychol. Rev.* 96:459. doi: 10.1037/0033-295X.96.3.459

Kiefer, M., and Trumpp, N. M. (2012). Embodiment theory and education: the foundations of cognition in perception and action. *Trends Neurosci. Educ.* 1, 15–20. doi: 10.1016/j.tine.2012.07.002

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

Laurencelle, L. (2005). *Abrégé sur les Méthodes de Recherche et la Recherche Expérimentale.* Sainte-Foy, QC: Presses de l'Université du Québec.

Lê, M., Quémart, P., Potocki, A., Gimenes, M., Chesnet, D., and Lambert, E. (2020). Rhythm in the blood: the influence of rhythm skills on literacy development in third graders. *J. Exp. Child Psychol.* 198:104880. doi: 10.1016/j.jecp.2020.104880

Lundetræ, K., and Thomson, J. M. (2018). Rhythm production at school entry as a predictor of poor reading and spelling at the end of first grade. *Read. Writ.* 31, 215–237. doi: 10.1007/s11145-017-9782-9

Martin, M. O., Mullis, I. V., and Hooper, M. (2017). *Methods and Procedures in PIRLS 2016.* Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement.

Ministère de la Famille (2019). *Plan Stratégique 2019–2023.* Québec, QC: Gouvernement du Québec.

Monier, F., and Droit-Volet, S. (2019). Development of sensorimotor synchronization abilities: motor and cognitive components. *Child Neuropsychol.* 25, 1043–1062. doi: 10.1080/09297049.2019.1569607

Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., and Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Read. Writ.* 26, 739–769. doi: 10.1007/s11145-012-9389-0

Oakhill, J., and Kyle, F. (2000). The relation between phonological awareness and working memory. *J. Exp. Child Psychol.* 75, 152–164. doi: 10.1006/jecp.1999.2529

OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*. Paris: PISA, Éditions OCDE.

Ouellette, G., and Sénéchal, M. (2008). Pathways to literacy: a study of invented spelling and its role in learning to read. *Child Dev.* 79, 899–913. doi: 10.1111/j.1467-8624.2008.01166.x

Ozernov-Palchik, O., Wolf, M., and Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *J. Exp. Child Psychol.* 167, 354–368. doi: 10.1016/j.jecp.2017.11.009

Pagliarini, E., Scocchia, L., Vernice, M., Zoppello, M., Balottin, U., Bouamama, S., et al. (2017). Children's first handwriting productions show a rhythmic structure. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-05105-6

Paquet, A. (2017). *Effets de la Pratique Musicale Sur le Rappel Actif Auditif de l'enfant: Une étude de Potentiels Évoqués [Master's Thesis]*. Montréal: Université du Québec à Montréal.

Patscheke, H., Degé, F., and Schwarzer, G. (2018). The effects of training in rhythm and pitch on phonological awareness in four-to six-year-old children. *Psychol. Music* 47, 376–391. doi: 10.1177/0305735618756763

Provasi, J., and Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2-1/2 and 4-year-old children. *Int. J. Behav. Dev.* 27, 220–231. doi: 10.1080/01650250244000290

Provasi, J., Doyère, V., Zélanti, P. S., Kieffer, V., Perdry, H., El Massioui, N., et al. (2014). Disrupted sensorimotor synchronization, but intact rhythm discrimination, in children treated for a cerebellar medulloblastoma. *Res. Dev. Disabil.* 35, 2053–2068. doi: 10.1016/j.ridd.2014.04.024

Pulido, L., and Morin, M. F. (2018). Invented spelling: what is the best way to improve literacy skills in kindergarten? *Educ. Psychol.* 38, 980–996. doi: 10.1080/01443410.2017.1414155

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349–351. doi: 10.1126/science.288.5464.349

Rautenberg, I. (2015). The effects of musical training on the decoding skills of German-speaking primary school children. *J. Res. Read.* 38, 1–17. doi: 10.1111/jrir.12010

Raven, J., Raven, J. C., and Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales.* Oxford, UK: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.

Sala, G., and Gobet, F. (2020). Cognitive and academic benefits of music training with children: a multilevel meta-analysis. *Mem. Cogn.* 48, 1429–1441. doi: 10.3758/s13421-020-01060-2

Schaefer, R. S. (2014). Auditory rhythmic cueing in movement rehabilitation: findings and possible mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130402. doi: 10.1098/rstb.2013.0402

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* 15, 511–514. doi: 10.1111/j.0956-7976.2004.00711

Thomson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol.* 102, 120–129. doi: 10.1016/j.jphysparis.2008.03.007

Tierney, A., and Kraus, N. (2013). Music training for the development of reading skills. *Prog. Brain Res.* 207, 209–241. doi: 10.1016/B978-0-444-63327-9.00008-4

Vaivre-Douret, L., Lopez, C., Dutruel, A., and Vaivre, S. (2021). Phenotyping features in the genesis of pre-scriptural gestures in children to assess handwriting developmental levels. *Sci. Rep.* 11, 1–13. doi: 10.1038/s41598-020-79315-w

Van Galen, G. P., and Teulings, H. L. (1983). The independent monitoring of form and scale factors in handwriting. *Acta Psychol.* 54, 9–22. doi: 10.1016/0001-6918(83)90020-3

Verney, J. P. (2013). *Rhythmic Perception and Entrainment in 5-Year-Old Children [Doctoral Dissertation]*. Cambridge: University of Cambridge.

Véron-Delor, L., Velay, J. L., Braibant, I., and Danna, J. (2017). Qu'apporte la musique à l'apprentissage de l'écriture? Étude de cas auprès d'un enfant en grande difficulté d'écriture. *Approche Neuropsychologique des Apprentissages chez l'Enfant (ANAE)* 151, 719–727.

Wechsler, D. (2005). *WISC-IV CDN-F–Échelle D'intelligence de Wechsler Pour Enfants – 4^{e} édition*. Version pour francophones du Canada. Pearson.

Wenk, B. J., and Wioland, F. (1982). Is French really syllable-timed? *J. Phon.* 10, 193–216. doi: 10.1016/S0095-4470(19)30957-X

Wiig, E., Secord, W., Semel, E., Boulianne, L., and Labelle, M. (2009). *Évaluation Clinique Des Notions Langagières Fondamentales: Version pour Francophones du Canada.* Toronto, ON: Pearson Canada Assessment.

Wolf, M., and Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *J. Educ. Psychol.* 91, 415–438. doi: 10.1037/0022-0663.91.3.415

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Read. Writ.* 15, 179–206. doi: 10.1023/A:1013880723925

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8, 547–558. doi: 10.1038/nrn2152

Zhang, Z., Sun, Y., Humphreys, G. W., and Song, Y. (2017). Different activity patterns for action and language within their shared neural areas: an fMRI study on action observation and language phonology. *Neuropsychologia* 99, 112–120. doi: 10.1016/j.neuropsychologia.2017.02.025

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One* 9:e99868. doi: 10.1371/journal.pone.0099868

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 13 March 2023
DOI 10.3389/fpsyg.2023.1027872

OPEN ACCESS

EDITED BY
Arnaud Leleu,
Université de Bourgogne,
France

REVIEWED BY
Fanny Poncet,
Université de Fribourg,
Switzerland
Sofie Vettori,
CNRS ISCMJ Lyon,
France

*CORRESPONDENCE
Julie Bertels
✉ julie.bertels@ulb.be

SPECIALTY SECTION
This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

RECEIVED 14 September 2022
ACCEPTED 13 February 2023
PUBLISHED 13 March 2023

CITATION
Bertels J, de Heering A, Bourguignon M, Cleeremans A and Destrebecqz A (2023) What determines the neural response to snakes in the infant brain? A systematic comparison of color and grayscale stimuli.
Front. Psychol. 14:1027872.
doi: 10.3389/fpsyg.2023.1027872

What determines the neural response to snakes in the infant brain? A systematic comparison of color and grayscale stimuli

Julie Bertels[1,2]*, Adelaïde de Heering[3], Mathieu Bourguignon[2,4], Axel Cleeremans[1] and Arnaud Destrebecqz[1]

[1]ULBabyLab, Consciousness, Cognition and Computation Group (CO3), Center for Research in Cognition and Neurosciences (CRCN), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles (ULB), Brussels, Belgium, [2]Laboratoire de Neuroanatomie et de Neuroimagerie Translationnelles (LN^{2}T), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles (ULB), Brussels, Belgium, [3]LulLABy, Unité de Recherche en Neurosciences Cognitives (UNESCOG), Center for Research in Cognition and Neurosciences (CRCN), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles (ULB), Brussels, Belgium, [4]Laboratory of Neurophysiology and Movement Biomechanics, ULB Neuroscience Institute (UNI), Université Libre de Bruxelles (ULB), Brussels, Belgium

Snakes and primates have coexisted for thousands of years. Given that snakes are the first of the major primate predators, natural selection may have favored primates whose snake detection abilities allowed for better defensive behavior. Aligning with this idea, we recently provided evidence for an inborn mechanism anchored in the human brain that promptly detects snakes, based on their characteristic visual features. What are the critical visual features driving human neural responses to snakes is an unresolved issue. While their prototypical curvilinear coiled shape seems of major importance, it remains possible that the brain responds to a blend of other visual features. Coloration, in particular, might be of major importance, as it has been shown to act as a powerful aposematic signal. Here, we specifically examine whether color impacts snake-specific responses in the naive, immature infant brain. For this purpose, we recorded the brain activity of 6-to 11-month-old infants using electroencephalography (EEG), while they watched sequences of color or grayscale animal pictures flickering at a periodic rate. We showed that glancing at colored and grayscale snakes generated specific neural responses in the occipital region of the brain. Color did not exert a major influence on the infant brain response but strongly increased the attention devoted to the visual streams. Remarkably, age predicted the strength of the snake-specific response. These results highlight that the expression of the brain-anchored reaction to coiled snakes bears on the refinement of the visual system.

KEYWORDS

infancy, snakes, steady-state visual evoked potential, color, EEG

Introduction

Snakes and primates have coexisted for thousands of years. Given that snakes are the first of the major primate predators (Isbell, 2009), natural selection may have favored primates with appropriate defensive behavior that increased their chances of survival. Such defensive behavior requires quick and efficient prior detection of the danger, and many studies have indeed demonstrated that human and non-human primates developed, accordingly, the propensity to rapidly detect snake-like visual cues. As a consequence, they detect snakes faster than non-snakes in a collection of pictures (Soares et al., 2014; Kawai and Koda, 2016). This predisposition is functional early in development, sensitive to snake characteristic features such as their coiled aspect, and subtended by a neurobiological substrate (Bertels et al., 2020), as evidenced notably by the existence of thalamic neurons in the macaque brain that selectively respond to snake pictures (Van Le et al., 2013).

Several studies support the idea of an inborn predisposition. As a matter of fact, human infants, who have no idea of how dangerous snakes can be and have never experienced these animals, are remarkable snake detectors, like their older peers and non-human primates (Deloache and Lobue, 2009; LoBue and DeLoache, 2010; Bertels et al., 2018). Recently, we provided electrophysiological evidence that glancing at snakes engenders specific occipital responses in the infant brain (Bertels et al., 2020). Indeed, we recorded the brain electrical activity of 7- to 10-month-olds when they watched a series of animal pictures flickering at 6 Hz. Depending on the sequences, snake, frog, or caterpillar images appeared every five images (i.e., at 1.2 Hz). We observed a snake-specific neural response at 1.2 Hz and its harmonics, which was larger in amplitude than that generated by frogs or caterpillars. These results support that humans are, very early on, equipped with a brain-anchored mechanism sensitive to snake prototypical features, functional from the first months of life and independent of any prior exposure to snakes.

What are the critical features driving that response to snakes in infants, and more generally in primates, is an unresolved issue. Their prototypical curvilinear coiled shape is of major importance (e.g., Lobue and Deloache, 2011; Van Strien et al., 2016; Gomes et al., 2018). Indeed, caterpillars, which are elongated as snakes, but not coiled, do not elicit any specific brain response in the occipital brain areas (Bertels et al., 2020). Nevertheless, it could still be that the infant brain responds to a mix of snake-like physical traits (Kawai, 2019), including their scale patterns (Isbell and Etting, 2017; Van Strien and Isbell, 2017), striking posture (Masataka et al., 2010) and coloration (Masataka et al., 2010; Hayakawa et al., 2011). These traits were mixed up in Bertels et al. (2020). In the present study, we aim at specifically examining whether, in human infants, color affects the neural response to coiled snakes.

Color information is a critical cue when processing natural scenes, for several reasons. When applied to images, color is not only capturing attention more than grayscale variations (Zhu et al., 2013) but it also facilitates their segmentation from the background (Delorme et al., 1999; Gegenfurtner and Rieger, 2000), eases discrimination between different categories of stimuli, and contributes to the generalization of variant exemplars of the same category (Or et al., 2019). Accordingly, the importance of color cues for recognition and categorization of natural scenes has already been demonstrated in adults (Gegenfurtner and Rieger, 2000; Oliva and Schyns, 2000; Goffaux et al., 2005). At a neural level, Goffaux et al. (2005) evidenced, for example, that chromatic information speeds up early scene categorization. More recently, Or et al. (2019) showed that color contributes to the rapid detection of faces among natural images, eliciting larger brain responses over occipitotemporal areas than grayscale face pictures.

Regarding snake detection, color information might be especially important. Although snakes do not share a single diagnostic color, their bright, conspicuous coloration could act as an aposematic signal, serving to repel predators and warn preys (Ruxton et al., 2004; Stevens and Ruxton, 2012; Souchet and Aubret, 2016; Prokop et al., 2018). Also, snakes' coloration might, in conjunction with their diagnostic shape, act as a trigger for the rapid detection and the fear responses in primates (Souchet and Aubret, 2016). However, the exact contribution of color cues to snake detection is mixed. Although children detect aposematically colored snakes faster than cryptically colored snakes (Hayakawa et al., 2011; Fančovičová et al., 2020), their rapid detection also operates for grayscale pictures in children, adults, and monkeys (e.g., Shibasaki and Kawai, 2009; Masataka et al., 2010; Hayakawa et al., 2011; Lobue and Deloache, 2011; He et al., 2014). Color would therefore not be necessary for children to rapidly detect snakes. In infants, the question remains unsolved. As they are naïve subjects with immature visual systems and who have not yet learned to fear these reptiles, they could especially benefit from combined visual cues, including color, being indicative of danger.

The present study investigates whether color information contributes to the specific neural response to snakes in the infant brain, using fast periodic visual stimulation combined with electroencephalography, in trichromat infants, i.e., who discriminate colors in the green-red part of the spectrum. We first aimed at replicating Bertels et al.'s (2020) findings of a specific response to fleeting pictures of snakes in their natural background, in comparison to threat-irrelevant creatures. We then assessed how the colorful nature of the stimuli impacted these responses, and finally evaluated to what extent these responses develop with age.

Methods

Participants

Twenty-two full-term infants (6–11 months old; mean ± SD age, 271 ± 43 days; 9 males) with no known neurodevelopmental disorder were included in the study. Infants from the color group ($n = 11$) were presented with the color versions of the pictures, while those of the grayscale group ($n = 11$) were presented with their grayscale versions. Questions to the parents revealed they were not especially familiar with snakes or frogs. The parents gave informed consent prior to testing. The CUB Hôpital Erasme Ethics Committee approved the experimental protocol. The experiments were carried out in accordance with the approved guidelines and regulations.

Stimuli and procedure

Infants were presented with pictures of animals in their natural habitat, from various angles, as in Bertels et al. (2020) (see Figure 1). Infants viewed either color or grayscale versions of the pictures. Both sets were equalized in terms of luminance and contrast using Matlab (Mathworks, United States) and further resized to 200 × 200 pixels.

Snake and frog pictures (29 exemplars of each category) were taken from a set of pictures from LoBue and DeLoache (2008). Snakes were all depicted coiled, none in an attack posture. Frogs were used as non-snake control stimuli since they resemble snakes in texture, brightness, and color, and since they are as unfamiliar to infants as snakes are (Lobue and DeLoache, 2008). Other animal

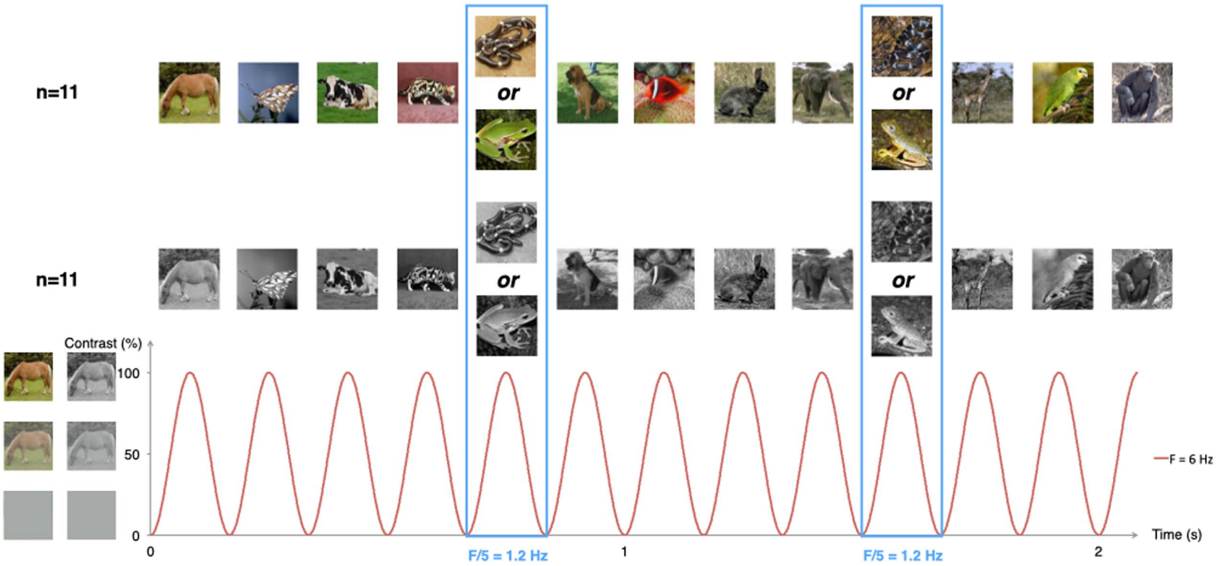


FIGURE 1
Schematic illustration of the experimental paradigm. Animal pictures were presented by sinusoidal contrast modulation at a rate of 6 per second (*F*=6Hz). Snake or frog pictures were presented every fifth stimulus (*F*=6/5=1.2Hz), in different trial sequences. Half of the infants were presented with color pictures, the other half with grayscale versions of the same pictures.

pictures were taken from a set of 130 pictures of 13 different species.[1]

Pictures were presented centrally on a 800 × 600 pixels resolution monitor, on a light grey background. Monitor refreshment rate was 60 Hz. Viewed from 40 cm of distance, images spanned a visual angle of ~13 × 13 degrees.

Visual stimulation was performed using the Psychtoolbox 3.0.9 software for Windows running on Matlab 7.6 (MathWorks Inc.). Images appeared on the screen within 24-s sequences, including a 2-s fade-in and a 2-s fade-out. There was a total of 144 different images per sequence. Each image lasted for 166.7 ms, yielding a presentation rate of 6 Hz (stimulation frequency). The intensity at every pixel was modulated by a squared-sine wave with null contrast at the beginning and end of stimulus presentation and reaching maximal contrast at 83.35 ms (see Figure 1).

Sequences were built as alternations of 4 non-frog non-snake animal pictures followed by 1 frog or snake picture, in frog and snake sequences, respectively. Half of the infants within each group started with a snake sequence. In this setting, the repetition rate of frog or snake images was therefore of 1.2 Hz (6 Hz/5), hereafter referred to as the category-selective frequency. Discrimination by the infant's visual system of frogs or snakes would therefore lead to the identification, in the EEG spectrum, of a response at 1.2 Hz.

Infants were tested in a dimly lit and quiet room. They were sitting in a car seat, with a caregiver behind. The experimenter monitored their looking behavior with a webcam and started each sequence once the infant looked at the screen. Infants watched as many sequences as they could ($M = 13$, SD = 2.8, range = 9–17 in the color group; $M = 8$, SD = 2.72, range 5–12 in the grayscale group). Testing lasted 3–10 min.

Electroencephalography acquisition

Electroencephalography recordings were obtained using a 32-channel BioSemi Active2 system (Amsterdam, Netherlands), with electrodes including standard 10–20 system locations and 2 additional reference electrodes. Signal was recorded at 1024 Hz. The offset of each electrode was kept between ± 25 μV by injecting each electrode with saline gel.

Triggers were used to indicate, at the beginning of each sequence, whether infants viewed a frog or a snake sequence. They were also sent at each image presentation, when its contrast was of 0%.

Electroencephalography pre-processing

To pre-process EEG signals, we used Letswave 6[2] running on MATLAB R2017a (Mathworks Inc.), and followed a standard procedure (e.g., Liu-Shuang et al., 2014).

Individual EEG signals were first band-pass filtered (0.1–100 Hz) and resampled to 250 Hz. Trials of 28 s were extracted from the continuous data, from 2 s before sequence onset to 2 s after sequence offset (see, e.g., de Heering and Rossion, 2015). Trials were further examined for possible channel artifacts. Noisy channels (maximum 3 per infant) were reconstructed by linear interpolation of 2 surrounding clean channels. This was done for six infants. EEG signals were then re-referenced to their common average, and trimmed to exclude the fade-in and fade-out. This resulted in 20-s stimulation sequences.

We then applied a Fast Fourier Transform (FFT) to these 20-s sequences (frequency resolution, 1/20s = 0.05 Hz). The Fourier

---

[1] https://calphotos.berkeley.edu/fauna

[2] http://letswave.org

coefficients of each sequence were further converted into signal-to-noise ratios (SNRs) by taking the ratio between the amplitude at a frequency bin of interest and the mean amplitude at the 12 surrounding frequency bins (6 on each side, excluding the immediately adjacent bins, see Peykarjou et al., 2017; Barry-Anwar et al., 2018).

Given that the infant brain should not synchronize to the stimulation frequency if they are not watching the pictures, sequences with an SNR below 2 at 6 Hz (stimulation frequency) at all the medial occipital electrodes (O1, O2, and Oz) were discarded (see de Heering and Rossion, 2015; Peykarjou et al., 2017; Bertels et al., 2020). This step led to rejecting 2.55 ± 2.36 (mean ± SD; range, 0–7) sequences per infant. The number of sequences kept in the frog (4.14 ± 1.75; range 1–8) and snake condition (3.95 ± 1.86; range 1–8) did not differ significantly ($p > 0.40$). The number of sequences retained was 87 for snake (55 color and 32 grayscale) and 91 for frog sequences (58 color and 33 grayscale). Four frog sequences (3 color and 1 grayscale) were also randomly discarded to ensure comparability of SNR measures between conditions.

Frequency domain analyses

Custom-made MATLAB scripts were used for the analyses. For each remaining sequence, corresponding Fourier coefficients were divided by a single normalization factor taken as the mean amplitude of the Fourier coefficients for frequencies within 0.6–1.8 Hz, a frequency range that surrounded the category-selective frequency (1.2 Hz). This procedure ensured all sequences were given about the same weight at 1.2 Hz, even when they contained excessive movement artifacts. To restore the usual units and scales of the Fourier coefficients, they were all multiplied by the median across sequences of the normalization factor. Raw amplitude spectra were obtained for each condition and electrode as the modulus of the averaged sequences of Fourier coefficients, at the subject level and the group level (averaging the 55 color sequences, and the 32 grayscale sequences, disregarding their origin in terms of participants' identity). These raw spectra at each channel were further corrected by subtracting from the amplitude at each frequency bin, that averaged across the 12 surrounding bins (see Rekow et al., 2021; Supplementary Figure S1). In addition, raw amplitude spectra were converted to SNR responses as described above. SNR responses were further averaged across the 3 occipital electrodes (O1, O2, and Oz) where the responses were observed in Bertels et al. (2020).

In view of statistical appraisal, we also estimated *Z*-scores at the group and the individual level. *Z*-scores were obtained as the difference between amplitude at each frequency bin and mean amplitude at the 12 surrounding frequency bins (excluding the immediately adjacent bins, see below) divided by the standard deviation of the amplitude at these 12 surrounding bins. We hypothesized that stronger responses will be generated at 1.2 Hz and harmonics, for snake than frog pictures.

Statistics

We tested the statistical significance of responses at category-selective frequency (1.2 Hz) and harmonics under 12 Hz (i.e., 1.2, 2.4, 3.6, 4.8, 7.2, 8.4, 9.6, and 10.8 Hz, see Retter et al., 2021). The significance of category-selective responses in infants indeed never exceeds the first harmonics (Peykarjou, 2022). To do so, we used a statistical test akin to a permutation test (Nichols and Holmes, 2002). The test was applied to both group-and subject-level responses for frog and snake sequences separately. The null hypothesis under testing was that frog and snake stimuli would elicit responses of similar amplitude as non-frog non-snake stimuli, leading to a *Z*-score at 1.2 Hz (and harmonics) that does not depart significantly from values expected by chance. To test this hypothesis, we used a previously published test specifically designed to overcome the difficulty linked to the fact that a single *Z*-score was obtained for all participants' data. The test was applied to amplitude at O1, Oz and O2, considered here as the channels of interest (Bertels et al., 2020). The starting point was to re-estimate the *Z*-scores at each tested frequency, averaged across the three occipital electrodes, based on sequences in which either the first or last cycle was removed. In that framework, a permutation distribution (1,000 permutations) for that *Z*-score was built from sequences trimmed in a way that randomizes the position of the frog or snake images while preserving synchrony in image presentation. Namely, for each permutation, we removed the data corresponding to the n first images and $5-n$ last images, n being a random integer between 0 and 4. With this approach, the phase-locking of possible responses specific to frog or snake images was disrupted. We considered the proportion of values in the permutation distribution that were above the observed value is a robust statistical estimate of response (and *Z*-score) significance.

When comparing the SNR between groups or conditions, it is generally statistically advantageous to consider the SNR response averaged across multiple harmonics. This approach is well grounded given that multiple harmonics do not have a direct meaning in terms of underlying pathophysiological processes (Heinrich, 2010; Norcia et al., 2015). In adults, a common approach is to consider all harmonics before the first non-significant one (e.g., Liu-Shuang et al., 2014). However, in infants, responses are usually confined to fewer harmonics (< 3, maximum 9, see Peykarjou, 2022, for a systematic review) that are not necessarily subsequent (e.g., Peykarjou et al., 2022). Therefore, in a second step, we considered the average of the responses across the successive harmonics of 1.2 Hz until the last significant response (i.e., until 7.2 Hz, see "Results" section). This selection of harmonics is warranted by the fact that considering non-significant harmonic responses has no detrimental effect on the quantification of the response (Rossion et al., 2020; Peykarjou et al., 2022).

To compare group-level responses between frog and snake sequences, separately at category-selective and stimulation (6 and 12 Hz) frequencies, we also used a statistical test akin to a permutation test (see above). In that test, the *Z*-scores were contrasted between conditions, and this contrast was compared to a permutation distribution (1,000 permutations) wherein the contrast value was obtained after having shuffled frog and snake sequences. The same procedure was used to compare group-level responses between colorful and grayscale sequences.

Finally, we used Spearman correlations to test the relationship between the infants' age and their neural responses to snakes and frogs.

Results

Stimulation responses (6 and 12 Hz)

Grand-averaged SNR spectra performed on the selected sequences showed clear responses at the stimulation frequencies (6 and 12 Hz), which attested for the successful synchronization of the infants' visual system to the fast presentation of animal pictures. For both snake and frog sequences, these responses were characterized by a medial occipital topography (mean SNRs over the occipital electrodes in the snake sequences at 6 and 12 Hz = 17.57 and 7.88, z-scores = 33.08 and 11.91; mean SNRs in the frog sequences = 19.08 and 11.74, z-scores = 40.37 and 24.81; see Figure 2A).

The comparison between the z-scores associated with the occipital SNR values in snake and frog sequences at the stimulation frequencies did not reveal any significant difference (both $p > 0.45$).

Correlational analyses revealed no significant relationship between the infants' age and the responses at the stimulation frequencies, neither for snake or frog sequences (at 6 Hz: $r_s = -0.003$, $p = 0.990$ and $r_s = -0.163$, $p = 0.468$; at 12 Hz: $r_s = 0.169$, $p = 0.451$ and $r_s = -0.167$, $p = 0.459$).

Category-selective responses (1.2Hz and harmonics)

Significant responses to snake pictures were observed at 1.2, 3.6, and 7.2 Hz in the occipital region (mean SNRs = 3.04, 1.59 and 2.56; $ps < 0.05$, z-scores = 2.35, 2.23 and 3.40; see Figure 2A). The same was however not true for frog sequences, at any of the harmonics before 12 Hz ($ps > 0.05$; see Figure 2A). As a consequence, the averaged brain response across the first 6 harmonics (i.e., until the last significant harmonic for either frog or snake sequences; see Methods) revealed a significant brain signal, in the occipital region, in response to snake (SNR = 1.78, $p < 0.001$; see Figure 2B) but not frog pictures (SNR = 0.91, $p > 0.70$; see Figure 2B). Importantly, the z-score associated with the SNR value in the occipital region averaged across the first 6 harmonics was significantly higher for snakes compared to frogs ($p < 0.01$).

These results thus replicate Bertels et al. (2020)'s findings of a specific response in the infant brain to pictures of snakes in their natural background, in comparison to similarly colorful and unfamiliar but threat-irrelevant animals.

We estimated the correlations between the infants' age and the category-selective response averaged on the first six harmonics, for snake and frog sequences (see Figure 3). These analyses revealed that the older the infant, the higher the response to snakes ($r_s = 0.426$, $p = 0.048$). This association was not significant for frog sequences ($r_s = -0.124$, $p = 0.583$).

Effect of color on category-selective responses

We further analyzed data separately for the color and grayscale infant groups to examine the role of color on the snake responses.

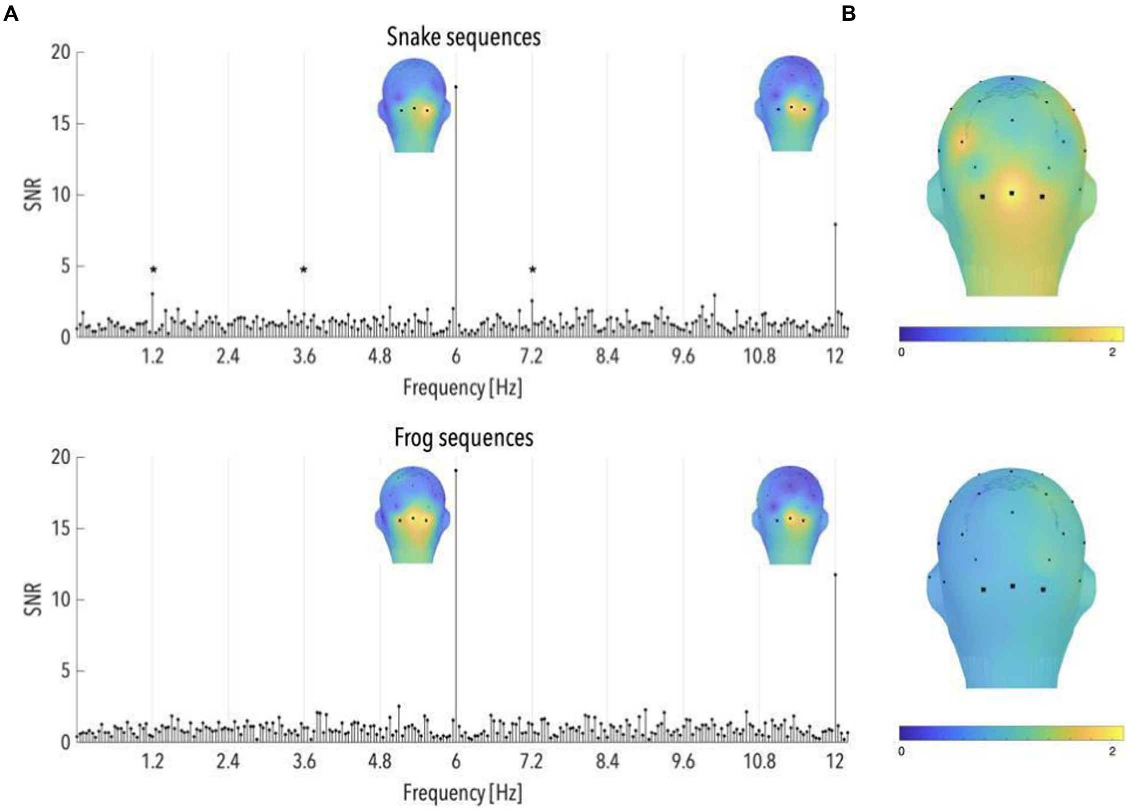


FIGURE 2
(A) Signal-to-noise ratio (SNR) spectra of category-selective and stimulation responses until 12Hz, in the occipital region (data have been averaged across O1, Oz, and O2), and topographical maps of SNR over posterior scalp regions at stimulation frequencies, for snake and frog sequences (n=87 and 91, respectively). The topographies at 6 and 12Hz are shown on their individual maximal color scales. The asterisk indicates a significant discrimination response in the occipital region, of colorful snake pictures. **(B)** Topographical maps of SNR averaged on the first six harmonics of the category-selective response, for snake and frog sequences (upper and lower part, respectively).

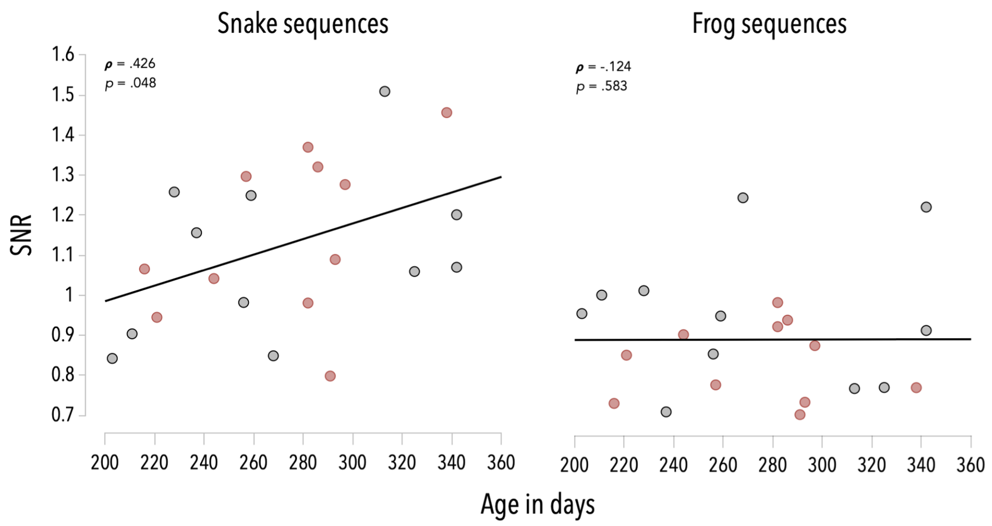


FIGURE 3
Associations between infants' age and SNR values averaged on the first six harmonics of the category-selective response, for snake and frog sequences. Red dots refer to infants in the color group, grey dots refer to infants in the grayscale group.

Analyses of colorful snake sequences revealed significant brain responses at 1.2 and 7.2 Hz (mean occipital SNRs = 3.39 and 2.10; *ps* < 0.03, *z*-scores = 3.69 and 2.02; see Figure 4A), and a significant averaged response across the first 6 harmonics (SNR = 1.74, $p < 0.001$; see Figure 4B). Grayscale snake sequences led to a similar pattern with significant occipital responses observed at 3.6 and 7.2 Hz (mean SNRs = 1.57 and 1.98; *ps* < 0.03, *z*-scores = 2.11 and 2.01; see Figure 4A), as well as a significant averaged response across the first 6 harmonics (SNR = 1.32, $p < 0.05$; see Figure 4B).

To quantify differences between colorful and grayscale snake-specific responses, we compared the z-scores associated with the occipital SNR values in colorful and grayscale snake sequences. As the number of snake sequences considered in each group differed (55 in the color vs. 32 in the grayscale group), we randomly discarded 23 colorful sequences to ensure comparability of SNR measures between conditions. Analyses were run on 20 random selections of sequences to ensure the observed pattern did not depend on the selection of colorful sequences. No significant difference emerged between SNR values averaged across the first six harmonics, nor at 3.6 and 7.2 Hz (*ps* > 0.10). Comparisons at 1.2 Hz revealed higher responses to colorful than to grayscale snake pictures but the significance of that difference depended on the selection of sequences, with about half the selections leading to a significant difference.

Associations between age and category-selective responses—though positive—did not reach significance when considering separately colorful and grayscale snake sequences ($r_s = 0.405$, $p = 0.216$ and $r_s = 0.337$, $p = 0.311$; see Figure 3).

Of note, no significant frog-selective responses were observed, either in the color or in the grayscale group (*ps* > 0.05; see Supplementary Figure S2). No association was observed between the infants' age and the non-significant category-selective responses to colorful or grayscale frogs (both $r_s < -0.10$, $p > 0.80$).

Discussion

Humans are remarkable snake detectors, and this ability would have evolved from the vital need to react adequately in the presence of these ancient, major predators. We recently demonstrated that this evolved predisposition to rapidly detect snakes is brain-anchored and already effective in infants (Bertels et al., 2020). As a matter of fact, snake pictures elicited specific neural responses in the infant occipital brain regions compared to similarly unfamiliar and elongated but uncoiled animals. However, in Bertels et al. (2020) the coiled shape of snakes was blended with other possibly critical visual features, notably their coloration. Yet, coloration has been shown to contribute to the rapid processing of faces and natural scenes (Goffaux et al., 2005; Or et al., 2019), and acts as a powerful aposematic signal (Stevens and Ruxton, 2012). Here, we aimed at clarifying the role of color information in the infants' brain responses to snakes by contrasting the effects of colorful vs. grayscale images.

While we replicated Bertels et al. (2020)'s findings of a selective neural response to snakes, the results do not provide convincing evidence that color information plays a critical role in this response. As a matter of fact, snake-specific responses were not only observed when snakes were depicted in color, but also when grayscale versions of these reptiles were presented to the infants. These results further support that the shape of snakes is a crucial factor in their rapid detection (Lobue and Deloache, 2011). They mirror previous findings in children, adults, and non-human primates that color is not necessary to detect snakes faster than non-snake controls (Shibasaki and Kawai, 2009; Masataka et al., 2010; Hayakawa et al., 2011; Lobue and Deloache, 2011; Kawai and Koda, 2016). Using neurophysiological measures, our results extend these findings to human infants below 1 year of age.

In our study, responses were however stronger—though not significantly so—when infants were exposed to colorful rather than grayscale snakes. Color information could therefore contribute, albeit

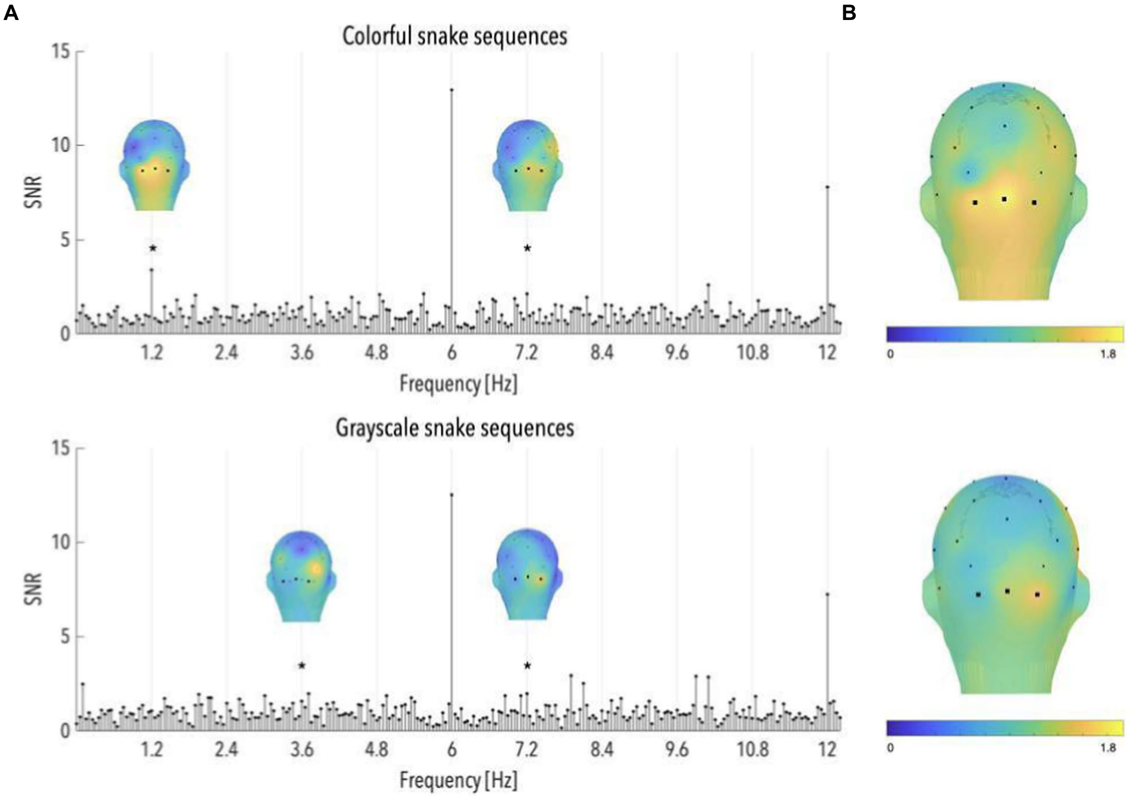


FIGURE 4
(A) Signal-to-noise ratio (SNR) spectra of category-selective and stimulation responses until 12Hz, in the occipital region (data have been averaged across O1, Oz, and O2), and topographical maps of SNR over posterior scalp regions at significant category-selective frequencies, for colorful and grayscale snake sequences (*n*=55 and 32, respectively). The topographies at the harmonics are shown on their individual maximal color scales. The asterisk indicates a significant discrimination response in the occipital region, of colorful and grayscale snake pictures. **(B)** Topographical maps of SNR averaged on the first six harmonics of the category-selective response, for colorful and grayscale snake sequences (upper and lower part, respectively).

slightly, to the rapid categorization of snakes, as it does for faces that are also highly relevant stimuli for the survival of the species (Or et al., 2019). This observation echoes recent findings that trichromat humans—as are most representatives of our species by 3 months (Teller, 1998)—outperform dichromats in detecting predators (Pessoa et al., 2014; de Moraes et al., 2021). These findings have been taken as evidence in favor of the *trichromatic advantage for predator detection hypothesis* (Pessoa et al., 2014) according to which trichromatic color vision evolved in humans and closely related primates due to its evolutionary advantage in detecting predators. In this view, predation pressure, as well as the urge to detect food (Sumner and Mollon, 2003) and mate (Changizi et al., 2006), acted as selective factors to favor trichromacy in humans and Old World monkeys. This hypothesis and its supporting evidence further emphasize the role that color, and color vision more generally, plays in predator detection.

Color does not have a decisive influence on the detection of frogs either. As a matter of fact, we did not observe any frog-selective response whether frogs were presented in color or in black and white. Of note, the absence of a significant response to colorful frogs contrasts with previous findings by Bertels et al. (2020). Yet the number of participants and frog sequences included in both studies is comparable ($n = 22$ vs. 26, and $n = 91$ vs. 109). We hypothesize this result to rise from the fact that the neural responses to frogs are less robust than those to snakes, in infants.

While color only slightly influenced snake detection in our study, and had no effect on neural responses to frogs and other animals, it nevertheless largely impacted the overall amount of attention that infants devoted to flickering stimuli. In fact, infants attentively looked at almost twice as many colored sequences as grayscale ones. This result is in line with studies showing that color stimuli grab more attention than grayscale stimuli (Zhu et al., 2013) and that, from 4 months of age, infants prefer colorful to grayscale exemplars of the same stimuli (Spears, 1966). Color also eases figure-ground segmentation (Delorme et al., 1999; Gegenfurtner and Rieger, 2000), which probably made our pictures more interesting to look at when presented in their color than grayscale versions.

Remarkably, the neural categorical responses to colorful and grayscale snakes increased with age during the second half of the first year, with older infants showing stronger responses to snakes. This finding could be attributed to the general maturation of the visual system. Indeed, even though infants discriminate black-and-white patterns from their first days of life (Fantz and Miranda, 1975), detect complex visual stimuli such as animal or human faces in grayscale image arrays already at 3 months (Simpson et al., 2019), and have an advanced perception of color (Skelton et al., 2022) and visual acuity close to adult levels at 8 months (Courage and Adams, 1990), visual discrimination takes years to refine (Skelton et al., 2022). Obviously, infants also gained more experience during that period, notably with animals, being progressively able to recognize them and discriminate between species. However, if the growing experience was the main cause for that increased response to snakes, we should have observed a similar increase in the brain responses to frogs, which we did not. Rather, there seems to be a specific development in the ability to detect snakes, although throughout the age range tested infants remain naïve about their potential danger. We would therefore argue that the observed improvement in the ability to detect snakes between 6 and 11 months of age relates to the specific development of the visual system with respect to evolutionary-relevant shapes embedded in complex backgrounds, rather than to increased experience with animals, in general.

Overall, this study reveals that color information is not necessary for infants to swiftly detect snakes in natural scenes. Together with previous studies on infants and older individuals, it corroborates that the coiled shape of snakes is the critical feature driving fast response to snakes in primates. It also provides evidence that snake-specific neural responses increase as the visual system matures. Future studies should investigate larger age ranges, from birth on, and explore, using stimuli adapted to the infants' visual abilities (e.g., contrasting stimuli with saturated colors and a background mask), how these responses evolve with developing visual systems, growing life experience, accumulated learning and possible fear experience.

Data availability statement

The datasets generated and analyzed during the current study are available upon reasonable request from the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by CUB Hôpital Erasme Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

JB, AdH, and AD conceived the study. AdH implemented the experiment. JB and AdH performed the study. JB and MB analyzed the data. JB wrote the manuscript. AD, MB, AdH, and AC edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by a FRS–FNRS grant (F.4524.10). JB and AdH were Post-Doctoral Researchers of the Fonds de la Recherche Scientifique-FNRS.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1027872/full#supplementary-material

References

Barry-Anwar, R., Hadley, H., Conte, S., Keil, A., and Scott, L. S. (2018). The developmental time course and topographic distribution of individual-level monkey face discrimination in the infant brain. *Neuropsychologia* 108, 25–31. doi: 10.1016/j.neuropsychologia.2017.11.019

Bertels, J., Bayard, C., Floccia, C., and Destrebecqz, A. (2018). Rapid detection of snakes modulates spatial orienting in infancy. *Int. J. Behav. Dev.* 42, 381–387. doi: 10.1177/0165025417693955

Bertels, J., Bourguignon, M., de Heering, A., Chetail, F., De Tiège, X., Cleeremans, A., et al. (2020). Snakes elicit specific neural responses in the human infant brain. *Sci. Rep.* 10:7443. doi: 10.1038/s41598-020-63619-y

Changizi, M. A., Zhang, Q., and Shimojo, S. (2006). Bare skin, blood and the evolution of primate colour vision. *Biol. Lett.* 2, 217–221. doi: 10.1098/rsbl.2006.0440

Courage, M. L., and Adams, R. J. (1990). Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples. *Am. J. Optom. Physiol. Optic* 67, 713–718. doi: 10.1097/00006324-199009000-00011

de Heering, A., and Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *elife* 4:e06564. doi: 10.7554/eLife.06564

de Moraes, P. Z., Diniz, P., Spyrides, M. H. C., and Pessoa, D. M. A. (2021). The effect of pelage, background, and distance on predator detection and the evolution of primate color vision. *Am. J. Primatol.* 83:e23230. doi: 10.1002/ajp.23230

Deloache, J. S., and Lobue, V. (2009). The narrow fellow in the grass: human infants associate snakes and fear. *Dev. Sci.* 12, 201–207. doi: 10.1111/j.1467-7687.2008.00753.x

Delorme, A., Richard, G., and Fabre-Thorpe, M. (1999). Rapid processing of complex natural scenes: a role for the magnocellular visual pathways? *Neurocomputing* 26-27, 663–670. doi: 10.1016/s0925-2312(98)00158-1

Fančovičová, J., Prokop, P., Szikhart, M., and Pazda, A. (2020). Snake coloration does not influence children's detection time. *Hum. Dimens. Wildl.* 25, 489–497. doi: 10.1080/10871209.2020.1758252

Fantz, R. L., and Miranda, S. B. (1975). Newborn infant attention to form of contour. *Child Dev.* 46, 224–228. doi: 10.2307/1128853

Gegenfurtner, K. R., and Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Curr. Biol.* 10, 805–808. doi: 10.1016/S0960-9822(00)00563-7

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., and Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Vis. Cogn.* 12, 878–892. doi: 10.1080/13506280444000562

Gomes, N., Soares, S. C., Silva, S., and Silva, C. F. (2018). Mind the snake: fear detection relies on low spatial frequencies. *Emotion* 18, 886–895. doi: 10.1037/emo0000391

Hayakawa, S., Kawai, N., and Masataka, N. (2011). The influence of color on snake detection in visual search in human children. *Sci. Rep.* 1:1. doi: 10.1038/srep00080

He, H., Kubo, K., and Kawai, N. (2014). Spiders do not evoke greater early posterior negativity in the event-related potential as snakes. *Neuroreport* 25, 1049–1053. doi: 10.1097/WNR.0000000000000227

Heinrich, S. P. (2010). Some thoughts on the interpretation of steady-state evoked potentials. *Doc. Ophthalmol.* 120, 205–214. doi: 10.1007/s10633-010-9212-7

Isbell, L. A. (2009). *The Fruit, the Tree, and the Serpent*. Cambridge, MA: Harvard University Press.

Isbell, L. A., and Etting, S. F. (2017). Scales drive detection, attention, and memory of snakes in wild vervet monkeys (*Chlorocebus pygerythrus*). *Primates* 58, 121–129. doi: 10.1007/s10329-016-0562-y

Kawai, N. (2019). *The Fear of Snakes. The Science of the Mind*. Berlin: Springer.

Kawai, N., and Koda, H. (2016). Japanese monkeys (*Macaca fuscata*) quickly detect snakes but not spiders: evolutionary origins of fear-relevant animals. *J. Comp. Psychol.* 130, 299–303. doi: 10.1037/com0000032

Liu-Shuang, J., Norcia, A. M., and Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* 52, 57–72. doi: 10.1016/j.neuropsychologia.2013.10.022

Lobue, V., and DeLoache, J. S. (2008). Detecting the snake in the grass: attention to fear-relevant stimuli by adults and young children. *Psychol. Sci.* 19, 284–289. doi: 10.1111/j.1467-9280.2008.02081.x

LoBue, V., and DeLoache, J. S. (2010). Superior detection of threat-relevant stimuli in infancy. *Dev. Sci.* 13, 221–228. doi: 10.1111/j.1467-7687.2009.00872.x

Lobue, V., and Deloache, J. S. (2011). What's so special about slithering serpents? Children and adults rapidly detect snakes based on their simple features. *Vis. Cogn.* 19, 129–143. doi: 10.1080/13506285.2010.522216

Masataka, N., Hayakawa, S., and Kawai, N. (2010). Human young children as well as adults demonstrate "superior" rapid snake detection when typical striking posture is displayed by the snake. *PLoS One* 5:e15122. doi: 10.1371/journal.pone.0015122

Nichols, T. E., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25. doi: 10.1002/hbm.1058

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., and Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *J. Vis.* 15:4. doi: 10.1167/15.6.4

Oliva, A., and Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cogn. Psychol.* 41, 176–210. doi: 10.1006/cogp.1999.0728

Or, C. C.-F., Retter, T. L., and Rossion, B. (2019). The contribution of color information to rapid face categorization in natural scenes. *J. Vis.* 19:20. doi: 10.1167/19.5.20

Pessoa, D. M. A., Maia, R., de Albuquerque Ajuz, R. C., De Moraes, P. Z. P. M. R., Spyrides, M. H. C., and Pessoa, V. F. (2014). The adaptive value of primate color vision for predator detection. *Am. J. Primatol.* 76, 721–729. doi: 10.1002/ajp.22264

Peykarjou, S. (2022). Frequency tagging with infants: the visual oddball paradigm. *Front. Psychol.* 13:1015611. doi: 10.3389/fpsyg.2022.1015611

Peykarjou, S., Hoehl, S., Pauen, S., and Rossion, B. (2017). Rapid categorization of human and ape faces in 9-month-old infants revealed by fast periodic visual stimulation. *Sci. Rep.* 7:12526. doi: 10.1038/s41598-017-12760-2

Peykarjou, S., Langeloh, M., Baccolo, E., Rossion, B., and Pauen, S. (2022). Superior neural individuation of mother's than stanger's faces by five months of age. *Cortex* 155, 264–276. doi: 10.1016/j.cortex.2022.07.011

Prokop, P., Fančovičová, J., and Kučerová, A. (2018). Aposematic colouration does not explain fear of snakes in humans. *J. Ethol.* 36, 35–41. doi: 10.1007/s10164-017-0533-9

Rekow, D., Baudouin, J.-Y., Poncet, F., Damon, F., Durand, K., Schaal, B., et al. (2021). Odor-driven face-like categorization in the human infant brain. *PNAS* 118:e2014979118. doi: 10.1073/pnas.2014979118

Retter, T. L., Rossion, B., and Schiltz, C. (2021). Harmonic amplitude summation for frequency-tagging analysis. *J. Cogn. Neurosci.* 33, 2372–2393. doi: 10.1162/jocn_a_01763

Rossion, B., Retter, T. L., and Liu-Shang, J. (2020). Understanding human individuation of unfamiliar faces with oddball fast periodic visual stimulation and electroencephalography. *Eur. J. Neurosci.* 52, 4283–4344.

Ruxton, G. D., Sherratt, T. N., and Speed, M. P. (2004). *Avoiding Attack*. Oxford: Oxford University Press.

Shibasaki, M., and Kawai, N. (2009). Rapid detection of snakes by Japanese monkeys (*Macaca fuscata*): an evolutionarily predisposed visual system. *J. Comp. Psychol.* 123, 131–135. doi: 10.1037/a0015095

Simpson, E. A., Maylott, S. E., Leonard, K., Lazo, R. J., and Jakobsen, K. V. (2019). Face detection in infants and adults: Effects of orientation and color. *J. Exp. Child Psychol.* 186, 17–32.

Skelton, A. E., Maule, J., and Franklin, A. (2022). Infant color perception: insight into perceptual development. *Child Dev. Perspect.* 16, 90–95. doi: 10.1111/cdep.12447

Soares, S. C., Lindström, B., Esteves, F., and Ohman, A. (2014). The hidden Snake in the grass: superior detection of snakes in challenging attentional conditions. *PLoS One* 9:e114724. doi: 10.1371/journal.pone.0114724

Souchet, J., and Aubret, F. (2016). Revisiting the fear of snakes in children: the role of aposematic signalling. *Sci. Rep.* 6:37619. doi: 10.1038/srep37619

Spears, W. C. (1966). Visual preference in the four-month old infant. *Psychon. Sci.* 4, 237–238. doi: 10.3758/BF03342270

Stevens, M., and Ruxton, G. D. (2012). Linking the evolution and form of warning coloration in nature. *Proc. Biol. Sci.* 279, 417–426. doi: 10.1098/rspb.2011.1932

Sumner, P., and Mollon, J. D. (2003). Colors of primate pelage and skin: objective assessment of conspicuousness. *Am. J. Primatol.* 59, 67–91. doi: 10.1002/ajp.10066

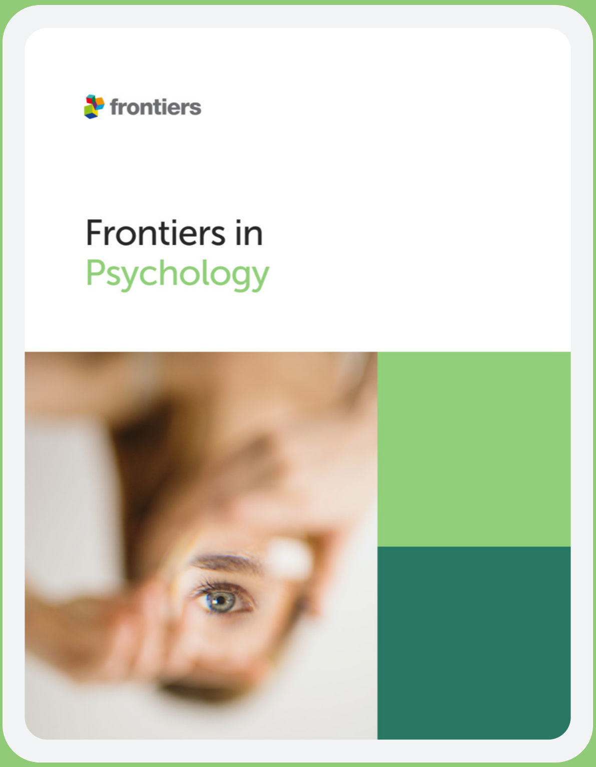
Teller, D. Y. (1998). Spatial and temporal aspects of infant color vision. *Vis. Res.* 38, 3275–3282. doi: 10.1016/S0042-6989(97)00468-9

Van Le, Q., Isbell, L. A., Matsumoto, J., Nguyen, M., Hori, E., Maior, R. S., et al. (2013). Pulvinar neurons reveal neurobiological evidence of past selection for rapid detection of snakes. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19000–19005. doi: 10.1073/pnas.1312648110

Van Strien, J. W., Christiaans, G., Franken, I. H. A., and Huijding, J. (2016). Curvilinear shapes and the snake detection hypothesis: an ERP study. *Psychophysiology* 53, 252–257. doi: 10.1111/psyp.12564

Van Strien, J. W., and Isbell, L. A. (2017). Snake scales, partial exposure, and the Snake detection theory: a human event-related potentials study. *Sci. Rep.* 7:46331. doi: 10.1038/srep46331

Zhu, W., Drewes, J., and Gegenfurtner, K. R. (2013). Animal detection in natural images: effects of color and image database. *PLoS One* 8:e75816. doi: 10.1371/journal.pone.0075816


frontiers
Frontiers in
Psychology