universität wien

DISSERTATION | DOCTORAL THESIS

Titel | Title

Acoustic Properties and Developmental Outcomes of Infant-directed Singing and Music

verfasst von | submitted by

Susanne Reisner BSc MSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 298
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Psych. Dr. Stefanie Höhl

Acknowledgements / Danksagung

I would like to thank the people who have supported me greatly throughout my PhD.

I first want to thank my wonderful supervisors, Prof. Stefanie Höhl and Dr Gabriela Markova. Steffi, thank you for supervising me with so much support, empathy, and openness over the years. Gabi, thank you for your enthusiasm, guidance, and faith in me. Thank you for believing in me, I really could not have asked for better mentors. You inspire me greatly, and you made my PhD experience wonderful!

I would also like to thank my co-authors Dr Trinh Nguyen, Dr Pierre Labendzki, Anja Lueger, Prof. Sam Wass, Dr Félix Bigand, Dr Atesh Koul, Dr Roberta Bianco, and Prof. Giacomo Novembre. Thank you for the interesting work we did together. I am looking forward to our continued collaboration!

A big thank you goes out to the whole Kinderstudien team – you have created such a nurturing and healthy work environment. I treasure our lunchtime discussions and thank you for supporting me through the ups and downs of academia and life. I would also like to thank all the master's students, student assistants, volunteers, and, of course, our participants who make our research possible.

Thank you to Prof. Sam Wass and the whole ISEY team for welcoming me to the University of East London and sharing your expertise so generously. I continue to learn so much from all of you. Special thanks to Pierre, Majo, Gio, Tom D, and Marta for making me feel so welcome in London from the start.

Thank you to the DK CogCom2 cohort. I loved learning from your diverse interests and getting to connect with you.

A huge thank you to all the friends and family who supported me throughout this journey. Ein großes Dankeschön an meine Mutter Hannelore für deine unendliche Unterstützung und Liebe. Danke, dass du mir deine Wissbegierde weitergegeben hast und ich immer auf dich zählen kann. Danke an Marie für deine großartige Begeisterung und bedingungslose Unterstützung. Danke an Eva für deine wunderbare Offenheit und deine ehrlichen Ratschläge. Merci beaucoup à Pierre for having my back in science and in life. Your insights, enthusiasm, unconditional support, and joie de vivre bring me strength, inspiration, happiness, and sunshine every day.

Table of Contents

List of Scientific Publications for the Cumulative Thesis

Accepted for publishing in a journal/book/anthology:

1) Nguyen, T., Bigand, F., Reisner, S., Koul, A., Bianco, R., Markova, G., Hoehl, S., & Novembre, G. (2026). Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year. *bioRxiv*, 2025.05.04.649695. https://doi.org/10.1101/2025.05.04.649695 [1]
2) Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313
3) Reisner, S., Nguyen, T., Labendzki, P., Hoehl, S., & Markova, G. (2025). The reciprocal relationship between maternal infant-directed singing and infant gaze. *Musicae Scientiae, 0*(0). https://doi.org/10.1177/10298649251385676

Notes.

[1] This manuscript has been accepted for publication in *eLife* and is currently available on the journal homepage as a reviewed, not revised preprint: https://doi.org/10.7554/eLife.107088.1

1. Infant-Directed Music and Singing: Acoustic Basis and Developmental Outcomes

1.1. General Introduction and Thesis Overview

Infant-directed (ID) songs are proposed to be specially tailored to the needs of infants. ID music and singing differ from their non-ID equivalents in some acoustic and performance aspects (Trainor, 1996; Trehub, 2017; Trehub et al., 1997). Caregivers widely and intuitively use them to meet various needs, such as entertaining, soothing, or capturing and holding attention (Rock et al., 1999; for a review, see Sharman et al., 2023). Like in speech (Cooper & Aslin, 1990), infants prefer the infant-directed version of music and singing to the adult-directed one (Trainor, 1996). So far, we know that infants neurally respond to rhythm (Edalati et al., 2023) and pitch (Carral et al., 2005) and that they respond to recorded musical stimuli with body movements (Ilari, 2015; Zentner & Eerola, 2010).

However, in order to investigate whether the ID musical input has the intended communicative effects, we need to examine how infants actually engage with musical stimuli. To shed light on how infants engage with music, we first need to investigate how they process different acoustic features and musical structures at the neural level, and how they react to them at the behavioural level. Then we can investigate the multimodal (i.e., neural and behavioural) reactions to ID musical input in an interactive setting and zoom in on the moment-to-moment timescale to elucidate how infants actually engage in these interactions. These questions are especially important to investigate on multiple modalities, both in well-controlled experiments and in live interactions. It is largely unknown which acoustic qualities in ID music and singing evoke the strongest neural and behavioural responses, how different modalities are connected, and which are most engaging in musical interactions. This would provide insight into reactions to musical stimuli, as well as the mechanisms underlying music processing and the role they play in musical interactions. A perspective on the moment-to-moment timescale of how attention flows during reciprocal interactions would shed light on how caregivers and infants structure musical interactions in different contexts depending on infants' needs and how infants engage in musical interactions.

Therefore, the present thesis aims to investigate the following questions:

- How do infants process ID music and singing at the neural level? Do musical structure or acoustic qualities like pitch, tempo, loudness, or rhythmicity play a role in neural processing?

- How do infants behaviourally respond to ID music and singing? Do musical structure, acoustic qualities, or certain types of songs affect their movements, attention, or language outcomes?
- How do infants and caregivers interact during live ID singing interactions? Are caregivers or infants more responsive to each other during certain types of songs? How do caregivers adjust their singing contingently to infant attention in real time?

The following studies aim to answer these questions: Study 1 investigates how infants process ID music via neural and sensorimotor responses to controlled ID musical stimuli over the first postnatal year of life. It examines whether musical structure and pitch influence neural processing and motor output. Study 2 explores how infants engage with ID music by investigating their neural tracking and motor responses to two distinct types of live ID singing at 7 months, as well as their language outcomes at 20 months. Study 3 extends Study 2 by investigating the relationship between infant attention and maternal acoustic variability, giving a much-needed interactional perspective on ID singing and how it can scaffold infant attention. Together, these studies shed light on how infants process and respond to ID music and singing across multiple modalities, how caregivers scaffold infants' attention during ID singing, and how caregivers and infants meaningfully engage with each other in the context of ID singing.

1.2. Music and Singing in Development

1.2.1. What Are Infant-Directed Music and Singing, and When Are They Used?

Infant-directed (ID) music and singing are special forms of ID communication. They refer to musical content that caregivers intuitively and routinely use in interaction with infants all over the world (Ilari, 2005; Steinberg et al., 2021; Trehub, 2019; Trehub et al., 1997), who are especially receptive to ID communication in their first year of life (Trainor, 1996; Tsang et al., 2017). Singing to infants seems to have been a ubiquitous part of human culture throughout history (Mehr et al., 2019; Savage et al., 2021). Nowadays, the majority of U.S. American infants' music exposure consists of recorded vocal and/or instrumental music, though their musical soundscape includes a mix of live and recorded vocal and instrumental music (Mendoza & Fausey, 2021). While ID singing has the advantage for caregivers of being readily available without preparation, recorded ID music can entertain infants without caregivers having to take on an active, productive role in the interaction. Depending on the context, it might be more convenient for caregivers to use recorded ID music in certain

caregiving situations, for example, when the focus is less on the interaction itself, but rather on other caregiving goals like entertainment, distraction, or routine-building.

Still, the communal context of music is perhaps one of its most important ones, as it fosters social connection (Savage et al., 2021), which extends to singing as a potential means of infant-caregiver bonding (Trehub et al., 1997). Especially in the early years of life, singing to infants is a widespread method to foster bonding with caregivers (Ilari, 2005). Besides bonding with their infants, caregivers use ID singing and music in a myriad of ways: To promote affect attunement between infant and caregiver, to aid infant emotional regulation, to assure caregivers of their role, and to attract, maintain or diffuse infant attention (Rock et al., 1999; for a review, see Sharman et al., 2023). Especially the interactional nature of ID singing may even help to strengthen the infant-caregiver relationship.

Culture influences the type of singing interaction caregivers use. Caregivers in cultures where infants are carried in close physical contact for most of the day and co-sleep at night predominantly sing soothing melodies to their infants (Trehub & Gudmundsdottir, 2015). Singing interactions in urban Western cultures, where infants have less physical contact with their caregivers (Hewlett & Lamb, 2002; Trehub, 2019), tend to be in a playful face-to-face context, while soothing lullabies are reserved for putting infants to sleep (Trehub, 2019). Of course, nowadays, non-Western cultures are not cut off from Western cultural influences and music (Trehub & Cirelli, 2018). In conclusion, ID singing and music are common caregiving practices and have diverse uses. Not only does the practice of ID singing unite humans, but it is constituted by shared acoustic qualities.

Acoustic properties.

ID music and singing are characterised by specific acoustic properties. Most research quantifying the acoustic qualities of ID songs has focused on ID singing. Compared to ID speech, ID singing is more fixed in its form and content, can contain words or non-lyrical melodies, and has more repetitions, and consistent pitch and tempo (Bergeson & Trehub, 2002; Longhi, 2009; Nakata & Trehub, 2011; Trehub, 1990). ID singing has a slower tempo, longer pauses, higher and more variable pitch, less timing variations, and more accentuated rhythm compared to self- (Nakata & Trehub, 2011; Trainor, 1996; Trainor et al., 1997; Trehub et al., 1997) or adult-directed singing (Hilton et al., 2022). Those physically quantifiable acoustic qualities are complemented by qualitative ratings about voice quality, such as emotional engagement, loving tone, warmth, or involvement (Trainor, 1996; Trehub et al., 1997). These acoustic qualities are accentuated by the physical presence of an infant.

Caregivers (and even older siblings) are rated to be more emotionally expressive when singing live to an infant (Trehub et al., 1994, 2016), and even untrained adult listeners recognise renditions of songs when infants are present vs absent (Trainor, 1996). The presence of infants incentivises caregivers to sing with a more accentuated rhythm and loving tone. Possibly because of those acoustic quality differences, infants prefer recordings of infant-present over infant-absent ID singing (Trainor, 1996).

While recorded ID music does not change based on infant presence or absence, it has the added option for multiple parts or melodies, as well as the advantage of being easy to control in experiments, which opens the possibility for thorough examination of which acoustic qualities infants prefer, attend to, or engage with the most. Live singing, on the other hand, is much more variable and allows for multimodal interaction. For example, adults adjust their eye movements to highlight salient information (Lense et al., 2022) and exhibit positive facial affect, energetic movements (Cirelli et al., 2020; Trehub et al., 2016), and idiosyncratic gestures during ID singing (Eckerdal & Merker, 2008)

Two Distinct Types of ID Music and Singing.

Based on its acoustic qualities and intended function, ID music is often divided into two distinct categories: playsongs and lullabies. When sung, playsongs are generally faster, higher-pitched, and more dynamically variable than lullabies (Cirelli et al., 2020; Rock et al., 1999; Trainor, 1996; Trainor et al., 1997). Infants seem to reinforce these acoustic differences via their behaviour: Tested in head-turn and preferential looking paradigms, 6-7-month-olds prefer lower pitch in lullabies and higher pitch and faster tempo in playsongs (Conrad et al., 2011; Tsang & Conrad, 2010; Volkova et al., 2006). The instances in which caregivers use those distinct types depend not only, as mentioned above, on cultural influences, but also on the infant's state and the goal the caregiver wants to achieve. To emotionally regulate infants and direct their attention with ID singing, caregivers often use lullabies to soothe infants, and playsongs to arouse and attract attention (Rock et al., 1999; Trainor, 1996), though the choice of song may be dependent on the infant's emotional and attentional state. Caregivers more often choose playsongs over lullabies to calm upset infants (Cirelli & Trehub, 2020), while lullabies soothe already calm infants even more (Cirelli et al., 2020). It seems that playsongs have an attention-grabbing effect, while lullabies are attention-diffusing. To determine whether those effects actually occur in infants, one needs to investigate how infants perceive and respond to musical stimuli.

1.2.2. How Do Infants Perceive and React to Musical Stimuli?

Neural Processing.

Infants' ability to perceive musical stimuli can be measured via their neural responses. Infants are already sensitive to various musical components from before birth: At 30 to 33 gestational weeks, infants can already selectively process musical beat and tempo (Edalati et al., 2023). During violation-of-expectation electroencephalography (EEG) experiments, newborns' brains show mismatch negativity, an event-related potential (ERP) to a rare deviant stimulus, to tempo changes (Háden, Honing, et al., 2015), downbeat omission (Winkler et al., 2009), pitch direction changes (Carral et al., 2005), and pitch intervals in musical stimuli (Stefanics et al., 2009). FMRI experiments further show that newborns are sensitive to key changes and consonance/dissonance (Perani et al., 2010). These findings point to fairly advanced musical processing already at birth.

As infants mature, their processing abilities improve to take on more complex tasks, such as recognising familiar melodies (Plantinga & Trainor, 2009) and rhythms (Demany et al., 1977) at around 2 months of age, measured via head-turn or preferential looking paradigms. While these abilities are thought to be culturally universal, culture-specific musical scales and rhythms also influence infants. Sensitivity to some rules of Western harmonic scales might already be present and measurable via event-related potentials (ERPs) in newborns, such as the sensitivity to dissonant and minor chords (Virtala et al., 2013). This cultural sensitivity can be accelerated by appropriate musical training, even in infants as young as 6 months of age (Gerry et al., 2012). Neural responses to other cultural-musical aspects, including harmony or key membership, however, seem to develop towards an adult-like response around 4 years of age (Corrigall & Trainor, 2014, 2019). Similarly, temporal processing skills, such as phase-locking to pure tones, develop with age (Shahin et al., 2010). While basic music processing abilities seem to be present from very early in life and develop with age, other modalities, such as attentional and behavioural responses, may support their further development.

Attentional Responses.

Measuring attention provides insight into early cognitive development by highlighting important, salient signals for infants. Attentional responses early in development can even predict later cognitive and social outcomes (Braithwaite et al., 2020; Pons et al., 2019; Werchan et al., 2025). Infant attention can be recorded via looking behaviour, neural signals,

or physiology. For example, looking behaviour can be tracked via preferential looking or look duration (Pons et al., 2019; Werchan et al., 2025), neural signals via theta power (Braithwaite et al., 2020; Perapoch Amadó et al., 2023; Xie et al., 2018), and physiological signals via heart rate (Perapoch Amadó et al., 2023; Zhang et al., 2025). Importantly, those modalities are interconnected with respect to attention and developmental change. In live interactions, autonomic physiological arousal precedes shifts in attentive looking in 5-month-olds (Perapoch Amadó et al., 2023). Only by 10 months of age does neural activity (theta power) follow visual attention shifts, predict the lengths of sustained attention, and modulate physiological arousal changes (Perapoch Amadó et al., 2023).

Attentional responses to musical stimuli can be observed very early in life: Infants seem to remember musical stimuli they repeatedly heard in the womb (Partanen et al., 2013). As development progresses after birth, infants can react in more sophisticated ways. Overall, in their first year of life, infants pay more attention (measured via looking time) to ID singing than to ID speech (Nakata & Trehub, 2004; Tsang et al., 2017). In preferential looking paradigms, infants aged four to thirteen months look longer towards positive, loving affect in ID singing (Corbeil et al., 2013; Trainor, 1996) and to familiar, live-performed melodies (Kragness, Johnson, et al., 2022; Mehr et al., 2016).

Furthermore, playsongs are more attention-grabbing than lullabies: During playsongs, 6-month-old infants look longer towards their caregiver, whereas during lullabies, they look longer towards their own bodies or objects they are interacting with. This could be connected to playsongs' brilliance, rhythmicity, and pitch variability, since the acoustic qualities of playsongs are suggested to convey their intention to be attention-grabbing and fun (Rock et al., 1999). Additionally, playsongs could invite more interpersonal connection, which is supported by higher caregiver emotional expressivity, especially positive caregiver affect (Markova et al., 2020; Trehub et al., 2016). Further corroborating this idea, the multimodality of singing elicits higher attention (measured via preferential looking) than audio-only versions of the same ID songs in 5-6-month-olds (Trehub et al., 2016). Additionally, rhythm seems to play an important role for infant attention: During video playbacks, infants look towards the singer's eyes on the beat of singing already at two months of age (Lense et al., 2022).

Similar to neural processing, culture-specific perceptual narrowing also develops in the attentional domain. Infants, similar to adults, pay more attention to rhythmic irregularities in music of their own culture vs an unfamiliar culture at 12 months but not at 6 months of age

(Hannon & Trehub, 2005a, 2005b). While 6-month-olds do not differentiate between preserved and disrupted rhythmic structure in music from familiar and unfamiliar cultures (Hannon & Trehub, 2005a), 12-month-olds, like adults, have lost the ability to distinguish between them. However, this lost sensitivity can be regained by brief post-experimental exposure to unfamiliar music (Hannon & Trehub, 2005b).

ID singing appears to have profound effects on infant attention, which could have roots in both its acoustic qualities and interactional aspects. To better understand how ID songs become such a powerful tool for regulating infant attention, one has to look at infants' behavioural responses.

Movement Responses.

Translating neural processing and attention to musical stimuli, infants can respond to ID singing and music with movement responses from quite early in development: Already at 35 gestational weeks, foetuses show increased movement in response to music vs speech exposure (Kisilevsky et al., 2004). Five- to twenty-four-year-old infants spontaneously and rhythmically move more to music than to speech (Ilari, 2015; Zentner & Eerola, 2010), and they show tempo flexibility: They adjust their movement speed to the tempo of the music, even though they cannot perfectly synchronise to a beat yet (Rocha & Mareschal, 2017; Zentner & Eerola, 2010). Only later in toddlerhood, around 4 years of age, are children able to better synchronise their movements to rhythmic stimuli (Provasi & Bobin-Bègue, 2003). As children grow older, rhythmic complexity such as syncopation (i.e., the emphasis of non-dominant beats) affects movements: Similar to adults, 3-6-year-old children prefer music with medium syncopation for dancing (Cameron et al., 2023).

Interestingly, motor and vestibular processes might play an important role in music processing. Before infants master their own motor system properly and actively, they experience passive vestibular movements, which has an impact on their own movements. For example, infant spontaneous drumming not only becomes more regular and faster with age, it also correlates with caregiver body height. This correlation of infant spontaneous drumming with parent body height, rather than with the infant's own body height, suggests that passive vestibular experiences are impactful from 5 months to 3 years of age (Rocha et al., 2021a). Corroborating this conclusion, manipulating infant-carrying caregivers' walking speeds has an effect on 10-month-old infants' spontaneous drumming (Rocha et al., 2021b). Together, these studies on spontaneous motor tempo, i.e., the natural rate of rhythmic movement (Fraisse, 1982), suggest that the caregivers' walking pace influences their infants'

spontaneous body movements via the passive rhythmic experience of being carried (Rocha et al., 2021a, 2021b). Furthermore, the rhythm at which infants are bounced to a metrically ambiguous rhythm influences their preferences. They prefer the rhythm at which they had been bounced (Phillips-Silver & Trainor, 2005).

It seems that movement responses to musical stimuli, like aspects of neural or attentional responses, are present from early in life. How these three aspects interconnect between each other could give important insights into the underlying processing mechanisms and the engagement with musical stimuli.

1.3. Section Summary

ID music and singing are almost universally present, distinct musical forms of ID communication, characterised by their special acoustic qualities and behavioural, neural, and physiological effects on infants. The acoustic qualities seem to connect to certain behavioural and neural outcomes. Infants neurally respond to rhythm and pitch, and develop culture-specific responses with age. Behavioural responses to musical stimuli include movements and attention.

Examining how music processing and the reaction to musical stimuli interconnect between different modalities would broaden our understanding of the importance of ID musical communication for infant engagement with musical input and could have implications for early cognitive and behavioural outcomes, as well as long-term language and social development. To reach a well-founded understanding, it is necessary to investigate multimodal reactions to both well-controlled, yet complex, recorded ID musical stimuli and live, face-to-face musical interactions. The studies presented in this dissertation attempt to do exactly that. The studies presented here focus on 1) the developmental trajectory of the connection between neural and sensorimotor processing of controlled ID musical stimuli, 2) the interplay between neural responses, rhythmic movement, and language outcomes following a live ID singing interaction, and 3) the interplay of different modalities between caregivers and infants during live ID singing interactions.

2. Studies

2.1. Study 1. Development of Auditory and Spontaneous Movement Responses to Music Over the First Postnatal Year

Engaging with music through movement is considered a core feature of musicality (Honing et al., 2015; Schachner et al., 2009; Trehub et al., 2015), which is the human propensity to perceive, appreciate, and produce music(Honing, 2018; Trehub, 2003). In Study 1, we investigated the developmental trajectory of two fundamental neurocognitive components of musical engagement: the sensory component and the motor component. The sensory component comprises the ability to perceive and recognise music (Brown, 2022; Trehub, 2003; Trevarthen, 1999). For example, infants can neurally process beat structure, pitch deviations, and regularities in tone intervals (Bianco et al., 2025; Edalati et al., 2023; Háden et al., 2009, 2024; Háden, Németh, et al., 2015; Stefanics et al., 2009; Winkler et al., 2009). One way to observe the sensory response to music is to record auditory event-related potentials (ERPs), such as the P1, a positive, phase-locked neural response occurring 200-300ms after an auditory stimulus (Chen et al., 2016; Kushnerenko et al., 2002; Wunderlich et al., 2006). Auditory evoked potentials can also be captured via frequency-domain analyses (Damsma et al., 2025; Novembre & Iannetti, 2018), such as auditory steady-state responses (ASSRs), also known as steady-state evoked potentials (SSEP; e.g., Cirelli et al., 2016; Nave et al., 2022).

The motor component comprises the ability to vocalise, produce percussive actions, and move to music in a temporarily aligned way (Brown, 2022; Fitch, 2015; Honing et al., 2015; Trehub et al., 2015). Infants have been found to respond to music with movements from as young as 28-35 gestational weeks. While 5-24-month-olds (Zentner & Eerola, 2010) show more spontaneous rhythmic movement to classical and child-directed music than to speech, these movements are not beat-synchronised, at least in younger infants (3-4 months old; Fujii et al., 2014; Zentner & Eerola, 2010).

We further sought to answer what could modulate the sensory and motor components and decided to investigate the role of pitch. While higher pitch is more prominent in daily ID speech and music for infants (Fernald & Simon, 1984; Nakata & Trehub, 2011) and infants neurally encode higher pitches better than lower pitches (Marie & Trainor, 2013, 2014), adults tend to move more in response to lower pitches (Cameron et al., 2022; Stupacher et al., 2013, 2016; Van Dyck et al., 2013). Studying the effect of pitch on the neural encoding of

music and its impact on movement could shed further light on the development of the sensory and motor components of musical engagement.

To date, the sensory and motor components have been studied separately, often without a control condition to test whether infants responded to musical structure or purely to acoustic stimuli (i.e., original songs with intact musical structure vs shuffled pitch and rhythm sequences), and not across development. Study 1 investigates whether sensory and motor reactions to musical stimuli are specific to musical structure by exposing infants to songs where musical structure was intact (music condition) or disrupted (shuffled music condition). Further, Study 2 investigated whether pitch could play an attenuating or strengthening role (high pitch vs low pitch condition), how the sensory and motor components intertwine, and how responses to musical stimuli develop over the first postnatal year of life. The concurrent investigation of both components across development could shed light on how perception translates into action in the context of musical engagement, and how this translation changes over development.

The hypotheses of Study 1 were threefold: 1) Auditory neural responses were expected to be stronger in the music condition compared to the shuffled music condition, because musical structure, notably absent in the shuffled music condition, is essential in drawing attention to predictable events (Kouider et al., 2015; Lense et al., 2022). 2) The music condition would elicit more spontaneous rhythmic movements than the shuffled music condition, based on previous findings of music eliciting more movement than speech or silence in 3-24 month-olds (Fujii et al., 2014; Zentner & Eerola, 2010). 3) Neural responses would be enhanced in response to high-pitched music compared to low-pitched music, since infants show high voice superiority from 3 months of age (Marie & Trainor, 2013, 2014).

Seventy-nine infants aged 3 months (n = 26), 6 months (n = 26), and 12-13 months (n = 27), and an adult control sample (n = 26) were exposed to prerecorded playsong refrains consisting of a melody line and a bass line. Infants listened to four conditions: music, shuffled music, high-pitched music, and low-pitched music. The shuffled music condition was in the same pitch range as the music condition, but notes were shuffled and inter-note-onset intervals were randomised to manipulate musical structure. The bass and melody line were shuffled independently from each other. The high-pitched and low-pitched music conditions followed the same musical structure as the music condition; only the melody line was transposed to one octave higher in the high-pitched music condition, and the bass line

was transposed to one octave lower in the low-pitched music condition compared to the music condition.

Infant and adult sensory, neural responses were recorded via EEG, from which event-related potentials (ERP, time-domain analysis) and auditory steady-state responses (ASSR, frequency-domain analysis) were extracted. ERPs can show time-locked neural responses to tones, specifically the latency and amplitude of the P1, a neural response peaking at 200-300ms after an auditory stimulus in infants (Chen et al., 2016; Kushnerenko et al., 2002; Wunderlich et al., 2006). ASSR were a complementary frequency-domain analysis to explore the neural mechanism underlying responses to musical stimuli. Infant spontaneous rhythmic movement patterns were estimated via markerless pose estimation from video recordings via DeepLabCut (Mathis et al., 2018; Nath et al., 2019) and quantified with principal component analysis (PCA) into 10 principal movements (PMs; Bigand et al., 2024) . These PMs were compared between conditions for a difference in Quantity of Movement (QoM). Furthermore, Granger Causality Analyses shed light on auditory-motor coupling by investigating whether changes in intensity of the auditory stimulus influenced spontaneous rhythmic movements. To investigate whether changes in music intensity evoked phase-locked movements (i.e., if movements were synchronised to the beat), event-related analyses were performed on the same data used for the Granger Causality Analyses.

EEG data showed that infants across all age groups had enhanced auditory responses to music compared to shuffled music, both in ERPs and ASSR. The only exception was that ASSR to music vs shuffled music were only a trend in 6-month-olds. Cluster-based permutation analyses of ERPs showed that this heightened neural response to the music condition became more adult-like over development. Specifically, while all infants exhibited the enhanced positivity response known as the infantile P1 to music but not to shuffled music, only 12-month-olds showed a second positivity, the infantile P2. Furthermore, P1 peaks increased in amplitude and decreased in latency with age. Comparing the high- and low-pitch music conditions, infants showed a higher infantile P1 response to higher pitch only at 6 months, and showed no preference between high and low pitch at 3 and 12 months.

Movement responses were present in all age groups, but only 12-month-olds moved more (higher QoM) in response to music compared to shuffled music, primarily driven by an increase in upper-body movements. There was no difference in QoM between the high- and low-pitch conditions at all ages. As a measure of auditory-movement coupling, Granger causality analyses revealed that intensity changes in music were a better predictor of

movement velocity than shuffled music at all ages. Additionally, the high-pitch music condition predicted movement velocity better than the low-pitch music condition. Rhythmic movements were uncoupled to musical structure in all age groups, meaning that movements were not synchronised to specific musical events, and movement periodicity was not significantly different between music and shuffled music.

This study demonstrated that enhanced sensory-motor responses to music over shuffled music are present as early as 3 months, and they mature with age. Neural responses become more adult-like over the first year of postnatal life, both becoming quicker and higher in amplitude. These results fit into existing literature postulating that infants are sensitive to rhythmic or pitch regularities at various ages in infancy (e.g., Cirelli et al., 2016; Edalati et al., 2023; Háden et al., 2009; Stefanics et al., 2009; Trevarthen, 1999), which may be possible due to auditory predictive processing (Vuust & Witek, 2014). Two neural response analysis approaches – ERPs and ASSR – yielded similar results: enhanced neural responses to music compared to shuffled music. This suggests that both these analysis approaches measured evoked responses (Capilla et al., 2011; Novembre & Iannetti, 2018). High pitch being easier to track only for 6-month-olds could be linked to the detection of socially relevant stimuli. At this age, face-to-face interactions peak (Beebe et al., 2016; Feldman, 2007), and since ID communication is characterised by high pitch (Hilton et al., 2022; Tsang & Conrad, 2010), the observed neural processing advantage of high pitch might reflect infants' current developmental phase of musical and communicative processing (Cooper et al., 1997; Newman & Hussain, 2006).

Music not only elicited specific neural responses, but it also led to movement responses at all ages. Movement velocity was better predicted by music compared to shuffled music in all age groups, but only 12-month-olds moved more in response to music compared to shuffled music. This, together with the result that no age group was able to synchronise their movements to musical events like the beat, suggests that movements to auditory stimuli are present from a very young age, but the motor system develops with age. A potential explanation for the increase in QoM by 12 months could be the infants' seated position (as opposed to lying in a car seat at younger ages) or the increase in postural control typically achieved by 9-10 months of age (Hadders-Algra, 2005). The lack of synchronised movements aligns with previous literature, which suggests a gradual refinement of motor control from individual muscle control to an integration of different muscles that allows coordinated movements (Hinnekens et al., 2023). The fact that high-pitched music, rather than low-

pitched music, led to better auditory-motor coupling is difficult to explain, but one possible explanation could be that higher pitch elicited higher arousal, which may have enhanced auditory-motor coupling.

This study demonstrates that both neural and movement responses to musical stimuli mature over the first postnatal year of life. The advantage of having multimodal measurements allows us to consider both neural and movement results together: While audio-motor coupling triggered by musical stimuli emerges early in development, neural tracking precedes the development of phase-locked motor responses. This suggests that the lack of synchronised movement to musical stimuli in the first postnatal year of life is not due to a lack of sensitivity to rhythm in the infant brain.

Nguyen, T., Bigand, F., Reisner, S., Koul, A., Bianco, R., Markova, G., Hoehl, S., & Novembre, G. (2026). Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year. *bioRxiv*, 2025.05.04.649695. https://doi.org/10.1101/2025.05.04.649695 [1]

Notes.

[1] This manuscript has been accepted for publication in *eLife* and is currently available on the journal homepage as a reviewed, not revised preprint: https://doi.org/10.7554/eLife.107088.1

2.2. Study 2. Sing to Me, Baby: Infants Show Neural Tracking and Rhythmic Movements to Live and Dynamic Maternal Singing

Infants prefer ID communication, both in the speech and the singing domain (Cooper & Aslin, 1990; Trainor, 1996). On a neural level, infants respond to ID speech with coherence between volume fluctuations and neural activity (Attaheri et al., 2024; Menn, Michel, et al., 2022). This temporal alignment of an exogenous acoustic signal and endogenous neural activity is called neural tracking (Obleser & Kayser, 2019). While infants are capable of neurally tracking pure tone musical stimuli (Cirelli et al., 2016), playback recordings (Attaheri et al., 2024), and ID speech (Menn, Michel, et al., 2022), it is yet unknown whether infants can neurally track the volume fluctuations of live singing. Investigating live singing is important because it is experienced differently compared to playback stimuli (Trehub, 2017). Live singing performances are more engaging than playback recordings (de l'Etoile, 2006), elicit more interpersonal arousal synchronisation (Kragness, Eitel, et al., 2023) and give more opportunities for attunement to infant feedback, such as by increasing positive caregiver affect (Markova et al., 2020; Trehub et al., 2016).

Neural tracking of live communication might be supported by body movements. While, as found in Study 1 of this dissertation, neural responses precede body movements regarding maturity, movement might still support neural tracking because rhythmic movements and passive vestibular stimulation guide auditory perception (Phillips-Silver & Trainor, 2005, 2007). Especially rhythmic movement coordination to musical stimuli could influence behaviour and neural activity in a way that they entrain to each other, that is, lock onto a common periodicity (Vuilleumier & Trost, 2015). The influence of body movements could expand our perspective on neural tracking of musical stimuli.

Rhythmic motor abilities are related to language abilities. Rhythmic movement teaches infants timing, which is required for speech . Herein lies the connection between language acquisition and music via rhythmic movements. On the level of acoustic qualities, music exposure during infancy and preschool seems to be beneficial for language acquisition (e.g., Iverson, 2010; Iverson et al., 2007; Papadimitriou et al., 2021; Putkinen et al., 2013; Zhao & Kuhl, 2016). Especially the musical aspects of language (timbre, rhythm, prosody) are thought to support the development of semantic and syntactic understanding and production (Brandt et al., 2012). On a neural level, neural tracking of nursery rhymes positively correlates with later language outcomes (Attaheri et al., 2024; Menn, Ward, et al., 2022). Instrumental music and ID speech elicit similar neural activity in newborns (Kotilahti

et al., 2010), so spoken language could even be viewed as a form of music from a developmental perspective (Brandt et al., 2012).

This relationship between music processing on a neural level, rhythmic movement responses and language development could shed new light on the interplay between these three domains. Investigating the neural and movement responses to live ID singing, as well as their connection to language development, could have important implications for understanding the possibility of using live musical communication to facilitate infants' communicative development.

Study 2 investigates three effects of live and dynamic ID singing: 1) whether infants show neural tracking of live and dynamic ID singing, 2) the association between song type and rhythmic movement, and 3) the contributions of neural tracking and rhythmic movement to ID singing, respectively, to language development. These questions were addressed in two groups of 7-month-old infants and their mothers, who sang eight verses of a well-known playsong and lullaby to their infants in a counterbalanced order and semi-naturalistic face-to-face context. In the first group, neural tracking was measured via EEG ($N = 30$). These infants sat in an infant car seat to reduce movement artefacts. A temporal response function (TRF) analysis was applied to the EEG data to determine how much the sound envelope of ID singing could predict infant neural activity. Higher encoding accuracy happened when the infant's brain tracked singing more closely, which in turn meant that infant neural activity could be better predicted from the sound envelope of maternal ID singing. In the second group, rhythmic movement was assessed in infants ($N = 40$) seated in a highchair to allow for less restricted movement. Rhythmic movement was manually coded after Thelen (Thelen, 1979): at least three repetitions of the same movements with a maximum of one second between them. Language outcomes were assessed via expressive vocabulary size in 20-month-olds of both testing groups ($N = 60$). The acoustic features pitch (fundamental frequency), amplitude (perceived loudness), rhythmicity (pulse clarity), and tempo (bpm) were extracted from maternal singing via the MIRtoolbox in Matlab (Lartillot et al., 2008).

It was hypothesised that infant neural responses (i.e., neural tracking measured via TRF) would be more coordinated with the slower, soothing lullabies than with the more complex playsongs based on previous findings that neural tracking is enhanced by melodic expectations and in response to slower music (Di Liberto et al., 2020; Weineck et al., 2022). Furthermore, infants were expected to move more rhythmically to playsongs than to lullabies due to playsongs' higher expected rhythmic complexity, which has been reported to increase

movement and the inclination to move in children and adults (Cameron et al., 2023; Kragness, Anderson, et al., 2023). Additionally, it was hypothesised that better neural tracking and more rhythmic movements at 7 months would predict a larger vocabulary size at 20 months of age because neural tracking of nursery rhymes has previously been shown to be positively correlated to later language outcomes (Attaheri et al., 2024; Menn, Ward, et al., 2022).

The present study showed that infants 1) were able to neurally track music, and that lullabies were easier to track than playsongs, 2) that infants rhythmically moved more to playsongs than to lullabies, and 3) that vocabulary size positively correlated with better neural tracking and more rhythmic movement in playsongs alone. Some acoustic parameters influenced neural tracking. Neural tracking was enhanced with less tempo variation. Furthermore, playsongs were easier to neurally track with lower perceived loudness and loudness variation, and higher pulse clarity. Lullabies were more accurate to neurally track with a slower tempo and lower pitch variability.

This study highlights the multivariate effects of live and dynamic ID singing on infant neural activity and behaviour, and its relation to language acquisition. The results show that infants can neurally track live ID musical communication beyond pure tones (Cirelli et al., 2016), playback recordings (Attaheri et al., 2024), and ID speech (Menn, Michel, et al., 2022), which could be connected to attentional processes. The slower, more regular rhythmicity and higher tempo and pitch predictability of lullabies might help neural tracking. Similarly, mothers might have counterbalanced the complexity of playsongs by highlighting their rhythmic structure (Zentner & Eerola, 2010) and lowering loudness (Calignano et al., 2021). Playsongs' higher rhythmic variability, on the other hand, might elicit more rhythmic movement and help language learning by highlighting prosodic information (Hannon & Johnson, 2005). These findings corroborate the assumption that ID singing has inherent acoustic characteristics that help infants coordinate to interactive rhythms (Markova et al., 2019, 2020), and that certain acoustic features in ID musical communication might be linked to certain caregiving goals (Cirelli et al., 2020; Rock et al., 1999; Trainor, 1996).

Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313

2.3. Study 3. The Reciprocal Relationship between Maternal Infant-directed Singing and Infant Gaze

Previous literature has investigated the acoustic qualities of ID singing and infant reactions to ID singing separately. However, ID singing is not a one-way performance, but an interactive form of communication, so caregivers and infants likely adjust to each other during ID singing interactions on multiple modalities and potentially across modalities. Such adjustments may even be needed to reach song function. During an ID singing performance, the caregiver's primary modality is the acoustic domain. As described in Section 1.2.2, previous studies have shown that infants are sensitive to rhythm and pitch changes in musical stimuli. Caregivers choose different types of ID songs with distinct acoustic qualities to achieve different functions: Playsongs to engage infants and direct their attention, and lullabies to soothe infants and divert their attention (Cirelli et al., 2020; Rock et al., 1999; Trehub, 2018; Trehub & Trainor, 1998). Furthermore, caregivers seem to adjust the acoustic variability to their communicative intent, at least in ID speech. Caregivers of premature newborns speak less acoustically variable to awake infants, and more acoustically variable to asleep infants (Saliba et al., 2020). This has been interpreted as minimising stimulation for awake infants with less acoustic variability, and engaging sleeping infants with more acoustically variable speech. Caregivers seem to adjust the acoustic qualities of their communication to their goals (Nguyen et al., 2023), and it stands to reason that they should be sensitive to the efficacy of their performance. Preverbal infants cannot use words to communicate their needs, but visual attention is a clear communication tool that caregivers can register and respond to. However, no study to date has investigated how a change in acoustic variability modulates infant attention specifically in an ID singing context, and how this attention, in turn, affects acoustic variability in caregiver ID singing.

Study 3 expands Study 2 of this dissertation by testing whether infants are sensitive to acoustic variability in caregiver ID singing, whether caregivers respond to infant attention by adjusting the acoustic variability of their singing to fit the song function, and whether those changes co-occur. Seventy-four mothers sang eight verses of a well-known playsong and a lullaby to their 7-month-old infants. The acoustic variability of maternal singing was measured via spectral flux, an acoustic variable describing amplitude changes in the frequencies comprising a spectrum (Müller, 2015) and measuring acoustic novelty (Müller & Chiu, 2024). Infant attention was measured via social gaze towards the mother's face for longer than one second.

It was hypothesised that maternal singing would show increased acoustic variability (i.e., spectral flux) around times of infant visual attention (i.e., social gaze), either to attract infant attention or as a response to it. Furthermore, it was hypothesised that playsongs would contain higher spectral flux and elicit more social gaze than lullabies because of the attention-attracting nature of playsongs (Cirelli et al., 2020).

The frequency and duration (absolute and relative to song length) of infant gaze towards the mother's face (henceforth referred to as infant social gaze) were manually coded. The following acoustic qualities of maternal singing were analysed with custom scripts: spectral flux, amplitude envelope, and pitch via F0 extraction with the Python packages numpy and parselmouth (Harris et al., 2020; Jadoul et al., 2018; Python Software Foundation, 2023), and tempo via bpm with the MIRtoolbox in Matlab (Lartillot et al., 2008). First, overall condition differences of infant social gaze and maternal acoustic qualities were tested to see if mothers and infants differentiated between playsongs and lullabies. Furthermore, the correlation and temporal relation between infant social gaze and mean spectral flux was tested between playsongs and lullabies.

Spectral flux was calculated as the sum over frequency (from 0 to 3000 Hz) of the absolute difference in amplitude between consecutive spectra at a temporal resolution of 10 ms. Singing and speech contain most energy in the chosen frequency band (Lindblom & Sundberg, 2007). To assess the changes in acoustic variability around infant social gaze onset, spectral flux around those moments was compared to surrogate spectral flux at random moments without infant social gaze (N = 1000 per ID and condition) in a permutation analysis (N_{perm} = 1000, p-values below the 5th/2 percentile). The p-values were Bonferroni corrected to the $5^{th}/2$ percentile because of the two singing conditions. Supplementary permutation analyses were conducted for amplitude envelope, pitch (F0), and spectral flux around infant social gaze offset.

This study revealed two main results. First, infants showed more social gaze towards their mothers during playsongs than during lullabies, and mothers sang playsongs more acoustically variable (i.e., with higher spectral flux) than lullabies. Second, changes in maternal spectral flux were present both before and after infant social gaze in both song types. However, spectral flux peaks were especially pronounced as a response to infant social gaze in playsongs. Further manipulation checks showed that mothers sang playsongs longer, faster, louder, and with higher spectral flux than lullabies, though there was no pitch

difference between conditions. Mean spectral flux also positively correlated with the absolute duration of infant social gaze.

The results of Study 3 suggest that mothers and infants reciprocally adjust to each other during an ID singing interaction, and that this adjustment is influenced by song type. By increasing their spectral flux during playsongs, mothers may have tried to elicit and maintain infant attention, which was reflected in the especially pronounced spectral flux increases after infant social gaze onset, and met with longer and more frequent infant social gaze. These results are consistent with research on the attention-getting nature of playsongs (Cirelli et al., 2020; Rock et al., 1999). Furthermore, playsongs' acoustic variability could have induced uncertainty, which might have been attention-eliciting in itself (Kidd et al., 2014; Poli et al., 2020). Mothers could have sung lullabies more predictably to fulfil their soothing function (Nguyen et al., 2023; Schwartz, 2004), consistent with research showing that infants shift their focus inwards during lullabies (Rock et al., 1999). Its live interactional setting is an important aspect of this study, since social presence elicits more pronounced responses (Kragness, Eitel, et al., 2023; Rochat, 2001) and could induce a feedback loop between an actively participating infant (Beebe et al., 2016) and a contingently responding mother, enhancing the communicative and regulatory effects of ID singing.

Reisner, S., Nguyen, T., Labendzki, P., Hoehl, S., & Markova, G. (2025). The reciprocal relationship between maternal infant-directed singing and infant gaze. *Musicae Scientiae, 0*(0). https://doi.org/10.1177/10298649251385676

3. General Discussion

The present thesis aims to provide evidence on how infants process and react to infant-directed music and live singing across multiple ages in the first postnatal year of life across multiple modalities (neural, movement, and attentional responses, and language outcomes), and on how caregivers and infants engage in the context of ID singing. Across three studies, I demonstrate that infants are sensitive to musical stimuli, including live ID singing, from early on, how they neurally process and behaviourally react to musical stimuli, how these responses are influenced by acoustic properties of the musical input and how they mature over the first postnatal year of life, and that live ID singing creates opportunities to interact across modalities.

Study 1 is, to date, the only study to link neural and motor processing of musical stimuli in infants. It demonstrated that 3-, 6-, and 12-month-old infants are neurally sensitive to musical structure, that neural and movement responses become more adult-like and sophisticated over the first year of postnatal life, though no movement synchronisation was found. High pitch was neurally processed better than low pitch only by 6-month-olds, but elicited more auditory-movement coupling in 12-month-olds.

Study 2 was the first study to investigate neural tracking of live ID singing and link rhythmic movement responses to it in 7-month-olds. It demonstrated neural tracking of live singing, with better neural tracking of lullabies, more rhythmic movement responses to playsongs, and positive correlations between vocabulary size at 20 months to neural tracking of playsongs and rhythmic movement to playsongs at 7 months, respectively.

Study 3 demonstrated how infants and caregivers interact on multiple modalities during live ID singing. It showed that infant attention elicited maternal variability in ID singing especially in playsongs, not in lullabies, and that infants paid more attention to playsongs than to lullabies.

3.1. Expansion of Existing Literature on ID Songs

The studies compiled in this dissertation contribute to and expand the existing literature on ID music and singing, using diverse investigation approaches (neural: ERPs, mTRF; movement: markerless motion tracking, manual microcoding; time-series analysis of interactions; questionnaires) across different modalities (neural responses, movement, acoustic qualities, attention, language outcomes), and across passive playback listening and active live interactions.

3.1.1. Neural Responses to ID Songs: Sensitivity to Musical Structure, Influence of Acoustic Qualities, Maturation of Responses over Time

Studies 1 and 2 demonstrated that 1) infants are sensitive to musical structure (i.e., original vs shuffled songs), 2) infants can neurally track live singing, 3) the acoustic qualities of ID songs impact how well infants can neurally process them, and 4) neural responses to musical stimuli mature over the first postnatal year of life.

Study 1 showed that, indeed, musical structure is important, not purely the individual tones making up a song: Infants as young as 3 months show distinct ERPs to playsong-style musical stimuli and have stronger responses to musical notes that are part of an intact musical structure compared to scrambled notes. Study 2 expanded these results by showing that infants are able to neurally track live singing beyond pure tones (Cirelli et al., 2016), playback recordings (Attaheri et al., 2024), and ID speech (Menn, Michel, et al., 2022).

Interestingly, pitch influenced neural processing only in 6-month-olds: Higher pitch elicited stronger ERPs than lower pitch in Study 1. This high pitch superiority (Marie & Trainor, 2013, 2014) disappeared again by 12 months of age, and furthermore could not be demonstrated in Study 2, where playsongs and lullabies lacked pitch differences, potentially because they were sung in close succession. However, lullabies were still easier to neurally track than playsongs for 7-month-olds. Hence, other acoustic qualities besides pitch impacted neural processing:

Decreased loudness, lower volume fluctuations, and higher rhythmicity enhanced neural tracking accuracy only in playsongs, while slower tempo and lower pitch variability enhanced it only in lullabies. These results lead to the conclusion that, depending on the song type (playsong vs lullaby), a combination of different acoustic qualities aids neural processing of musical stimuli. Caregivers could even intuitively adjust how they present their songs. This idea is corroborated by Study 3, which demonstrated that caregivers adjust their acoustic variability in ID singing depending on whether they sing a playsong or a lullaby. This difference in acoustic variability also affected infant attention, which will be discussed in Section 3.1.4.

The idea that caregivers adjust the acoustic qualities of their singing performance to fit the purpose of what they want to convey also fits into an active learning framework, wherein caregivers present information to their child in an age-appropriate manner, balancing repetition and complexity, with information complexity increasing with age (Tal et al., 2024). While lexical information complexity was fairly fixed in our paradigm (Studies 2 and 3),

since the choice of song was predefined, mothers were still free to choose how to perform each of those songs – they may have counterbalanced their lack of lexical freedom by varying their performance of the playsong and the lullaby in the musical domain. Further underlining the potential of this active learning framework, infants' vocal (Elmlinger et al., 2019, 2023) and attentional signals (Study 3) have been shown to influence how adults communicate with them, indicating the active role that infants play in modulating adults' acoustic signals.

3.1.2. Rhythmic Movement Extraction

Studies 1 and 2 both used rhythmic movement as an outcome measure to ID songs, but different analysis approaches were used. While Study 2 applied a comparably more coarse manual coding scheme (Thelen, 1979), Study 1 utilised a novel method to extract not only the Quantity of Movement (comparable to the manual coding in Study 2), but also precise movement patterns (Principal Movements; Bigand et al., 2024) and movement-beat synchronisation. This was the first time such an elaborate movement analysis had been performed on infant data. The methods of both Studies 1 and 2 provided a similar birds-eye view of infant movements: Infants rhythmically move to musical stimuli. However, when we zoom in on the results, the picture becomes more differentiated: At 7 months of age, infants move more to playsongs than to lullabies (Study 2), but only by 12 months do they move more to musical structure compared to scrambled sounds (Study 1). There are several possible explanations for this: 1) Infants could be sensitive to subtle differences in the acoustic qualities of songs before they are as sensitive to the presence or absence of musical structure. 2) The live ID singing context differs from a playback experiment in terms of infant behavioural outcomes. Live singing performances are phenomenologically different from and more engaging than playback recordings (de l'Etoile, 2006; Trehub, 2017). The presence of additional modalities displayed by mothers (e.g., facial expressions, eye contact, affect, body sway, etc.) could have made the difference in eliciting infant rhythmic movement responses between playsongs and lullabies. Those modalities were all absent in the recorded stimuli of Study 1. 3) Song familiarity in Study 2 encouraged more movement in comparison to the unfamiliar songs of Study 1, and therefore, more nuanced movement responses were possible in Study 2 depending on the song type. Song familiarity has a variety of documented effects, including eliciting positive emotions, movements, and attentive listening in infants (Dou & Cirelli, 2025), being preferred by infants (Kragness, Ullah, et al., 2022), more effectively reducing infant distress (Cirelli & Trehub, 2020), and increased proximity-seeking to and

prosociality towards the singer (Cirelli & Trehub, 2018). 5) Playsongs' activating effects have been well documented (Cirelli et al., 2020; Rock et al., 1999; Trainor, 1996; Trainor et al., 1997). Therefore, it is not surprising that we found higher rhythmic movements to playsongs compared to lullabies in Study 2. Study 1 also utilised playsongs, and the scrambled condition may still have been more similar to playsongs than to the more calming lullabies. It is also important to note that, while both Studies 1 and 2 examined the Quantity of Movement, the level of fine-grained analysis differed between the two studies: Study 1's sophisticated automated analysis could have been much more thorough than Study 2's manual coding.

Examining the influence of pitch on infant movement responses, the results of Studies 1 and 2 present an equally complex picture: Pitch differences did not influence the amount of movement (Studies 1 and 2), but did influence movement velocity, with high-pitched music eliciting higher movement velocity compared to low-pitched music (Study 1). Infants did, however, move more in response to playsongs than to lullabies, despite no examined acoustic feature (i.e., pitch, tempo, perceived loudness, perceived rhythm) correlating with this result (Study 2). It seems that rather than low pitch being the single driver for increased infant movements, like it is in adults (Cameron et al., 2022; Stupacher et al., 2013, 2016; Van Dyck et al., 2013), other factors could make playsongs the superior choice for eliciting more infant movement: Maybe it is a combination of acoustic features, or infants' previous association of playsongs with playful, engaging situations that trigger increased movement. When and how the adult preference for low pitches develops remains to be discovered.

The finding in Study 1 that infants showed greater movement velocity to high-pitched compared to low-pitched music from 3 to 12 months of age adds another layer of complexity: As mentioned above in Section 2.1, it remains challenging to explain these results. One possibility is that infants' neural high voice superiority (Marie & Trainor, 2013, 2014) also manifests in the behavioural domain across a wider developmental window: A predisposition to better neurally process higher pitches could have led infants to show higher movement velocity in response to those higher pitches. This would, however, only explain movement responses at a time when high pitches elicit enhanced neural responses at around 6 months of age.

3.1.3. Language Outcomes

Only Study 2 investigated the correlation between live ID singing and infant language outcomes, finding a positive correlation between neural and movement responses to playsongs at 7 months and vocabulary size at 20 months. While the positive correlation between music exposure and language outcomes has been shown before (Boyne et al., 2025; Franco et al., 2022; Papadimitriou et al., 2021), Study 2 was the first to show a correlation with song-specific neural processing and movement responses. The level of neural and movement engagement with musical stimuli may be directly related to infants' engagement with language: Neural tracking might be connected to the extraction of prosody, which supports word segmentation (Menn, Michel, et al., 2022) and is especially pronounced in ID singing (Nakata & Trehub, 2011; Suppanen et al., 2019; Trainor et al., 1997).

It is important to note that these results do not imply causality. Infants who already have a better affinity for perceiving rhythmic structures may have an advantage in language learning. However, conclusions on whether these individual differences persist beyond the early toddler years cannot be drawn from the current evidence. Studies on clinical NICU populations do not support this idea, though: While singing is recommended to support preterm-born infants' development of auditory processing by term-age (Kostilainen et al., 2021), music therapy during the neonatal period does not seem to have a long-term effect on language outcomes at 2-3 years of age (Kostilainen et al., 2023).

3.1.4. Attentional Responses and Interactional Effects

Study 3 expanded on existing literature on infant attentional effects of ID songs by providing a nuanced attentional and interactional perspective. While previous studies had already shown that playsongs seem to have an attention-grabbing effect, and lullabies seem to be more attention-diffusing (Cirelli et al., 2020; Rock et al., 1999; Trehub, 2018; Trehub & Trainor, 1998), Study 3 of this dissertation expands current knowledge with the interactional responses following infant attention. Not only do infants pay attention differently to playsongs and lullabies, but caregivers also contingently react to this attention depending on song type – keeping a stable performance in lullabies, while introducing more acoustic variability in playsongs, which results in a more unpredictable and engaging interaction. This fits previous literature of how playsongs prompt infants to shift their attention more outward (Rock et al., 1999) and connects back to the increased movement responses to playsongs in Study 2: Maybe acoustic variability was a driver of the greater rhythmic movement responses to playsongs?

3.2. In context

3.2.1. Maturation of Infant Responses: Neural Responses were Observable before Movement Responses, Timeline of Attentional Development Still Unclear

Whether movements supported neural processing of ID songs is difficult to say. In Study 2, we postulated that due to playsongs eliciting both stronger neural and movement responses, movements may have aided neural processing. However, Study 1 showed that neural responses preceded sophisticated motor responses, which could be linked to the maturation of the motor system (for a review, see Adolph & Franchak, 2017). The first year of postnatal life seems to be a suitable time period to observe the maturation of neural and simple motor responses to musical stimuli, but it apparently only captures the beginnings of complex motor response maturation.

While movement in response to music has been observed even in the fetal stage (Kisilevsky et al., 2004) and some abilities like tempo flexibility are present from as early as 5 months of age (Zentner & Eerola, 2010), it is important to note that previous studies did not control for whether infants moved in response to music specifically or to auditory input in general. As the results of Studies 1 and 2 suggest, a preference for musical structure and distinguished responses to different song types may develop within the first postnatal year of life. Rather than movements aiding neural processing, neural responses appear to be a prerequisite for movement responses.

Attentional responses to ID songs are likely present from the neonatal period (Epstein et al., 2025; Palazzi et al., 2021), but given current evidence, we cannot say from which time point onward we can observe distinctions in infant attention between playsongs and lullabies. Culture and song familiarity could play a major role here. Since mothers in Study 3 sang familiar ID songs, which infants are known to respond to more strongly than to unfamiliar ones (Kragness, Johnson, et al., 2022; Mehr et al., 2016), it is likely that infants may learn the purposes of familiar playsongs and lullabies and have already learned to respond accordingly earlier than at 7 months of age.

3.2.2. Engagement with Both Recorded and Live Stimuli

Infant engagement with both recorded and live musical stimuli was shown across the three studies of this dissertation. This validates the use of recorded stimuli for studies on musical engagement and shows that recorded and live stimuli have distinct applications in research: recorded stimuli can provide insight into which acoustic qualities drive specific

infant responses. Live interactions need to be studied as multimodal activities: Infants might engage more readily in live, dynamic interactions, especially when caregivers may intuitively adjust their singing to subtle displays of infant behaviour. Our results suggest that infant engagement is related to caregiving goals. While live singing, as well as infant responses to it, differed between song types even in a lab setting (Studies 2 and 3), in the future, the multimodality of such interactions needs to be studied in an even more ecologically valid way in infants' everyday lives, to further draw conclusion on how the different modalities of ID singing interactions, infant engagement, caregiving goals, and developmental outcomes intertwine.

3.3. Limitations and Open Questions

The studies in this thesis have limitations. While Study 1 is arguably the least naturalistic of the three studies, it is the most controlled. The recorded songs allowed for fine-grained, well-controlled stimulus manipulation, but were not as richly multimodal as the live songs in Studies 2 and 3. The distinction of ID singing into lullabies and playsongs in Studies 2 and 3 was based on previous literature making those clear distinctions (e.g., Trehub & Trainor, 1998). While those two categories are well established in research on ID songs, they might not strictly reflect what infants actually listen to in daily life. A first comprehensive study of the musical soundscape of U.S. American infants found that a third of daily music exposure consisted of exclusively recorded stimuli, and a quarter was sung live (Mendoza & Fausey, 2021). Recorded stimuli might fit a clearer distinction into playsongs and lullabies than live-sung songs, especially when the latter are improvised. Future research is needed on the validity of the lullaby-playsong dichotomy. In collaboration with Prof. Stefanie Höhl and Dr Gabriela Markova, I am working on data collected in the home environment with wearable devices to map out the home ID singing environment. This project also includes fathers and therefore aims to close the knowledge gap of current research regarding father ID singing. Currently, very few studies have included fathers as participants in ID singing studies (but see Trehub et al., 1997). It would be insightful to better understand the roles of all active caregivers.

Another open question regarding ID singing is its physiological effect. Current empirical research is sparse and yields conflicting results: While playsongs and lullabies are often said to be arousing and calming, respectively, current research either cannot corroborate those claims or highlights the importance of the infant's initial state. One study showed that caregivers of upset infants tend to calm them with playsongs (Cirelli & Trehub, 2020), while

another study shows that calm infants show a decrease in physiological arousal in response to lullabies, while playsongs have no effect (Cirelli et al., 2020). To add to this sparse literature on the physiological effects of ID singing, I am working on a project expanding on the data analysis from Studies 2 and 3 to determine the physiological effects of live playsongs and lullabies on calm infants.

Furthermore, it would be interesting to investigate how infants and caregivers make music together. While infants do not have the musical skills to participate in musical interactions as actively as adults do, Study 3 showed that they do so through their attentional behaviour, similar to what they do in other social contexts (for a review, see Begus & Southgate, 2018). One could investigate the influence of other infant behavioural modalities, such as body movements, clapping, vocalisations, or affect, to get a more complete picture of how infants influence their caregivers during musical interactions.

Lastly, the long-term developmental outcomes of ID singing have often only been implied, correlated, or studied through questionnaires. Study 2 of this dissertation, for example, has shown a positive correlation between vocabulary size (via questionnaire data) and neural and movement responses to playsongs. This is a first step in investigating the mechanisms that could drive the impact of ID singing on later language development. However, whether and how caregivers could improve their infants' neurobehavioural processing of music and boost their language learning by singing to them more often, or whether the acoustic qualities matter, is still to be determined. Initial evidence from preterm-born infants has found a positive short-term effect of ID singing on infants' language learning by term age (Kostilainen et al., 2021), which disappeared by 2-3 years of age (Kostilainen et al., 2023). However, music interventions in 4- to 6-year-old children have been found to affect language skills at the word and phoneme levels (Patscheke et al., 2016). There may be specific time windows in development in which music exposure could boost language learning in the short term. The results of Study 2 suggest that the developmental period of high voice superiority warrants further investigation into whether music exposure aids language learning.

It is furthermore possible that any positive effects of ID singing on language outcomes are supported by the social interaction itself and by caregiver responsiveness (for a review, see Smith & Kong, 2025). Whether musical interactions furthermore support caregiver-infant bonding remains to be systematically investigated.

4. Conclusion

In this thesis, I have presented three studies on how infants perceive and react to musical stimuli, both in recorded form and in live interactions, and on how infants and caregivers engage with each other across multiple modalities during live interactions. I have presented evidence that the neural sensitivity and behavioural responses to musical stimuli are modified by the musical structure and acoustic qualities of the musical input, that those perceptions and actions mature over the first postnatal year of life, and that live caregiver singing is interactional across modalities. Taken together, these studies highlight the importance of investigating the impact of infant-directed songs across multiple modalities in controlled, yet complex passive listening and face-to-face interactive paradigms.

5. References

Adolph, K. E., & Franchak, J. M. (2017). The development of motor behavior. *WIREs Cognitive Science*, *8*(1–2), e1430. https://doi.org/10.1002/wcs.1430

Attaheri, A., Ní Choisdealbha, Á., Rocha, S., Brusini, P., Di Liberto, G. M., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Alfaro E Oliveira, M., Brough, C., Flanagan, S., & Goswami, U. (2024). Infant low-frequency EEG cortical power, cortical tracking and phase-amplitude coupling predicts language a year later. *PLOS ONE*, *19*(12), e0313274. https://doi.org/10.1371/journal.pone.0313274

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother–infant face-to-face communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085

Begus, K., & Southgate, V. (2018). Curious Learners: How Infants' Motivation to Learn Shapes and Is Shaped by Infants' Interactions with the Social World. In M. M. Saylor & P. A. Ganea (Eds), *Active Learning from Infancy to Childhood: Social Motivation, Cognition, and Linguistic Mechanisms* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-77182-3_2

Bergeson, T. R., & Trehub, S. E. (2002). Absolute Pitch and Tempo in Mothers' Songs to Infants. *Psychological Science*, *13*(1), 72–75. https://doi.org/10.1111/1467-9280.00413

Bianco, R., Tóth, B., Bigand, F., Nguyen, T., Sziller, I., Háden, G. P., Winkler, I., & Novembre, G. (2025). *Human newborns form musical predictions based on rhythmic but not melodic structure* (p. 2025.02.19.639016). bioRxiv. https://doi.org/10.1101/2025.02.19.639016

Bigand, F., Bianco, R., Abalde, S. F., & Novembre, G. (2024). The geometry of interpersonal synchrony in human dance. *Current Biology*, *34*(13), 3011-3019.e4. https://doi.org/10.1016/j.cub.2024.05.055

Boyne, A. S., Alviar, C., & Lense, M. (2025). Parental Social and Musical Characteristics, the Home Music Environment, and Child Language Development in Infancy. *Infancy*, *30*(2), e70008. https://doi.org/10.1111/infa.70008

Braithwaite, E. K., Jones, E. J. H., Johnson, M. H., & Holmboe, K. (2020). Dynamic modulation of frontal theta power predicts cognitive ability in infancy. *Developmental Cognitive Neuroscience*, *45*, 100818. https://doi.org/10.1016/j.dcn.2020.100818

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and Early Language Acquisition. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00327

Brown, S. (2022). Group dancing as the evolutionary origin of rhythmic entrainment in humans. *New Ideas in Psychology*, *64*, 100902. https://doi.org/10.1016/j.newideapsych.2021.100902

Calignano, G., Dispaldro, M., Russo, S., & Valenza, E. (2021). Attentional engagement during syllable discrimination: The role of salient prosodic cues in 6- to 8-month-old infants. *Infant Behavior and Development*, *62*, 101504. https://doi.org/10.1016/j.infbeh.2020.101504

Cameron, D. J., Caldarone, N., Psaris, M., Carrillo, C., & Trainor, L. J. (2023). The complexity-aesthetics relationship for musical rhythm is more fixed than flexible: Evidence from children and expert dancers. *Developmental Science*, *26*(5), e13360. https://doi.org/10.1111/desc.13360

Cameron, D. J., Dotov, D., Flaten, E., Bosnyak, D., Hove, M. J., & Trainor, L. J. (2022). Undetectable very-low frequency sound increases dancing at a live concert. *Current Biology*, *32*(21), R1222–R1223. https://doi.org/10.1016/j.cub.2022.09.035

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., & Gross, J. (2011). Steady-State Visual Evoked Potentials Can Be Explained by Temporal Superposition of Transient Event-Related Responses. *PLoS ONE*, *6*(1), e14543. https://doi.org/10.1371/journal.pone.0014543

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *European Journal of Neuroscience*, *21*(11), 3201–3204. https://doi.org/10.1111/j.1460-9568.2005.04144.x

Chen, A., Peter, V., & Burnham, D. (2016). Auditory ERP response to successive stimuli in infancy. *PeerJ*, *4*, e1580. https://doi.org/10.7717/peerj.1580

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of Maternal Singing Style on Mother–Infant Arousal and Behavior. *Journal of Cognitive Neuroscience*, *32*(7), 1213–1220. https://doi.org/10.1162/jocn_a_01402

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring Neural Entrainment to Beat and Meter in Infants: Effects of Music Background. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00229

Cirelli, L. K., & Trehub, S. E. (2018). Infants help singers of familiar songs. *Music & Science*, *1*, 2059204318761622. https://doi.org/10.1177/2059204318761622

Cirelli, L. K., & Trehub, S. E. (2020). Familiar Songs Reduce Infant Distress. *Developmental Psychology*, 8.

Conrad, N. J., Walsh, J., Allen, J. M., & Tsang, C. D. (2011). Examining infants' preferences for tempo in lullabies and playsongs. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *65*(3), 168–172. https://doi.org/10.1037/a0023296

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488. https://doi.org/10.1016/S0163-6383(97)90037-0

Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-Directed Speech in the First Month after Birth. *Child Development*, *61*(5), 1584. https://doi.org/10.2307/1130766

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142–158. https://doi.org/10.1111/desc.12100

Corrigall, K. A., & Trainor, L. J. (2019). Electrophysiological Correlates of Key and Harmony Processing in 3-year-old Children. *Music Perception*, *36*(5), 435–447. https://doi.org/10.1525/mp.2019.36.5.435

Damsma, A., de Roo, M., Doelling, K., Bazin, P.-L., & Bouwer, F. L. (2025). Tempo-dependent selective enhancement of neural responses at the beat frequency can be mimicked by both an oscillator and an evoked model. *bioRxiv*, 2024.07.11.603023. https://doi.org/10.1101/2024.07.11.603023

de l'Etoile, S. K. (2006). Infant-Directed Singing: A Theory for Clinical Intervention. *Music Therapy Perspectives*, *24*(1), 22–29. https://doi.org/10.1093/mtp/24.1.22

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*(5604), 718–719. https://doi.org/10.1038/266718a0

Di Liberto, G. M., Pelofi, C., Bianco, R., Patel, P., Mehta, A. D., Herrero, J. L., De Cheveigné, A., Shamma, S., & Mesgarani, N. (2020). Cortical encoding of melodic

expectations in human temporal cortex. *eLife*, *9*, e51784. https://doi.org/10.7554/eLife.51784

Dou, A., & Cirelli, L. K. (2025). The Active Infant's Developing Role in Musical Interactions: Insights From an Online Parent Questionnaire. *Infancy*, *30*(1), e12648. https://doi.org/10.1111/infa.12648

Eckerdal, P., & Merker, B. (2008). 'Music' and the 'action song' in infant development: An interpretation. In S. Malloch & C. Trevarthen (Eds), *Communicative Musicality* (pp. 241–262). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198566281.003.0011

Edalati, M., Wallois, F., Safaie, J., Ghostine, G., Kongolo, G., Trainor, L. J., & Moghimi, S. (2023). Rhythm in the Premature Neonate Brain: Very Early Processing of Auditory Beat and Meter. *The Journal of Neuroscience*, *43*(15), 2794–2802. https://doi.org/10.1523/JNEUROSCI.1100-22.2023

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*(5), 998–1011. https://doi.org/10.1017/S0305000919000291

Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2023). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science*, *26*(2), e13296. https://doi.org/10.1111/desc.13296

Epstein, S., Arnon, S., Markova, G., Nguyen, T., Hoehl, S., Eitan, L., Bauer-Rusek, S., Yakobson, D., & Gold, C. (2025). Mother–Preterm Infant Contingent Interactions During Supported Infant-Directed Singing in the NICU—A Feasibility Study. *Children*, *12*(9), 1273. https://doi.org/10.3390/children12091273

Feldman, R. (2007). Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345. https://doi.org/10.1111/j.1467-8721.2007.00532.x

Fernald, A., & Simon, T. (1984). Expanded Intonation Contours in Mothers' Speech to Newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140091. https://doi.org/10.1098/rstb.2014.0091

Fraisse, P. (1982). Rhythm and tempo. *The Psychology of Music.*, 149–180.

Franco, F., Suttora, C., Spinelli, M., Kozar, I., & Fasolo, M. (2022). Singing to infants matters: Early singing interactions affect musical preferences and facilitate vocabulary building. *Journal of Child Language*, *49*(3), 552–577. https://doi.org/10.1017/S0305000921000167

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of Dancing and Singing to Music in Three- to Four-Months-Old Infants. *PLoS ONE*, *9*(5), e97680. https://doi.org/10.1371/journal.pone.0097680

Gerry, D., Unrau, A., & Trainor, L. J. (2012). Active music classes in infancy enhance musical, communicative and social development. *Developmental Science*, *15*(3), 398–407. https://doi.org/10.1111/j.1467-7687.2012.01142.x

Hadders-Algra, M. (2005). Development of Postural Control During the First 18 Months of Life. *Neural Plasticity*, *12*(2–3), 695071. https://doi.org/10.1155/NP.2005.99

Háden, G. P., Bouwer, F. L., Honing, H., & Winkler, I. (2024). Beat processing in newborn infants cannot be explained by statistical learning based on transition probabilities. *Cognition*, *243*, 105670. https://doi.org/10.1016/j.cognition.2023.105670

Háden, G. P., Honing, H., Török, M., & Winkler, I. (2015). Detecting the temporal structure of sound sequences in newborn infants. *International Journal of Psychophysiology*, *96*(1), 23–28. https://doi.org/10.1016/j.ijpsycho.2015.02.024

Háden, G. P., Németh, R., Török, M., & Winkler, I. (2015). Predictive processing of pitch trends in newborn infants. *Brain Research*, *1626*, 14–20. https://doi.org/10.1016/j.brainres.2015.02.048

Háden, G. P., Stefanics, G., Vestergaard, M. D., Denham, S. L., Sziller, I., & Winkler, I. (2009). Timbre-independent extraction of pitch in newborn infants. *Psychophysiology*, *46*(1), 69–74. https://doi.org/10.1111/j.1469-8986.2008.00749.x

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology*, *50*(4), 354–377. https://doi.org/10.1016/j.cogpsych.2004.09.003

Hannon, E. E., & Trehub, S. E. (2005a). Metrical Categories in Infancy and Adulthood. *Psychological Science*, *16*(1), 48–55. https://doi.org/10.1111/j.0956-7976.2005.00779.x

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639–12643. https://doi.org/10.1073/pnas.0504254102

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hewlett, B. S., & Lamb, M. E. (2002). Integrating evolution, culture and developmental psychology: Explaining caregiver–infant proximity and responsiveness in central Africa and the USA. In H. Keller, Y. H. Poortinga, & A. Schölmerich (Eds), *Between Culture and Biology* (1st edn, pp. 241–269). Cambridge University Press. https://doi.org/10.1017/CBO9780511489853.012

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Hinnekens, E., Barbu-Roth, M., Do, M.-C., Berret, B., & Teulier, C. (2023). Generating variability from motor primitives during infant locomotor development. *eLife*, *12*, e87463. https://doi.org/10.7554/eLife.87463

Honing, H. (2018). *The Origins of Musicality*. MIT Press.

Honing, H., Ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140088. https://doi.org/10.1098/rstb.2014.0088

Ilari, B. (2005). On musical parenting of young children: Musical beliefs and behaviors of mothers and infants. *Early Child Development and Care*, *175*(7–8), 647–660. https://doi.org/10.1080/0300443042000302573

Ilari, B. (2015). Rhythmic Engagement With Music in Early Childhood: A Replication and Extension. *Journal of Research in Music Education*, *62*(4), 332–343. https://doi.org/10.1177/0022429414555984

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 229–261. https://doi.org/10.1017/s0305000909990432

Iverson, J. M., Hall, A. J., Nickel, L., & Wozniak, R. H. (2007). The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language*, *101*(3), 198–207. https://doi.org/10.1016/j.bandl.2006.11.004

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A.-Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). Maturation of fetal responses to music. *Developmental Science*, *7*(5), 550–559. https://doi.org/10.1111/j.1467-7687.2004.00379.x

Kostilainen, K., Hugoson, P., Haavisto, A., Partanen, E., Mikkola, K., Huotilainen, M., Pakarinen, S., Furmark, C., Ådén, U., & Fellman, V. (2023). No impact of parental singing during the neonatal period on cognition in preterm-born children at 2–3 years. *Acta Paediatrica*, *112*(7), 1471–1477. https://doi.org/10.1111/apa.16788

Kostilainen, K., Partanen, E., Mikkola, K., Wikström, V., Pakarinen, S., Fellman, V., & Huotilainen, M. (2021). Repeated Parental Singing During Kangaroo Care Improved Neural Processing of Speech Sound Changes in Preterm Infants at Term Age. *Frontiers in Neuroscience*, *15*, 686027. https://doi.org/10.3389/fnins.2021.686027

Kotilahti, K., Nissilä, I., Näsi, T., Lipiäinen, L., Noponen, T., Meriläinen, P., Huotilainen, M., & Fellman, V. (2010). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping*, *31*(4), 595–603. https://doi.org/10.1002/hbm.20890

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1), 8537. https://doi.org/10.1038/ncomms9537

Kragness, H. E., Anderson, L., Chow, E., Schmuckler, M., & Cirelli, L. K. (2023). Musical groove shapes children's free dancing. *Developmental Science*, *26*(1), e13249. https://doi.org/10.1111/desc.13249

Kragness, H. E., Eitel, M. J., Anantharajan, F., Gaudette-Leblanc, A., Berezowska, B., & Cirelli, L. K. (2023). An itsy bitsy audience: Live performance facilitates infants' attention and heart rate synchronization. *Psychology of Aesthetics, Creativity, and the Arts*, *19*(5), 1124–1133. https://doi.org/10.1037/aca0000597

Kragness, H. E., Johnson, E. K., & Cirelli, L. K. (2022). The song, not the singer: Infants prefer to listen to familiar songs, regardless of singer identity. *Developmental Science*, *25*(1). https://doi.org/10.1111/desc.13149

Kragness, H. E., Ullah, F., Chan, E., Moses, R., & Cirelli, L. K. (2022). Tiny Dancers: Effects of Musical Familiarity and Tempo on Children's Free Dancing. *Developmental Psychology*, *58*(7), 1277–1285. https://doi.org/10.1037/dev0001363

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport*, *13*(1), 47–51. https://doi.org/10.1097/00001756-200201210-00014

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab Toolbox for Music Information Retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds), *Data Analysis, Machine Learning and Applications* (pp. 261–268). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_31

Lense, M. D., Shultz, S., Astésano, C., & Jones, W. (2022). Music of infant-directed singing entrains infants' social visual behavior. *Proceedings of the National Academy of Sciences*, *119*(45), e2116967119. https://doi.org/10.1073/pnas.2116967119

Lindblom, B., & Sundberg, J. (2007). The Human Voice in Speech and Singing. In T. D. Rossing (Ed.), *Springer Handbook of Acoustics* (pp. 669–712). Springer New York. https://doi.org/10.1007/978-0-387-30425-0_16

Longhi, E. (2009). `Songese': Maternal structuring of musical interaction with infants. *Psychology of Music*, *37*(2), 195–213. https://doi.org/10.1177/0305735608097042

Marie, C., & Trainor, L. J. (2013). Development of Simultaneous Pitch Encoding: Infants Show a High Voice Superiority Effect. *Cerebral Cortex*, *23*(3), 660–669. https://doi.org/10.1093/cercor/bhs050

Marie, C., & Trainor, L. J. (2014). Early development of polyphonic sound encoding and the high voice superiority effect. *Neuropsychologia*, *57*, 50–58. https://doi.org/10.1016/j.neuropsychologia.2014.02.023

Markova, G., Nguyen, T., & Hoehl, S. (2019). Neurobehavioral Interpersonal Synchrony in Early Development: The Role of Interactional Rhythms. *Frontiers in Psychology*, *10*, 2078. https://doi.org/10.3389/fpsyg.2019.02078

Markova, G., Nguyen, T., Schätz, C., & de Eccher, M. (2020). Singing in Tune – Being in Tune: Relationship Between Maternal Playful Singing and Interpersonal Synchrony: *Enfance*, *N° 1*(1), 89–107. https://doi.org/10.3917/enf2.201.0089

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-Month-Old Infants, Melodies Are Social. *Psychological Science*, *27*(4), 486–501. https://doi.org/10.1177/0956797615626691

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6). https://doi.org/10.1111/desc.13122

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Menn, K. H., Ward, E. K., Braukmann, R., van den Boomen, C., Buitelaar, J., Hunnius, S., & Snijders, T. M. (2022). Neural Tracking in Infancy Predicts Language Development in Children With and Without Family History of Autism. *Neurobiology of Language*, *3*(3), 495–514. https://doi.org/10.1162/nol_a_00074

Müller, M. (2015). *Fundamentals of Music Processing*. Springer eBooks. https://doi.org/10.1007/978-3-319-21945-5

Müller, M., & Chiu, C.-Y. (2024). A Basic Tutorial on Novelty and Activation Functions for Music Signal Processing. *Transactions of the International Society for Music Information Retrieval*, *7*(1), 179–194. https://doi.org/10.5334/tismir.202

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

Nakata, T., & Trehub, S. E. (2011). Expressive timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicology: Music, Mind and Brain*, *21*(1–2), 45–53. https://doi.org/10.1037/h0094003

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Nave, K. M., Hannon, E. E., & Snyder, J. S. (2022). Steady state-evoked potentials of subjective beat perception in musical rhythms. *Psychophysiology*, *59*(2), e13963. https://doi.org/10.1111/psyp.13963

Newman, R. S., & Hussain, I. (2006). Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy*, *10*(1), 61–76. https://doi.org/10.1207/s15327078in1001_4

Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313

Novembre, G., & Iannetti, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences*, *115*(47). https://doi.org/10.1073/pnas.1815311115

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Palazzi, A., Filippa, M., Meschini, R., & Piccinini, C. A. (2021). Music therapy enhances preterm infant's signs of engagement and sustains maternal singing in the NICU. *Infant Behavior and Development*, *64*, 101596. https://doi.org/10.1016/j.infbeh.2021.101596

Papadimitriou, A., Smyth, C., Politimou, N., Franco, F., & Stewart, L. (2021). The impact of the home musical environment on infants' language development. *Infant Behavior and Development*, *65*, 101651. https://doi.org/10.1016/j.infbeh.2021.101651

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal Music Exposure Induces Long-Term Neural Effects. *PLoS ONE*, *8*(10), e78946. https://doi.org/10.1371/journal.pone.0078946

Patscheke, H., Degé, F., & Schwarzer, G. (2016). The Effects of Training in Music and Phonological Skills on Phonological Awareness in 4- to 6-Year-Old Children of Immigrant Families. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01647

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., & Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, *107*(10), 4758–4763. https://doi.org/10.1073/pnas.0909074107

Perapoch Amadó, M., Greenwood, E., Ives, J., Labendzki, P., Haresign, I. M., Northrop, T. J., Phillips, E. A. M., Viswanathan, N. K., Whitehorn, M., Jones, E. J. H., & Wass, S. V. (2023). *The neural and physiological substrates of real-world attention change across development.* https://doi.org/10.7554/elife.92171.2

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, *308*(5727), 1430–1430. https://doi.org/10.1126/science.1110922

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*(3), 533–546. https://doi.org/10.1016/j.cognition.2006.11.006

Plantinga, J., & Trainor, L. J. (2009). Melody recognition by two-month-old infants. *The Journal of the Acoustical Society of America*, *125*(2), EL58–EL62. https://doi.org/10.1121/1.3049583

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior and Development*, *54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in 2½- and 4-year-old children. *International Journal of Behavioral Development*, *27*(3), 220–231. https://doi.org/10.1080/01650250244000290

Putkinen, V., Saarikivi, K., & Tervaniemi, M. (2013). Do informal musical activities shape auditory skill development in preschool-age children? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00572

Python Software Foundation. (2023). *Python* (Version 3.11.5) [Computer software].

Rocha, S., & Mareschal, D. (2017). Getting into the Groove: The Development of Tempo-Flexibility Between 10 and 18 Months of Age. *Infancy*, *22*(4), 540–551. https://doi.org/10.1111/infa.12169

Rocha, S., Southgate, V., & Mareschal, D. (2021a). Infant Spontaneous Motor Tempo. *Developmental Science*, *24*(2), e13032. https://doi.org/10.1111/desc.13032

Rocha, S., Southgate, V., & Mareschal, D. (2021b). Rate of infant carrying impacts infant spontaneous motor tempo. *Royal Society Open Science*, *8*(9), 210608. https://doi.org/10.1098/rsos.210608

Rochat, P. R. (2001). Social contingency detection and infant development. *Bulletin of the Menninger Clinic*, *65*(3), 347–360. https://doi.org/10.1521/bumc.65.3.347.19847

Rock, A. M. L., Trainor, L. J., & Addison, T. L. (1999). Distinctive Messages in Infant-Directed Lullabies and Play Songs. *Developmental Psychology*, *35*(2), 527–534. https://doi.org/10.1037/0012-1649.35.2.527

Saliba, S., Esseily, R., Filippa, M., Gratier, M., & Grandjean, D. (2020). Changes in the vocal qualities of mothers and fathers are related to preterm infant's behavioural states. *Acta Paediatrica*, *109*(11), 2271–2277. https://doi.org/10.1111/apa.15238

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, *19*(10), 831–836. https://doi.org/10.1016/j.cub.2009.03.061

Schwartz, F. J. (2004). Medical music therapy for the premature baby—Research review. In ;M. Nöcker-Ribaupierre (Ed.), *Music therapy for premature and newborn infants* (pp. 85–96). Barcelona Publishers.

Shahin, A. J., Trainor, L. J., Roberts, L. E., Backer, K. C., & Miller, L. M. (2010). Development of Auditory Phase-Locked Activity for Music Sounds. *Journal of Neurophysiology*, *103*(1), 218–229. https://doi.org/10.1152/jn.00402.2009

Sharman, K. M., Meissel, K., Tait, J. E., Rudd, G., & Henderson, A. M. E. (2023). The effects of live parental infant-directed singing on infants, parents, and the parent-infant dyad: A systematic review of the literature. *Infant Behavior and Development*, *72*, 101859. https://doi.org/10.1016/j.infbeh.2023.101859

Smith, A. R., & Kong, K. L. (2025). Music enrichment programs may promote early language development by enhancing parent responsiveness: A narrative review. *Child Development Perspectives*, *19*(1), 20–29. https://doi.org/10.1111/cdep.12519

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, *120*(2), 304–308. https://doi.org/10.1016/j.clinph.2008.11.020

Steinberg, S., Liu, T., & Lense, M. D. (2021). Musical Engagement and Parent-Child Attachment in Families With Young Children During the Covid-19 Pandemic. *Frontiers in Psychology*, *12*, 641733. https://doi.org/10.3389/fpsyg.2021.641733

Stupacher, J., Hove, M. J., & Janata, P. (2016). Audio Features Underlying Perceived Groove and Sensorimotor Synchronization in Music. *Music Perception*, *33*(5), 571–589. https://doi.org/10.1525/mp.2016.33.5.571

Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: A TMS investigation. *Brain and Cognition*, *82*(2), 127–136. https://doi.org/10.1016/j.bandc.2013.03.003

Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior and Development*, *57*, 101346. https://doi.org/10.1016/j.infbeh.2019.101346

Tal, S., Grossman, E., & Arnon, I. (2024). Infant-directed speech becomes less redundant as infants grow: Implications for language learning. *Cognition*, *249*, 105817. https://doi.org/10.1016/j.cognition.2024.105817

Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour*, *27*, 699–715. https://doi.org/10.1016/0003-3472(79)90006-X

Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development*, *19*(1), 83–92. https://doi.org/10.1016/S0163-6383(96)90046-6

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, *20*(3), 383–396. https://doi.org/10.1016/S0163-6383(97)90009-6

Trehub, S. E. (1990). Human Infants' Perception of Auditory Patterns. *International Journal of Comparative Psychology*, *4*(1). https://doi.org/10.46867/c4hg68

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669–673. https://doi.org/10.1038/nn1084

Trehub, S. E. (2017). The Maternal Voice as a Special Signal for Infants. In M. Filippa, P. Kuhn, & B. Westrup (Eds), *Early Vocal Contact and Preterm Infant Brain Development: Bridging the Gaps Between Research and Practice* (pp. 39–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-65077-7_3

Trehub, S. E. (2018). Music lessons from infants. In *Oxford Handbook of Music Psychology*. Oxford University Press.

Trehub, S. E. (2019). Nurturing infants with music. In *International Journal of Music in Early Childhood* (Vol. 14, Issue 1, pp. 9–15). Intellect. https://doi.org/10.1386/ijmec.14.1.9_1

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096. https://doi.org/10.1098/rstb.2014.0096

Trehub, S. E., & Cirelli, L. K. (2018). Precursors to the performing arts in infancy and early childhood. In *Progress in Brain Research* (Vol. 237, pp. 225–242). Elsevier. https://doi.org/10.1016/bs.pbr.2018.03.008

Trehub, S. E., & Gudmundsdottir, H. R. (2015). *Mothers as singing mentors for infants*. online publication. Oxford University Press.

Trehub, S. E., Plantinga, J., & Russo, F. A. (2016). Maternal Vocal Interactions with Infants: Reciprocal Visual Influences. *Social Development*, *25*(3), 665–683. https://doi.org/10.1111/sode.12164

Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43–77.

Trehub, S. E., Unyk, A. M., & Henderson, J. L. (1994). Children's songs to infant siblings: Parallels with speech. *Journal of Child Language*, *21*(3), 735–744. https://doi.org/10.1017/S0305000900009533

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental Psychology*, *33*(3), 500–507. https://doi.org/10.1037/0012-1649.33.3.500

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae*, *3*(1_suppl), 155–215. https://doi.org/10.1177/10298649000030S109

Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior and Development*, *33*(1), 96–100. https://doi.org/10.1016/j.infbeh.2009.11.006

Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants Prefer Infant-Directed Song Over Speech. *Child Development*, *88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The Impact of the Bass Drum on Human Dance Movement. *Music Perception*, *30*(4), 349–359. https://doi.org/10.1525/mp.2013.30.4.349

Virtala, P., Huotilainen, M., Partanen, E., Fellman, V., & Tervaniemi, M. (2013). Newborn infants' auditory system is sensitive to Western music chord categories. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00492

Volkova, A., Trehub, S. E., & Schellenberg, E. G. (2006). Infants' memory for musical performances. *Developmental Science*, *9*(6), 583–589. https://doi.org/10.1111/j.1467-7687.2006.00536.x

Vuilleumier, P., & Trost, W. (2015). Music and emotions: From enchantment to entrainment: Musical emotions and entrainment. *Annals of the New York Academy of Sciences*, *1337*(1), 212–222. https://doi.org/10.1111/nyas.12676

Vuust, P., & Witek, M. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, 1111. https://doi.org/10.3389/fpsyg.2014.01111

Weineck, K., Wen, O. X., & Henry, M. J. (2022). Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *eLife*, *11*, e75515. https://doi.org/10.7554/eLife.75515

Werchan, D. M., Hume, A., Zhang, M., Vo, T., & Brito, N. H. (2025). From focus to function: Longitudinal insights into infant attention and emerging executive functions via remote webcam eye tracking. *Developmental Psychology*, *61*(5), 957–963. https://doi.org/10.1037/dev0001948

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468–2471. https://doi.org/10.1073/pnas.0809035106

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, *212*(1–2), 185–202. https://doi.org/10.1016/j.heares.2005.11.010

Xie, W., Mallin, B. M., & Richards, J. E. (2018). Development of infant sustained attention and its relation to EEG oscillations: An EEG and cortical source analysis study. *Developmental Science*, *21*(3). https://doi.org/10.1111/desc.12562

Zentner, M., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, *107*(13), 5768–5773. https://doi.org/10.1073/pnas.1000121107

Zhang, Y., Martinez-Cedillo, A. P., Mason, H. T., Vuong, Q. C., Garcia-de-Soria, M. C., Mullineaux, D., Knight, M. I., & Geangu, E. (2025). An automatic sustained attention prediction (ASAP) method for infants and toddlers using wearable device signals. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-96794-x

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, *113*(19), 5212–5217. https://doi.org/10.1073/pnas.1603984113

6. Annex

6.1. English Abstract

Infant-directed (ID) songs are widely used to meet various caregiving needs, such as entertainment, soothing, and capturing and holding attention. Infants are attentive to ID songs, neurally respond to musical elements like rhythm and pitch, and move in response to recorded musical stimuli. To examine whether ID songs have the effects intended by caregivers, infants' musical engagement with different acoustic features and musical structures needs to be investigated on multiple modalities, both in controlled playback experiments and live interactions.

The studies in this dissertation aim to elucidate which acoustic qualities in ID songs elicit strong neural, motor, and attentional responses, and how infants and caregivers engage musically across modalities. The first study investigates how infants neurally process and exhibit movement responses to controlled ID musical stimuli at 3, 6, and 12 months of age, and whether musical structure and pitch influence these responses. It underlines the importance of musical structure for neural processing of ID songs. It further shows that neural processing can be observed earlier in development than rhythmic movement responses, that only 6-month-olds neurally process high pitch better than low pitch, and that infants are unable to synchronise to a musical beat in their first year of life. The second and third studies examine live maternal singing of playsongs and lullabies to 7-month-old infants. The second study demonstrates better neural tracking of lullabies, more rhythmic movement to playsongs, and positive correlations between vocabulary size at 20 months to neural tracking and rhythmic movement to playsongs at 7 months. The third study highlights how infant attention and maternal acoustic variability in ID singing create a multimodal interaction dependent on song type. Playsongs elicit more infant attention than lullabies, and mothers react to infant attention with acoustic variability in playsongs, not lullabies.

Together, the three studies provide valuable insights into the neural processing, behavioural reactions to, and multimodal interactional nature of ID songs, as well as long-term language outcomes. Given the ubiquity of ID songs in caregiving, these results lay an important building block of our understanding of how infants perceive and interact with ID songs, with possible developmental implications.

6.2. Deutsche Zusammenfassung

Säuglings- und kindgerichtete Lieder werden für diverse Zwecke im Rahmen der Kinderbetreuung genutzt, beispielsweise zur Unterhaltung, zum Beruhigen oder um Aufmerksamkeit zu erregen und zu erhalten. Säuglinge schenken kindgerichteten Liedern Aufmerksamkeit, verarbeiten musikalische Elemente wie Rhythmus und Tonhöhe neuronal, und bewegen sich zu wiedergegebenen musikalischen Stimuli. Um zu untersuchen, ob kindgerichtete Lieder die von Bezugspersonen intendierten Effekte auf Säuglinge haben, ist es notwendig, die musikalische Auseinandersetzung von Säuglingen mit verschiedenen akustischen Merkmalen und musikalischen Strukturen auf mehreren Modalitäten zu untersuchen, sowohl in kontrollierten Playbackexperimenten als auch in Live-Interaktionen.

Die Studien dieser Dissertation versuchen darüber aufzuklären, welche akustischen Merkmale in kindgerichteten Liedern starke neuronale, motorische, und Aufmerksamkeitsreaktionen hervorrufen. Die erste Studie untersucht, wie 3, 6 und 12 Monate alte Säuglinge kontrollierte, kindgerichtete Musikstimuli neuronal und motorisch verarbeiten, und ob musikalische Struktur und Tonhöhe diese Prozesse beeinflussen. Diese Studie unterstreicht die Bedeutung musikalischer Struktur für die neuronale Verarbeitung kindgerichteter Lieder. Des Weiteren zeigt sie, dass die neuronale Verarbeitung musikalischer Stimuli in der Entwicklung früher beobachtet werden kann als rhythmische Bewegung als Reaktion auf ebendiese musikalischen Stimuli. Sie zeigt auch, dass nur im Alter von sechs Monaten höhere Tonhöhen besser neuronal verarbeitet werden als tiefere Tonhöhen, und dass Säuglinge in ihrem ersten Lebensjahr nicht mit Taktschlägen synchronisieren können. Die zweite und dritte Studie untersuchen mütterlichen Live-Gesang von Spielliedern und Schlafliedern zu sieben Monate alten Säuglingen. Die zweite Studie zeigt, dass Schlaflieder besser neuronal getrackt werden und Spiellieder mehr rhythmische Bewegung auslösen, und dass es positive Korrelationen zwischen der Größe des Wortschatzes im Alter von zwanzig Monaten und dem neuronalen Tracking sowie rhythmischen Bewegungen zu Spielliedern im Alter von sieben Monaten gibt. Die dritte Studie beschreibt, wie die Aufmerksamkeit von Säuglingen und die Variabilität im Singen von Müttern eine multimodale Interaktion abhängig vom Liedtyp kreieren. Spiellieder lösen mehr Aufmerksamkeit aus als Schlaflieder und Mütter reagieren auf Aufmerksamkeit der Säuglinge in Spielliedern, nicht in Schlafliedern.

Gemeinsam bieten diese drei Studien wertvolle Einblicke in die neuronale Verarbeitung, Verhaltensreaktionen, und multimodale, interaktive Natur kindgerichteter

Lieder, sowie Langzeitfolgen im sprachlichen Bereich. Angesichts der Allgegenwärtigkeit kindgerichteter Lieder im Betreuungskontext bilden diese Ergebnisse einen wichtigen Baustein im Verständnis davon, wie Säuglinge kindgerichtete Lieder wahrnehmen und mit ihnen interagieren, und zeigen mögliche Implikationen für die Entwicklung auf.

6.3. Scientific Publications for the Cumulative Thesis

Accepted for publishing in a journal/book/anthology:

1) Nguyen, T., Bigand, F., Reisner, S., Koul, A., Bianco, R., Markova, G., Hoehl, S., & Novembre, G. (2026). Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year. *bioRxiv*, 2025.05.04.649695. https://doi.org/10.1101/2025.05.04.649695
2) Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313
3) Reisner, S., Nguyen, T., Labendzki, P., Hoehl, S., & Markova, G. (2025). The reciprocal relationship between maternal infant-directed singing and infant gaze. *Musicae Scientiae, 0*(0). https://doi.org/10.1177/10298649251385676

Notes.

[1] This manuscript has been accepted for publication in *eLife* and is currently available on the journal homepage as a reviewed, not revised preprint: https://doi.org/10.7554/eLife.107088.1

1. Manuscript

Nguyen, T., Bigand, F., Reisner, S., Koul, A., Bianco, R., Markova, G., Hoehl, S., & Novembre, G. (2026). Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year. *bioRxiv*, 2025.05.04.649695. https://doi.org/10.1101/2025.05.04.649695 [1]

Notes.

[1] This manuscript has been accepted for publication in *eLife* and is currently available on the journal homepage as a reviewed, not revised preprint: https://doi.org/10.7554/eLife.107088.1

Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year

Trinh Nguyen[1,2,3*], Félix Bigand[1], Susanne Reisner[2,4], Atesh Koul[1], Roberta Bianco[1,5], Gabriela Markova[6], Stefanie Hoehl[2], & Giacomo Novembre[1*]

[1]Neuroscience of Perception and Action Lab, Italian Institute of Technology, Rome, Italy

[2]Department of Developmental and Educational Psychology, University of Vienna, Vienna, Austria

[3]Department of Developmental and Biological Psychology, Heidelberg University, Heidelberg, Germany

[4]Doctoral School Cognition, Behavior and Neuroscience, University of Vienna, Austria

[5]Department of Translational Research & New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy

[6]Institute for Early Life Care, Paracelsus Medical University, Salzburg, Austria.

*Corresponding Authors. Trinh Nguyen, Viale Regina Elena 291, 00161 Rome. Trinh.nguyen@iit.it. Giacomo Novembre, Viale Regina Elena 291, 00161 Rome. Giacomo.novembre@iit.it

Acknowledgements

We are grateful to Eluisa Nimpf, Flavia Arnese, Larissa Reitinger, Josefine Schürholz and Liesbeth Forsthuber for their support in data acquisition, processing, and video coding. Additionally, we thank all families who participated in the study and the Department of Obstetrics and Gynaecology of the Vienna General Hospital for supporting our participant recruitment.

Funding

This research has received funding from the European Research Council awarded to Giacomo Novembre (MUSICOM, 948186), the Austrian Science Fund (FWF) DK Grant "Cognition & Communication 2" W1262-B29 awarded to Stefanie Hoehl (10.55776/W1262) and has been supported by the Brain and Machines Flagship Programme of the Italian Institute of Technology. Trinh Nguyen and Roberta Bianco also acknowledge the support of Horizon Europe's Marie Skłodowska-Curie Actions (SYNCON, 101105726; PHYLOMUSIC, 101064334).

Credits

Trinh Nguyen: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data curation, Visualization, Writing - Original draft, Writing – Review & Editing, Project administration; **Félix Bigand**: Software, Validation, Data curation, Formal Analysis, Writing – Review & Editing. **Susanne Reisner**: Investigation, Data curation, Writing – Review & Editing, Project administration. **Atesh Koul**: Software, Writing – Review & Editing; **Roberta Bianco**: Validation, Software, Writing - Original draft, Writing – Review & Editing; **Gabriela Markova**: Conceptualization, Data curation, Writing - Review & Editing; **Stefanie Hoehl**: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition; **Giacomo Novembre**: Conceptualization, Methodology, Resources, Writing - Original draft, Writing - Review & Editing, Supervision, Funding acquisition.

Data availability

The data and stimuli reported in this manuscript are available in the following repository: https://doi.org/10.48557/DCSCFO. The codes used for analyses and figures are available on Github: https://github.com/tnguyen1992/tinydancer.

Abstract

Humans across cultures not only share the ability to recognise music but also respond to it through movement. While the sensory encoding of music is well-studied, when and how infants naturally start moving to music is largely unexplored. This study simultaneously investigates infants' neural (auditory) responses and spontaneous movements to music during the first postnatal year. Neural activity (EEG) and body kinematics (markerless pose estimation) were recorded from 79 infants (aged 3, 6, and 12 months) listening to refrains of children's music, along with shuffled, high-pitched, and low-pitched versions of the same songs. Neural data revealed that, across all ages, infants exhibit enhanced auditory responses to music compared to shuffled music, indicating that auditory encoding of music emerges early in development. Movement data revealed a different outcome. While coarse auditory-motor coupling is present at all ages, more complex structured movement patterns emerge in response to music only by 12 months. Notably, no age group demonstrated evidence of coordinated movements to music. Additionally, enhanced auditory responses to high vs low pitch were only evident at 6 months, while infants' movements were better predicted by high-pitched compared to low-pitched music at all ages. This study provides initial insights into how the developing brain gradually transforms music into spontaneous movements of increasing complexity.

Keywords: Music, EEG, movement, development, infancy, pitch

Introduction

Musicality – the biological predisposition to perceive, appreciate, and produce music (Honing, 2018; Trehub, 2003) – is increasingly recognised as a fundamental aspect of human nature. Numerous accounts suggest that engaging with music through movement is at the core of musicality (Honing et al., 2015; Schachner et al., 2009; Trehub et al., 2015). Functionally, such engagement can be broken down into two fundamental components of neurocognitive development: the ability to perceive and recognise music (*sensory component*), and the ability to produce movement responses that are temporally aligned with the musical structure, from coordinated vocalizations and percussive actions up to complex dance moves (*motor component*; Brown, 2022; Trehub, 2003a; Trevarthen, 1999). Despite this inherent predisposition toward music, the developmental trajectory of infants' musicality remains largely unknown (see Nguyen et al., 2023, for a review). While there is increasing research on infant music perception, including controlled manipulations of select musical features, we know less about the translation of perception into action, namely the ontogenesis of infants' spontaneous movements to music (see Fujii et al., 2014; Nguyen, Reisner, et al., 2023; Zentner & Eerola, 2010). Furthermore, making our understanding of music-driven motor engagement even more incomplete, no studies to date have looked at both brain activity and spontaneous body movements simultaneously, especially during the first year of life. Accordingly, how the processing of music and its features is transformed into organised motor responses remains underexplored.

The *sensory component* of musicality, namely music perception, can be measured using electroencephalography (EEG), specifically by recording cortical auditory evoked potentials (event-related potentials [ERP]). One of these responses is the infantile P1, a phase-locked EEG positivity peaking around 200-300 ms after an auditory stimulus (Chen et al., 2016; Kushnerenko et al., 2002; Wunderlich et al., 2006). The infantile P1 has been observed in response to both musical notes and speech segments. Auditory evoked potentials, when elicited isochronously, can also be captured using frequency domain analyses (Damsma et al., 2024; Novembre & Iannetti, 2018) such as auditory steady-state responses (ASSR), which are also called steady-state evoked potentials (SSEP, e.g., Cirelli et al., 2016; Nave et al., 2022). These neural responses can provide insight into the developing auditory system and its ability to encode musical structure. Using these neurophysiological measures, prior research has shown that newborns and infants are sensitive to beat structure, pitch deviants and tone interval regularities (Bianco et al., 2025; Edalati et al., 2023; Háden et al., 2009, 2015, 2022; Stefanics et al., 2009; Winkler et al., 2009). Despite these promising results, the

neurophysiology of early music processing – particularly its developmental trajectory – remains not fully understood. Here, our primary goal is to investigate infants' neural encoding of music utilizing both ERP and ASSR approaches to characterize how such neural responses change across the first year of life.

Another component of musicality is the capacity to move to music (*motor component;* Brown, 2022; Fitch, 2015; Honing et al., 2015; Trehub et al., 2015). This capacity is linked to infants not only recognising musical structure but also moving their bodies in response to it. Even though this capacity appears to develop precociously, as evidenced by the fact that even 28-35-week-old foetuses move to music (Kisilevsky et al., 2004), very few studies have systematically examined music-driven spontaneous body movements in infants. An influential paper by Zentner and Eerola (2010) reported that infants across a large age range (from 5 to 24 months) showed more spontaneous rhythmic movements in response to classical music and children's music compared to infant-directed speech. Importantly, their movements were not synchronized with the musical input, even though a small degree of tempo flexibility was observed (i.e., faster musical tempi evoked relatively faster movement periodicities). The lack of synchrony between music and body movements has also been reported in younger (i.e., 3-4 months old) infants listening to popular music (Fujii et al., 2014). Further, another study testing 7-month-old infants reported more movement in response to (sung) playsongs compared to lullabies but did not assess movement synchrony (Nguyen, Reisner, et al., 2023). Despite these initial investigations, it remains unclear when infants begin to move in response to music, which specific movements are evoked, and when these movements become coordinated with the music. Moreover, a critical limitation in existing research is the lack of a control condition to determine whether these movements are driven specifically by musical structure or reflect general motor activity in response to auditory input. As a second goal, this study is the first to systematically test the gradual development of music-induced movements in different age groups across the first year of life.

Music engages both sensory and motor systems, yet different musical features may differentially shape infants' engagement with music. While rhythm has been widely studied in early music cognition, pitch is another salient acoustic cue that could play a role in auditory-motor engagement, particularly in infancy. High pitch is a defining feature of infant-directed speech (Fernald & Simon, 1984), among other features such as exaggerated intonation, slower tempo, and simplified vocabulary (Fernald & Kuhl, 1987; Kuhl & Meltzoff, 1982). Similarly, infants most frequently listen to music characterized by high pitch (Costa-Giomi & Sun, 2016; Nakata & Trehub, 2011). Reflecting its prominence, high pitch is

found to be one of the most prominent features thought to effectively capture (Conrad et al., 2011; Eckerdal & Merker, 2009; Trainor, 1996; Trainor & Zacharias, 1998) and guide infants' attention (Lense et al., 2022; Trainor & Desjardins, 2002). On the neural level, infants are also better at encoding pitch deviances in the high voice of polyphonic music, thus showing *high voice superiority* from 3 months of age (Marie & Trainor, 2013, 2014). Taken together, these findings indicate that higher-pitch music would amplify infants' neural responses (i.e., sensory component) in comparison to lower-pitch music. On the other hand, we know that adults move more to music with greater energy in lower frequencies (Cameron et al., 2022; Stupacher et al., 2013, 2016; Van Dyck et al., 2013). Yet, it remains unknown whether low-pitch music elicits increased movement in infants, as it does in adults, or whether infants' attraction to high pitch also extends to enhance their motor responses. As a third goal, we thus investigate how musical pitch affects infants' sensory and motor components.

We presented infants, aged 3, 6, and 12 months, with instrumental refrains of children's songs (music), shuffled versions of the same songs (shuffled music), and transpositions of the songs that would either emphasize the melody (high pitch) or the bassline (low pitch). We recorded infants' neural activity using EEG and specifically extracted ERPs and ASSR as indices of infants' neural response to the various auditory stimuli. We also analysed spontaneous (full-body) movement kinematics using automated video-based motion tracking (DeepLabCut) and extracted *principal movements* using principal component analysis (see Fig. 1, c.f. Bigand et al., 2024, Toiviainen et al., 2010). By adopting a cross-sectional design, we aimed to characterize the maturation of both auditory and movement responses across infancy. We hypothesized that auditory responses would be enhanced when triggered by music compared to shuffled music. This hypothesis was based on the notion that musical structure, notably eroded in the shuffled musical stimuli, is essential to attract infants' attention towards predictable events (Kouider et al., 2015; Lense et al., 2022). Similarly, based on previous evidence comparing movement responses to music vs speech and silence (Fujii et al., 2014; Zentner & Eerola, 2010), we expected the presence of musical structure to increase the likelihood of spontaneous movements in response to music compared to shuffled music, but we did not have a specific hypothesis about which particular movements would be produced. We further hypothesized that infants would show enhanced neural responses to high- compared to low-pitch music and explored co-occurring differences in spontaneous movements. Generally, we aimed to characterize the maturation of both auditory and motor responses as infants get older. By studying both sensory and motor components of musicality, we aimed to deepen our understanding of when and how infants

learn to transform what they perceive into spontaneous movements, eventually leading to the emergence of synchronization to music (Brown, 2022; Fitch, 2015; Honing et al., 2015; Patel & Iversen, 2014).

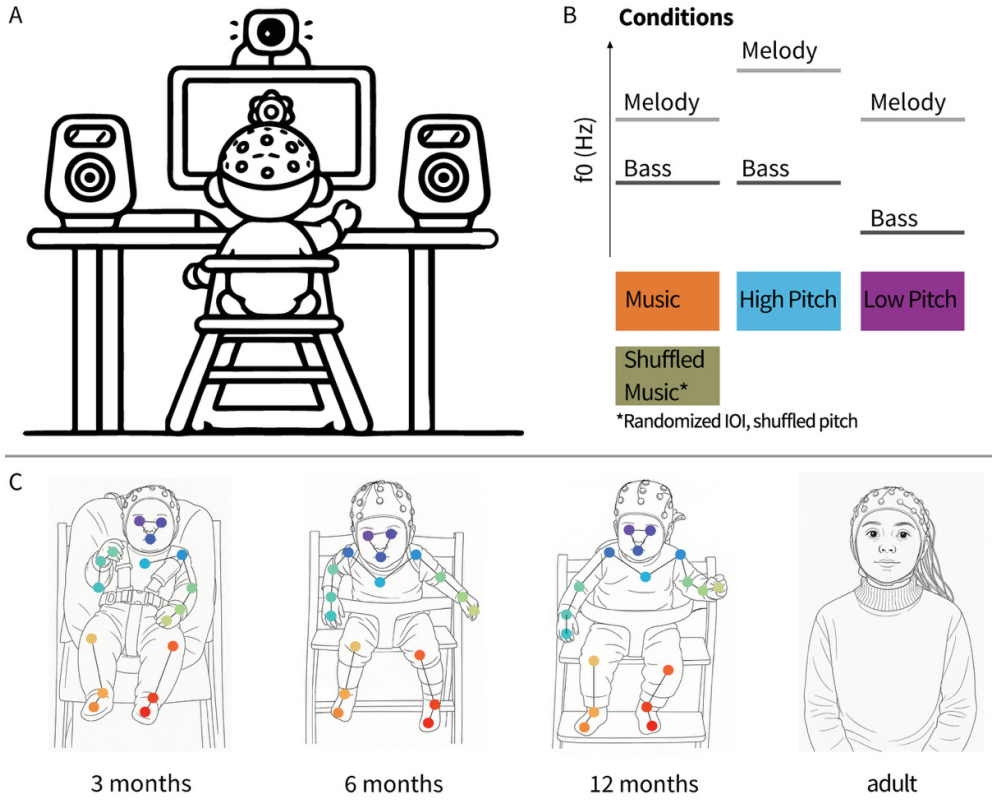


Figure 1. *Overview of the procedure (A), experimental conditions (B), and participant sample (C). (A) Infants sat in front of a screen with speakers on each side. The screen showed slowly blossoming flowers to attract infants' attention. Caregivers (not shown) sat behind the infants and wore noise-cancelling headphones. (B) Infants listened to polyphonic auditory stimuli consisting of a melody and a bassline in four different conditions. The music condition included two children's songs. The shuffled music condition included versions of the songs used in the music condition that were shuffled in pitch and randomized in inter-onset intervals (IOI). Stimuli belonging to the music and shuffled music conditions had the same pitches. In the high-pitch condition, the melody was shifted one octave higher than in the music condition. In the low-pitch condition, the bassline was shifted one octave lower than in the music condition. Hence, the two voices composing the high-pitch condition were one octave higher than those composing the low-pitch condition. (C) The sample included infants at 3 months (N=26), 6 months (N=26), 12 months (N=27), and an adult control sample (N=26). The dots overlaying the images represent the body parts whose movements were tracked using video-based kinematic analysis.*

Results

This study included EEG and movement measurements taken from 79 infants in the first postnatal year, as well as EEG measurements taken from a control sample of 26 adults. Participants were exposed to two polyphonic children's songs featuring a melody and a bassline, and their manipulated versions (see Table 1 and S1 for an acoustic characterization of the musical stimuli). We investigated neural responses and movement responses to music vs shuffled music (a control condition in which we shuffled the melody and randomized the inter-onset intervals (IOI) of the music). Additionally, we contrasted neural and movement responses to high- vs low-pitch music. Music vs shuffle conditions (manipulation of structure but not pitch height) and high pitch vs low pitch conditions (manipulation of pitch height but not structure) are contrasted separately to avoid comparing conditions differing in more than one variable. Below, we begin by reporting the neural results, followed by the movement results.

Table 1. Acoustic characteristics of the musical stimuli

Song name (Condition)	*TEMPO M±SD (BPM)*	*IOI RANGE (MS)*	*IPI M±SD (SEMITONES)*	*IPI RANGE (SEMITONES)*	*BASS NOTE AND PITCH RANGE (HZ)*	*MELODY NOTE AND PITCH RANGE (HZ)*
HOPP (MUSIC)	0.422±0.067	0.222-0.444	0.040±4.585	-10 - 7	E2 to E3 (82.41 – 164.82 Hz)	A3 to A4 (220.00 – 440.00 Hz)
HOPP (SHUFFLED MUSIC)	0.422±0.187	0.112-0.666	0.100±4.782	-7 - 12	E2 to E3 (82.41 – 164.82 Hz)	A3 to A4 (220.00 – 440.00 Hz)
HOPP (HIGH PITCH)	0.422±0.067	0.222-0.444	0.040±4.585	-10 - 7	E2 to E3 (82.41 – 164.81 Hz)	A4 to A5 (440.00 – 880.00 Hz)
HOPP (LOW PITCH)	0.422±0.067	0.222-0.444	0.040±4.585	-10 - 7	E1 to E2 (41.20 – 82.41 Hz)	A3 to A4 (220.00 – 440.00 Hz)
LOLA (MUSIC)	0.444±0.000	0.444-0.444	0.149±2.836	-7 - 7	E2 to E3 (82.41 – 164.82 Hz)	A3 to A4 (220.00 – 440.00 Hz)
LOLA (SHUFFLED MUSIC)	0.444±0.182	0.222-0.665	0.234±4.864	-12 - 11	E2 to E3 (82.41 – 164.82 Hz)	A3 to A4 (220.00 – 440.00 Hz)
LOLA (HIGH PITCH)	0.444±0.000	0.444-0.444	0.149±2.836	-7 - 7	E2 to E3 (82.41 – 164.81 Hz)	A4 to A5 (440.00 – 880.00 Hz)
LOLA (LOW PITCH)	0.444±0.000	0.444-0.444	0.149±2.836	-7 - 7	E1 to E2 (41.20 – 82.41 Hz)	A3 to A4 (220.00 – 440.00 Hz)

Note. M = mean, SD = standard deviation, BPM = beats per minute, IOI = inter-onset interval, IPI = inter-peak interval, Hz = hertz (frequency). IOI and IPI were computed across melody and bass notes. Hopp refers to the Hungarian playsong ("Hopp Juliska"), Lola refers to the Spanish playsong ("La vaca lola").

EEG: Event-related potentials (ERP)

To examine the neural responses, we analysed the average ERPs time-locked to the acoustic tone onsets of the bassline notes of the auditory stimuli (*Fig. 2*). Adults' ERPs, which served as a benchmark to interpret infants' responses, included an early positivity peaking at 37 ms post-stimulus (so-called "P50", here reaching an average amplitude of 1.05 μV), followed by a later negativity peaking at 87 ms post-stimulus (so-called "N100", here reaching an amplitude of -0.43 μV) and a second positivity peaking at 158 ms post-stimulus (so-called "P200", here reaching an amplitude of 0.85 μV). This triphasic EEG pattern has been widely observed in adults in response to fast-rising auditory stimuli and across different contexts (Novembre et al., 2018; Pratt et al., 2008; Remijn et al., 2014; Somervail et al., 2021). Cluster-based permutation analyses, contrasting the ERPs elicited by music vs shuffled music, revealed that the amplitude of the P50 – hereafter referred to as P1 – was larger in response to music compared to shuffled music, particularly within the time range comprised between -17 and 58 ms post-stimulus (*cluster-t*=366.16, *p*=.016). Similarly, the amplitude of the following P200 - hereafter referred to as P2 – was larger in response to music than to shuffled music, particularly within the time range of 114 to 190 ms post-stimulus (*cluster-t*=395.42, *p*=.016). Both P1 and P2 responses were observed over fronto-central electrodes, showing a medial distribution, in line with previous literature (e.g., Lijffijt et al., 2009).

All infants' ERPs showed a P1 response, while a P2 response was observed only in 12-month-old infants, albeit with a lower amplitude than the P1. The P1 latency decreased (X^2(2)=391.25, p<.001), and its amplitude increased (X^2(2)=8.59, p=.014) with age (*Fig. 2*, left column). Importantly, and in line with the adults' data, all infant groups exhibited enhanced P1 amplitudes in response to music compared to shuffled music. Cluster-based permutation (n_{Perm}=1000) testing revealed that 3-month-old infants' P1 amplitude was enhanced between 177 and 305 ms post-stimulus (*cluster-t*=1111.90, *p*=.002). Within this window, the P1 peaked at 212 ms and reached an average amplitude of 1.8 μV. The topography included a frontocentral cluster with a slight right lateralization. In 6-month-old infants, the amplitude of the P1 was enhanced between 116 and 284 ms post-stimulus (*cluster-t*=1401.60, *p*=.002), peaking at 165 ms and reaching an amplitude of 2.8 μV. The

topography included a few (centro-) parietal electrodes in addition to several frontocentral electrodes, with a bilateral activation. In 12-month-old infants, the amplitude of the neural response to music was enhanced in a two-peak cluster (*cluster-t*=1416.30, *p*=.002). The first peak, an infantile P1, occurred between 104 and 227 ms, peaked at 146 ms post-stimulus and reached an amplitude of 3.1 μV. Notably, 12-month-old infants exhibited an additional positivity, namely an infantile P2, possibly homologous to the P200 observed in adults. The P2 ranged from 307 to 325 ms post-stimulus and peaked at 316 ms, with an average amplitude of 1.026 μV. The topographies remained frontocentral but were more medial, similar to adults.

To rule out the possibility that increased neural responses in the music condition reflected differences in inter-tone temporal spacing (see IOI range in *Table 1*), we performed control analyses including only epochs from the shuffled music condition for which the prior or subsequent IOI exceeded the median IOI duration. The resulting ERPs in the shuffled music condition were highly similar to those reported above (compare *Fig. 2* with *Fig. S1*), and the difference between the music and shuffled music conditions remained significant across all age groups ($p<.05$). These findings indicate that the observed effects are not solely attributable to variations in inter-tone spacing but also reflect sensitivity to musical structure.

Next, we examined neural responses to the notes in the high- and low-pitch conditions (*Fig. 2,* right column). The morphology of both adults' and infants' ERPs was generally comparable to that elicited by the music condition. Cluster-based permutation (n_{Perm}=1000) testing revealed that the amplitude of the adults' ERPs was comparable across high- and low-pitch conditions ($p>.050$). This was also the case for both 3- and 12-month-old infants, but notably not for 6-month-old infants, who exhibited an enhanced P1 in response to high-pitch vs low-pitch conditions (*cluster-t*=763.84, *p*=.002). This enhanced positivity (178 and 332 ms) peaked at 204 ms and reached an average amplitude of 2.8 μV. Similar to the neural response elicited by the music condition, the topography included few (centro-) parietal electrodes in addition to several frontocentral electrodes, with a bilateral activation.

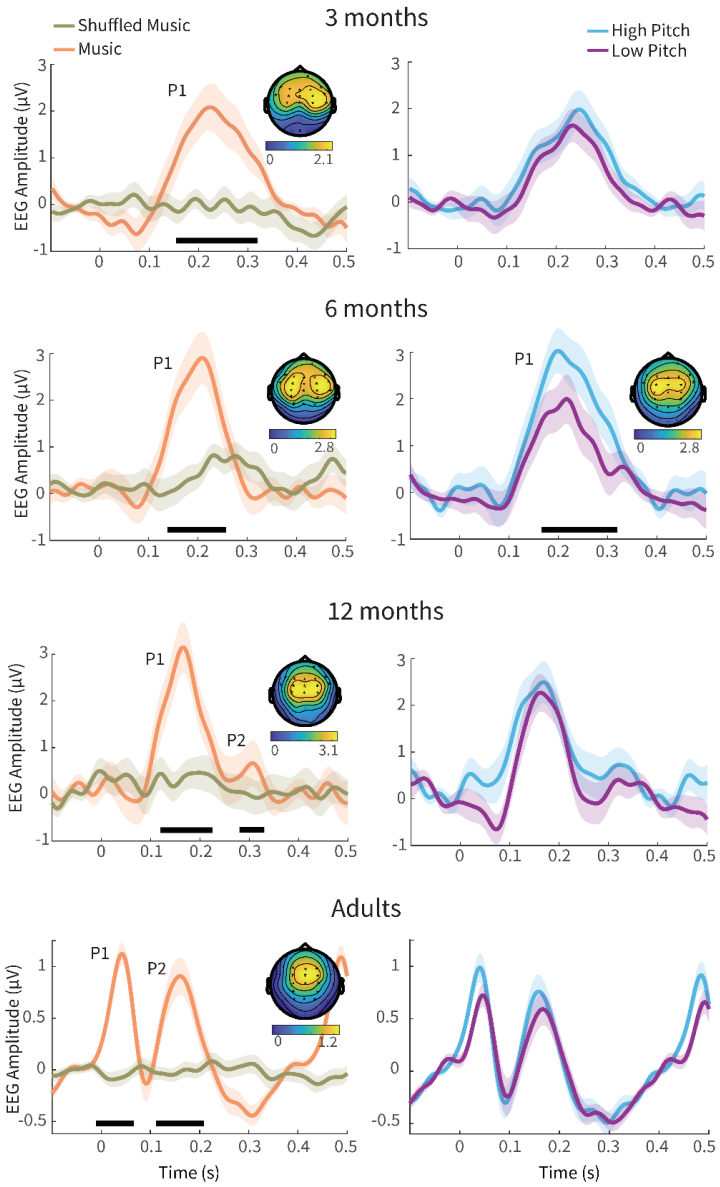


Figure 2. *Event-related potentials (ERPs) elicited by the notes comprised within the music (orange, left) vs shuffled music (khaki, left) as well as by the notes comprised within the high-pitch (light blue, right) vs low-pitch music (purple, right), across four groups of participants (plotted in ascending order of age, from top to bottom): 3-, 6-, 12-month-old infants (N=79) and adults (N=26). Grand-average ERPs are averaged across electrodes within the significant cluster of each age group in the music condition (except for pitch condition comparison in the 6-month-olds). Shaded areas indicate the standard error. ERPs show progressively shorter latencies with increasing age. All groups exhibited a P1 response, while only older infants (12-month-olds) and adults additionally exhibited a P2. Music elicited a larger P1 (and, when present, P2) amplitude compared to shuffled music, notably across all*

groups (time ranges associated with a significant difference are indicated by horizontal black lines). The topography of this neural response (averaged across the time window of the P1 cluster) in the music condition appears to shift more medially with increasing age. Colorbars beneath topography plots index EEG amplitude values.

EEG: Auditory Steady State Responses (ASSR)

Figure 3 shows bar plots indexing the relative power of ASSRs elicited by the auditory stimuli (power estimates were averaged across the electrodes comprised within the ERP clusters that were common to all age groups, i.e., FP2, F7, F3, Fz, F4, F8, FC7, FC3, FCz, FC4, FC8, C3, Cz, C4; see *Fig. 2*).

Our frequency of interest was 2.25 Hz, matching the musical beat as well as the presentation rate of the majority of the notes (15 out of 16 notes) in the auditory sequences (see methods, section "Stimuli"). We used linear mixed models, including power estimates as the dependent variable to contrast the different conditions and age groups (fixed and interaction effects). Participants were modelled as random intercepts. Power estimates were generally higher in response to music as opposed to shuffled music. Model outputs indicated that the power of the ASSR elicited by the music condition was significantly higher than that elicited by the shuffled music condition in 3-month-old infants ($F(1,50)=7.82$, $p=.007$), 12-month-old infants ($F(1,52)=12.03$, $p=.001$) and adults ($F(1,50)=13.49$, $p<.001$); while it only reached a trend for significance in 6-month-old infants ($F(1,50)=2.95$, $p=.092$). ASSR power estimates were not different across high-pitch and low-pitch conditions across all age groups ($p>.240$). These results indicated that most groups tracked the periodicity of the musical beat, leading to an enhancement of power in ASSR at a frequency matching the musical beat. This is generally in line with the ERP results showing stronger neural responses to music than shuffled music (even though 6-month-old infants only showed a marginal difference).

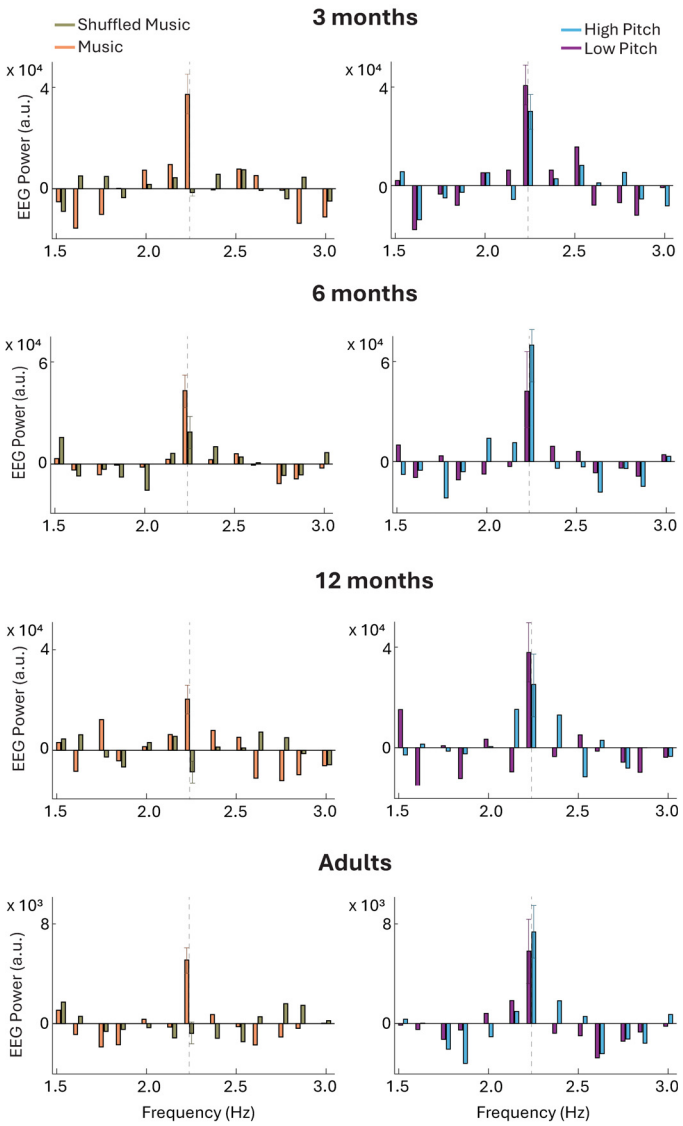


Figure 3. *Relative EEG Power (arbitrary unit [a.u.], y-axis) of the auditory steady-state responses (ASSR) elicited by music versus shuffled music (orange and khaki, left), and high-pitch versus low-pitch musical stimuli (blue and purple, right), across four groups of listeners: 3-month-olds (first row), 6-month-olds (second row), 12-month-olds (third row), and adults (fourth row). ASSR power estimates at the frequency (x-axis) matching the musical beat (2.25 Hz, highlighted by vertical dashed lines and including standard error bars) were statistically higher when elicited by music compared to shuffled music across nearly all participant groups (i.e., all but 6-month-olds). High- and low-pitch stimuli evoked similar ASSR (at 2.25 Hz). These results generally align with the ERP results (Fig. 2) across most infant groups and adults, except for 6-month-old infants for whom differences across conditions were either trending (music vs shuffled) or not significant (high vs low pitch).*

Extraction of Principal Movements and Estimation of Quantity of Movement

Using principal component analysis, we decomposed full-body kinematics into 10 Principal Movements (PMs), explaining 79.7 % of the whole kinematic variance. The PMs (depicted in Fig. 4) were reminiscent of common infant movements such as front-back rocking (PM1), side sway (PM2), proto-clapping (PM3), leg kicking (PM4), up-down rocking (PM5), arm pedalling (PM6), feet kicking (PM7), whole body wiggling (PM8), feet shuffling (PM9) and feet pedalling (PM10). Labels were assigned qualitatively, following visual inspection of the PMs.

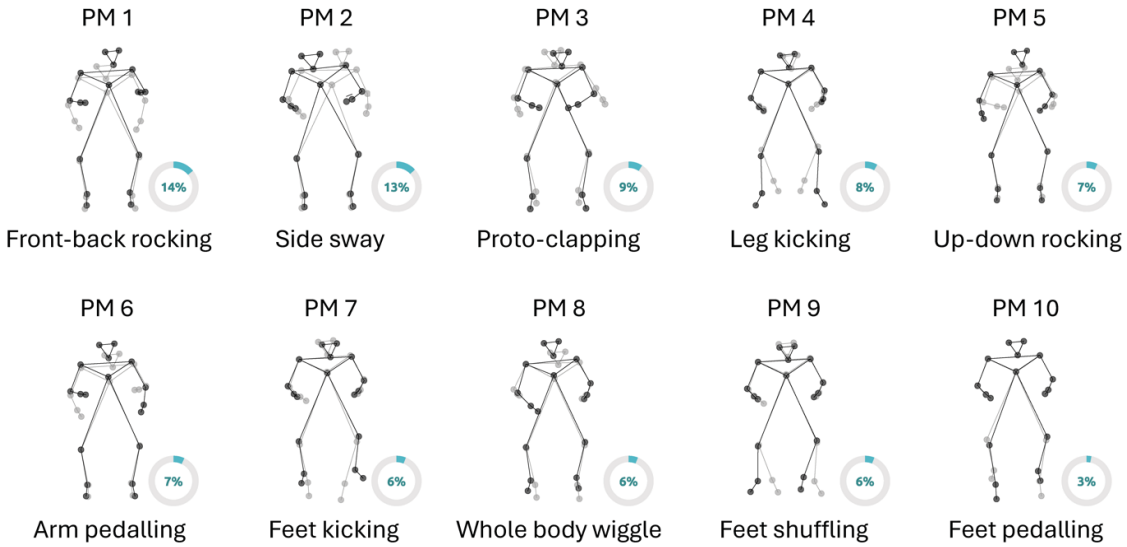


Figure 4. *Infants' principal movements (PMs). PMs are illustrated by showing the two most different body postures (min and max of the PM score, in grey and black, respectively) from the frontal perspective. The reader should interpret the PM as the kinematic displacement necessary to shift from one body posture (grey) to the other (black). Circle diagrams denote the proportion (%) of kinematic variance explained by each PM. Together, the ten PMs account for 79.7% of the total kinematic variance.*

Quantity of Movement (QoM) was estimated and compared across PMs, conditions, and groups (*Fig. 5*). QoM estimates were based on first-degree differentiation of the PM time series (see methods). A random intercept was included for each infant. Linear mixed effect modelling yielded a significant interaction between Condition and Age ($\chi^2(2)=16.76$, $p<.001$) indicating that only 12-month-old infants exhibited higher QoM in response to music compared to shuffled music across all PMs (i.e., when QoM was computed as an average across all PMs). Post-hoc contrasts comparing QoM averages between conditions revealed a

highly significant effect ($t(69.8)$=4.86, p<.001; see top left panel of Fig. 5 for QoM averaged across PMs).

Even though there was no interaction effect between Condition, Age and PMs, we still ran post-hoc comparisons to gain preliminary evidence about specific PMs driving the above-described effects. Results indicated that differences in 12-month-olds' QoM in response to music vs shuffled music were mostly driven by movements of the upper body and/or upper limbs. Specifically, front-back rocking (PM1), side sway (PM2), proto-clapping (PM3), up-down rocking (PM5) and arm pedalling (PM6) were linked with significantly higher QoM in response to music as opposed to shuffled music (PMs 1,2,3,5,6; ps<.050, corrected using the false-discovery rate). Contrarily, younger infants (3- and 6-month-olds) did not exhibit significantly different QoM in response to music vs shuffled music in any of the PMs (ps>.123). Further, when comparing infants' QoM to music at different pitches, we found no significant differences between conditions across PMs and age groups (ps>.295).

The model also yielded a significant interaction between PMs and Age ($\chi^2(18)$=181.575, p<.001), indicating that QoM generally increased with age but differently across PMs. Specifically, from 3 to 6 months, the PMs associated with a faster increase in QoM were front-back rocking (PM1), proto-clapping (PM3) and arm pedalling (PM6) (t(95.4)=2.34-2.60, p=.016-.032). From 3 to 12 months, front-back rocking (PM1), side sway (PM2), proto-clapping (PM3), up-down rocking (PM5) and arm pedalling (PM6) became more prevalent (t(95.4)=2.69-5.06, p=.001-.025). From 6 to 12 months, proto-clapping (PM3) became even more prevalent (t(95.4)=2.71, p=.012). Across the first year of life, infants seemed to consistently move their lower body while slowly increasing their capacity for upper-body and whole-body movements while seated.

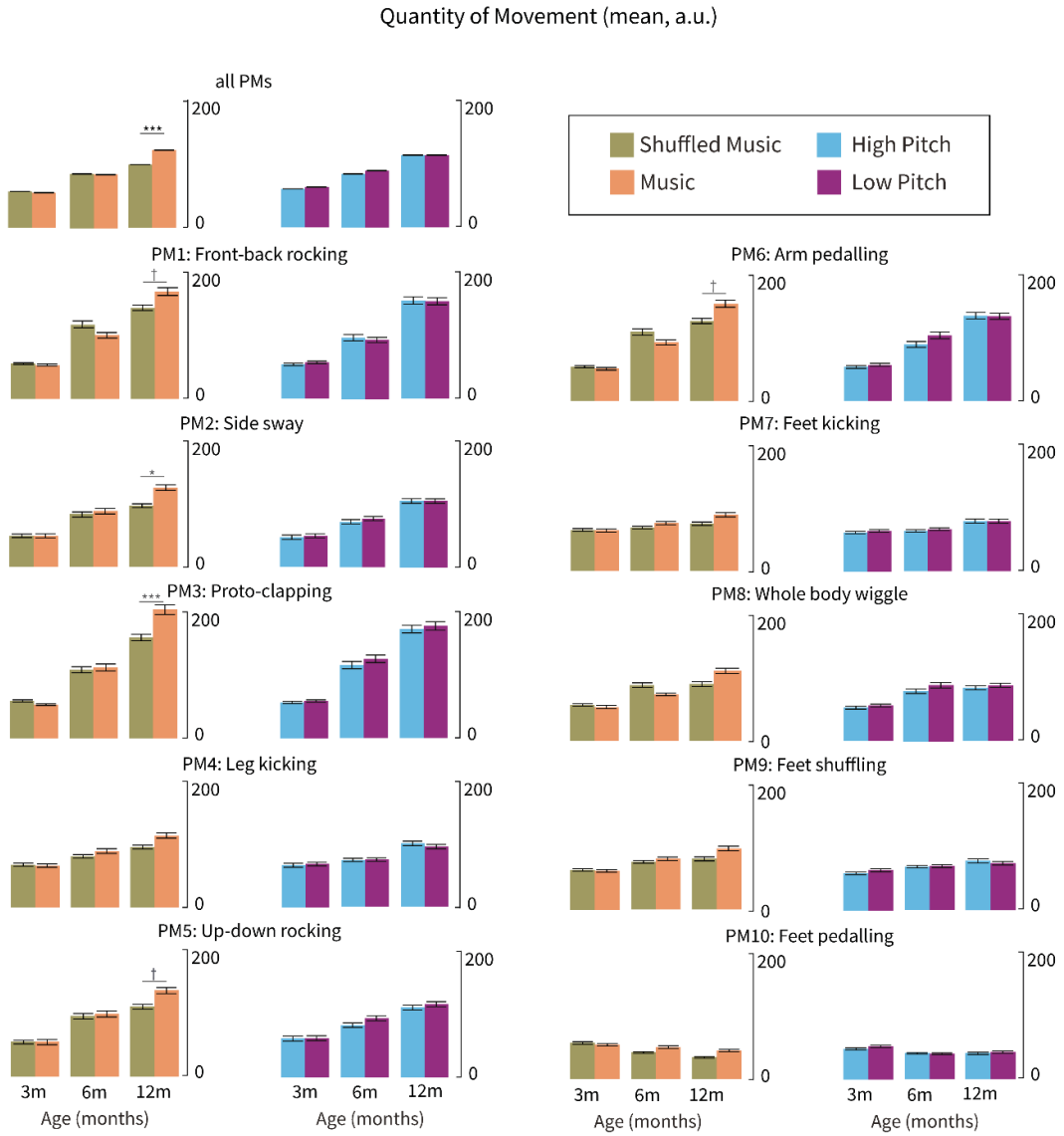


Figure 5. *Quantity of movement (mean, a.u.; [QoM]) elicited by music (orange) versus shuffled music (khaki) and high-pitch (blue) versus low-pitch music (purple) across different age groups (3-month-olds, 6-month-olds, 12-month-olds) and principal movements (PMs). Bar plots indicate the mean and standard error of QoM across different age groups, conditions, and PMs. Only twelve-month-old infants showed significantly increased QoM in response to music compared to shuffled music, specifically in PMs involving upper body movements (front-back rocking, side sway, proto-clapping, up-down rocking, and arm pedalling). No significant differences were observed between high- and low-pitch conditions. These results were also replicated in a supplementary analysis assessing differences in variance of (as opposed to mean) QoM (see Fig. S2 and Supplements for more details). † = p<.100, * = p<.050, ** = p<.010, *** = p<.001*

Movement: Granger Causality Analysis

Beyond looking at how much infants moved, we further investigated whether the spontaneous occurrence of infant movements could be explained by preceding changes in the intensity of the auditory stimuli. To do so, we used the sound envelope of the auditory stimuli (indexing changes in intensity over time) to predict infant movement velocity (time series representing changes in movement velocity, averaged across all PMs) and vice versa, using Granger-Causality analysis. Prediction estimates (Granger F-values) were yielded across different time lags, indexing the elapsed time necessary for a change in stimulus intensity to predict a change in movement velocity (i.e., highest F-values represent optimal prediction).

A preliminary analysis (sanity check) showed that musical stimuli predicted subsequent movement velocity better than vice versa across all age groups (*Fig. 6*, top-right corner; $F(1)=94.62$, $p<.001$). To identify the optimal temporal lags predicting sound-driven changes in movement, we conducted bootstrapped t-tests corrected for multiple comparisons across time points. Results (*Fig. 6*, left) showed that musical stimuli were better predictors of movement than shuffled stimuli at 3 months (lags of 160-200 ms, $t(40)=3.55$-3.98, $p<.001$), 6 months (lags of 120-240 ms, $t(37)=2.99$-4.49, $p<=.001$) and 12 months (lags of 160-240 ms, $t(31.6)=4.95$-5.03, $p<.001$). Conversely, there was a drop in prediction at later time lags in all groups (~320-360 ms; $t(43.6)=3.88$-8.25, $p<.001$). Together, these results suggest that infants' movement was related to intensity changes in the music, but not in the shuffled music.

Results associated with the high- and low-pitch conditions yielded similar Granger F-values to the music condition (*Fig. 6*, right). Notably, prediction estimates were generally higher for the high-pitch condition as compared to the low-pitch condition, indicating that high-pitch music was a better predictor of movement than low-pitch music. The difference between these two conditions was significant in one time-window in the 3-month-old infants (80-200 ms, $t(39.26)=2.66$-3.81, $p<.002$), while it encompassed two time windows in the 6-month-olds (120-200 ms, $t(45.64)=3.14$-4.21, $p<=.004$, and 320-360 ms, $t(45.7)=3.32$-3.82, $p<=.004$) and 12-month-olds (120-160ms, $t(40.4)=3.20$-3.73, $p<=.002$ and 520-600 ms, $t(39.9)=3.13$-3.41, $p<.004$) – perhaps suggesting that this effect grows stronger with age. Finally, Granger Causality statistics stratified by each PM and age group are detailed in the Supplements and *Fig. S2*.

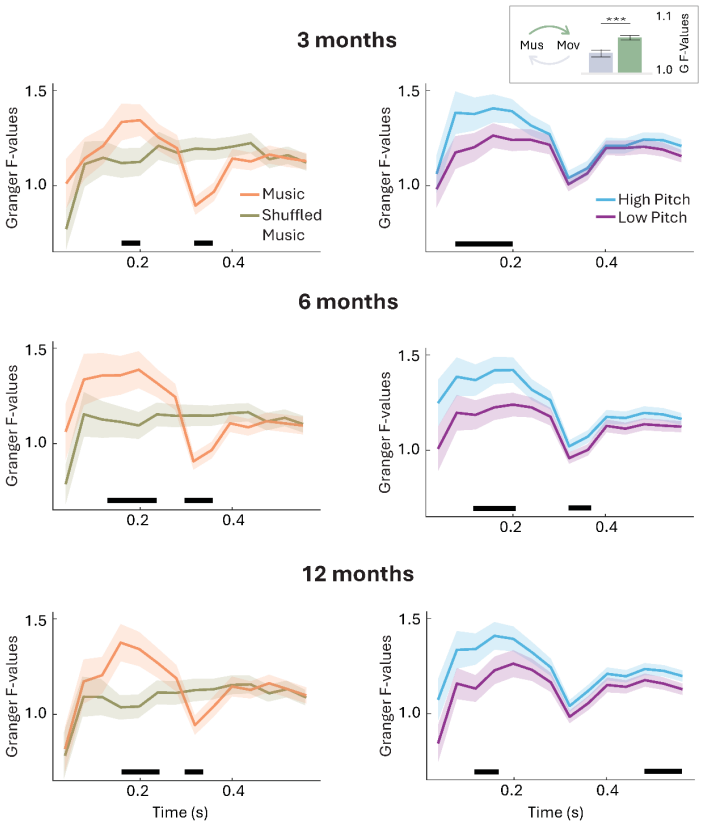


Figure 6. Music-driven movement *(Granger-Causality analysis). Top-right: A sanity check analysis showed that musical stimuli predicted subsequent movement velocity (green) better than vice versa (grey; p<.001). Left: Movement velocity was better predicted by music (orange) than by shuffled music (khaki), particularly with time lags of 160-200 ms (shaded areas indicate standard errors; horizontal black lines underline time ranges associated with a significant difference between conditions). Right: Movement was better predicted by high-pitch music (blue) compared to low-pitch music (purple).*

Movement: Phase-locked changes and periodicity

The Granger Causality analysis indicated that changes in music intensity drove movement in time, especially at a time lag comprised within a 200 ms delay. This result indirectly suggests that changes in music might evoke a phase-locked movement response. If so, we should be able to observe such phase-locked response when epoching movement data to peaks in the amplitude envelope of the auditory stimuli. To test this prediction, we ran supplementary event-related analyses on the movement velocity time series used for the previous analysis. Cluster-based permutation analyses revealed no significant clusters across age groups and conditions (*ps*>.050; see Supplements, *Fig. S3*), even though it should be

noted that 12-month-old infants exhibited a movement peak at ~200 ms, which was slightly but not significantly higher in response to music vs shuffled conditions. These results indicated that infants did not consistently exhibit phase-locked movement responses to musical events such as peaks in the amplitude envelope of the auditory stimuli.

Next, we also examined to what extent spontaneous movements were periodic and how so across conditions. This analysis builds upon the distribution of the highest coefficients yielded by auto-correlation analyses (similar to Zentner & Eerola, 2010). The results indicated that overall, spontaneous movements tended to be periodic, but such periodicity did not match the musical beat and was not different across conditions (see *Supplements*). In other words, movement periodicity did not result in coordination with music, and it was not modulated by whether infants were listening to music vs shuffled music or high- vs low-pitch music (*Fig. S4*). This result was further supported by an exploratory analysis in which we identified movement bursts by isolating epochs with high amplitude variation in the time series and excluding segments with low or no variation (c.f. Fujii et al., 2014). Together, these results suggest that while infants might generally exhibit rhythmic movements in response to sounds, the specificity of this behavioural response to music – and its coordination with musical structure – continues to develop beyond 12 months of age.

Discussion

This study examined the development of infants' neural and movement responses to music during the first year of life using a combination of neural measures, such as ERP and ASSR, alongside quantitative analyses of infants' movement. These approaches allowed us to explore both sensory and motor components of musicality, shedding light on how infants process and respond to musical stimuli. Below, we discuss 1) the development of neural auditory responses to music, 2) the emergence of music-driven movement patterns, and 3) the sensitivity of both components to pitch across early development.

Neural Responses to Music: Sensitivity to structured music across all ages

Much research suggests that the infants' auditory system is sensitive to structural features of music, such as timing and pitch regularities, from early on (Saffran et al., 1999; Thiessen & Saffran, 2009; see Nguyen et al., 2023 for a review). However, the developmental trajectory of such musical sensitivity across the first year of life is still underexplored. Here, we provide a detailed characterization of the progressive maturation of auditory evoked potentials, neurophysiological measures that generally capture rapid changes in the sensory

environment, such as the onset of musical notes (Kushnerenko et al., 2002; Somervail et al., 2021). Specifically, we observed ERPs with progressively shorter latency and larger amplitude throughout the first year of life. While these findings likely reflect a general maturation of sensory systems, they also highlight how the evoked potentials elicited by musical stimuli were enhanced in amplitude compared to those elicited by shuffled (i.e., structure-free) stimuli, an effect consistently observed across all age groups. This is in line with the idea that sensitivity to musical regularities emerges early and persists throughout infancy (Trevarthen, 1999). Indeed, sensitivity to simple rhythmic or pitch regularities has been observed in infants at various ages (Cirelli et al., 2016; Edalati et al., 2023; Flaten et al., 2022; Háden et al., 2009, 2024; Stefanics et al., 2009; Winkler et al., 2009), and, in the case of rhythm, even in premature neonates (Saadatmehr et al., 2025). Such sensitivity may be rooted in general auditory predictive processing, whereby the brain extracts regularities from past observations to generate and update predictions about incoming sensory information (Friston, 2010; Köster et al., 2020; Vuust & Witek, 2014). Ecologically valid music, such as the refrains used here, naturally incorporates rhythmic and pitch regularities that could trigger the generation of predictions – a process markedly dampened by shuffled music (Bianco et al., 2024, 2025; Lense et al., 2022).

Which specific musical regularities drove infants' predictions in our study? Previous EEG research suggests that in newborns, probabilistic auditory predictions are primarily driven by timing rather than pitch regularities when listening to ecologically valid music (Bianco et al., 2025). In contrast, adults rely on both timing and pitch regularities to generate predictions (Di Liberto et al., 2020). Given that our study examined infants older than newborns but not yet as mature as adults, it is likely that timing regularities played a key role for the youngest participants, while pitch regularities may have had a greater influence on the older age groups. However, it is important to note a limitation of such interpretation of our findings: our shuffling procedure permuted both the timing and pitch of the tones simultaneously, which prevents us from disentangling whether the stronger neural responses observed for the music condition reflect sensitivity to rhythmic regularity, melodic structure, or their combination. Hence, to clarify this issue, future research should investigate when, during development, the human brain begins incorporating pitch alongside timing regularities to generate musical predictions.

To explore the underlying neural mechanisms driving the above-described neural process, we analysed ASSR besides ERPs. This was done under the assumption that if the described results were driven by neural entrainment to the beat of the music, then ASSR

might better capture differences across conditions, as this measure can capture oscillatory activity. Instead, if the results were driven by evoked responses, then the ASSR results would generally align with ERP results or even be slightly less robust, given that frequency-domain analysis is less suited for capturing time-domain neural modulations. Results showed enhanced ASSR at a frequency matching the musical beat, with stronger power in response to music as opposed to shuffled stimuli and no differences between high- and low-pitch musical stimuli. Hence, ASSR and ERP results were very similar, a conclusion that is in line with previous accounts suggesting that on some occasions, the two measures might originate from the same signal, i.e., evoked responses (Capilla et al., 2011; Novembre & Iannetti, 2018). Further, building on evidence indicating that the infantile P1 originates from the auditory cortex (Chen et al., 2016; Musacchia et al., 2017; Riva et al., 2018), this result sheds light on one of several underlying neural structures that might be responsible for identifying musical structure in the auditory input.

Motor engagement with Music: Movement Sophistication develops with Age

In adults, music engages not only the auditory system but also the motor system (Fujioka et al., 2012; Grahn & Brett, 2007; Novembre & Keller, 2018; Phillips-Silver & Keller, 2012), often leading to spontaneous motor responses coordinated with the musical input (Hurley et al., 2014; Janata et al., 2012). While some evidence indicates that such spontaneous body movements are also exhibited by infants (Provasi et al., 2014), it remains unclear how this behaviour matures throughout development and how sophisticated it is.

First, as a coarse measure of auditory-motor coupling, we used Granger Causality to predict the time course of body movements using the time course of the auditory input (specifically the amplitude envelope). Taking this approach, we could predict body movements of infants from 3 to 12 months, specifically while listening to music as opposed to the shuffled conditions. This result indicates that recognising musical structures leads not only to distinct neural encoding patterns (as discussed in the previous section) but also to different levels of motor engagement. Further, as this result was observed across all ages, we might speculate that such audio-motor coupling, specifically triggered by music, emerges early in development and might be biologically predisposed. This conclusion aligns with previous findings showing that even foetuses as young as 28–35 weeks gestational age exhibit movement responses to musical sounds (Kisilevsky et al., 2014). However, it is important to note that these prenatal motor responses were not compared to motor responses to other auditory stimuli, as in our study, therefore raising questions about their specificity.

Next, taking a more complex measure of movement, we used principal component analysis to break down body kinematics into 10 independent principal movements (PMs; Bigand et al., 2024; Toiviainen et al., 2010) that explained nearly 80% of the whole kinematic variance. To the best of our knowledge, this approach has never been adopted for the analysis of infants' datasets, which normally do not distinguish between different kinds of movements (Fujii et al., 2014; Nguyen, Reisner, et al., 2023; Zentner & Eerola, 2010) or do so qualitatively (Thelen 1979, 1981). Building on this data-driven approach, we compared to what extent different movements were exhibited by infants across different ages and conditions. We found that only 12-month-old infants exhibited more movement in response to music compared to shuffled music. This result was driven by specific upper-body movements such as front-back rocking, side swaying, proto-clapping, and arm pedalling. It is not straightforward to interpret why these (and not other) movements were triggered by music, and why only at this age. Potential explanations include the development of refined postural control, typically achieved by 9–10 months (Hadders-Algra, 2005), and the fact that infants were tested in a seated position (a measure taken to simplify comparability across age groups). This seating arrangement likely facilitated upper body movements as the feet were resting. Further studies are needed to systematically characterize infants' movements to music across different contexts.

Finally, we examined movement coordination – a potential precursor of dance – and tested whether the periodicity of spontaneous movements matched the periodicity of the music. We did not find any evidence of movement coordination in none of the age groups. This result is in line with previous studies (Fujii et al., 2014; Zentner & Eerola, 2010) and suggests that sensorimotor transformation of music, specifically the ability to preserve its periodicity, develops after the first year of life. This delayed emergence may be due to motor control skills required for such transformation, which indeed are known to continue to mature during toddlerhood and even middle childhood (Kim & Schachner, 2022; Phillips-Silver et al., 2024). Hence, while coarse auditory-motor coupling is present at all ages, diversified movement patterns emerge only by 12 months, and spontaneous coordination with music likely continues to develop throughout infancy into childhood. Taking a broader perspective on infants' motor development, our findings align with research on locomotion across the first 14 months of life, which shows that as the number of motor primitives increases, their intrinsic variability decreases (Hinnekens et al., 2023). Viewed together, these patterns point toward a gradual refinement of motor control: the human motor system first develops the capacity to control individual muscles, and gradually the capacity to integrate them into

motor synergies that support complex, coordinated behaviours, such as locomotion, musical synchronization, and dance.

We suggest that the increasing complexity of infants' motor response to music is linked to the gradual maturation of the dorsal auditory stream, which connects posterior regions of the superior temporal gyrus with premotor cortices (e.g., Chen et al., 2009; Grahn & Brett, 2007; Kotz et al., 2018). Although this pathway is present at birth (Friederici, 2011; Perani et al., 2010), it has been suggested to play a crucial role in rhythmic entrainment and beat perception, functions that likely develop further as the pathway matures (Honing, 2018; Merchant & Honing, 2014; Patel & Iversen, 2014). This developmental trajectory aligns with neuroimaging evidence showing that while the ventral linguistic pathway (connecting temporal and frontal regions via the extreme capsule) is well-established at birth, the dorsal pathway—particularly the arcuate fasciculus connecting temporal regions to inferior frontal areas—continues maturing throughout the first postnatal months, with different maturational timelines for dorsal versus ventral connections (Dubois et al., 2016). Our suggestion is indirectly supported by comparative work on non-human primates, whose dorsal auditory stream is less developed than in humans (Merchant & Honing, 2014; Patel & Iversen, 2014). Indeed, research suggests that adult macaques struggle to recognise musical beats (Honing et al., 2012, 2018), while adult chimpanzees—though capable of adjusting their rhythmic sway periodicity in response to the beat (Hattori, 2021; Hattori & Tomonaga, 2020)—are unable to match it accurately, much like infants in our study and previous literature (Zentner & Eerola, 2010).

Pitch Sensitivity: Auditory and Movement preferences for High-Pitched music

At 6 months, infants exhibited enhanced neural responses to high-pitch music compared to low-pitch music, a specificity not observed at 12 months or in adults. This transient enhancement may reflect a critical period of heightened auditory plasticity, potentially supporting the detection of socially relevant stimuli (Fernald & Kuhl, 1987; Kuhl, 2010). Indeed, high-pitched sounds are prominent in infant-directed speech and singing, which are integral to caregiver-infant interactions and social communication during early infancy (Fernald & Kuhl, 1987; Hilton et al., 2022; Nakata & Trehub, 2011; Trainor & Zacharias, 1998; Tsang & Conrad, 2010). At six months, caregiver-infant face-to-face interactions peak (Beebe et al., 2016; Feldman, 2007), with infants relying more on behavioural cues, often conveyed through auditory modalities (Gratier et al., 2015; Nguyen, Zimmer, et al., 2023), and less on physical objects than in later developmental stages. Thus,

the enhanced sensitivity to high-pitched music at 6 months may reflect an essential phase in infants' developing abilities to process both musical and communicative signals (Shenfield et al., 2003). Toward the end of the first year, attentional focus widens (Cooper et al., 1997; Newman & Hussain, 2006) and face-to-face interactions decrease steeply (Jayaraman et al., 2015). At this stage, strengthening of lower-pitch processing likely leads to more balanced encoding of high- and low-pitch music by 12 months and beyond.

It should also be noted that our musical stimuli comprised polyphonic (two-voice) music, carrying sound frequencies falling within the typical range of infant-directed song (~200-400 Hz, Cirelli et al., 2020; Nguyen, Reisner, et al., 2023; Trainor & Zacharias, 1998). As such, our results might specifically speak for infants' ability to separate (and prioritize among) simultaneous communicative auditory streams (Marie & Trainor, 2013; Trainor, 2015). Indeed, other studies presenting one-voice pure tone sequences (single isochronous and isotonous tones) with high vs. low pitch - notably at frequencies outside our range (130 vs. 1237 Hz) - have reported stronger neural responses to relatively low frequencies (Lenc et al., 2023). Together, these contrasting observations suggest that pitch prioritization changes not only throughout development but also depends on the polyphonic complexity and spectral characteristics of the perceived stimuli. Further research might investigate this interesting issue further.

Moving to the movement results, it remains difficult to explain why high-pitch music was generally associated with enhanced auditory-motor coupling compared to low-pitch music. One possibility is that high-pitch music led to higher arousal, akin to infant-directed songs (Cirelli et al., 2020; Juslin & Laukka, 2003), which in turn strengthened the coupling between spontaneous movements and music. Alternatively, or complementarily, low-pitch music might have led to lower arousal or generally reduced attentional processes, thereby weakening auditory-motor coupling. If increased arousal were to result in greater overall movement, we would expect higher movement levels in the high pitch condition; however, this was not observed. QoM analyses based on the PMs did not reveal significant differences between the high pitch and low pitch conditions. This discrepancy may arise because Granger causality captures subtler temporal coupling between movement and music rather than gross movement quantity. Thus, high-pitch music may modulate the timing and coordination of motor responses without necessarily increasing the overall amount of movement. In line with prior work (e.g., Bigand et al., 2024), this interpretation emphasizes that musical coordination often involves changes in coupling strength rather than movement quantity per se. Either way, it is difficult to reconcile the movement results with the ERP results showing enhanced

auditory processing of high-pitch music only at 6 months. Perhaps, as infants mature, their sensitivity to high-pitched sounds may integrate with not only broader auditory but also motor processing mechanisms. Such a developmental trajectory might explain the reduced neural specificity to high-pitch music by 12 months, alongside a shift toward more coordinated motor responses to lower-pitched stimuli in adulthood (Cameron et al., 2022; Hove et al., 2014; Stupacher et al., 2016). These speculations could be addressed by future research exploring how musical pitch and rhythm interact to drive movement beyond infancy into adulthood to particularly examine *when* neural and motor engagement with low-pitch music becomes more prominent (Cameron et al., 2022; Stupacher et al., 2013, 2016).

Conclusion

Our study demonstrates that while robust auditory processing of music is present as early as 3 months, the translation of these sensory processes into organised motor behaviours unfolds gradually over and beyond the first year of life. Specifically, while coarse auditory-motor coupling is present at all ages, diversified movement patterns emerge only by 12 months, and spontaneous coordination with music likely develops throughout infancy and into childhood. Hence, our study provides evidence that, much like the auditory encoding of music, the propensity to move in response to music emerges early in development. This may reflect a biological or early-developing predisposition, eventually leading to dance-like behaviour. However, by 12 months, these motor responses remain relatively underdeveloped. Additionally, our study points to a previously unknown association between high-pitch music and auditory-motor coupling, extending the current research focus on infant-directed communication towards motor engagement and spontaneous behaviour. Together, these findings provide initial insights into how the developing brain gradually transforms music into spontaneous movements of increasing complexity. Future research should extend our characterisation of music-induced movement beyond the first year of life and further explore its (to-date mysterious) functional significance (Hoehl et al., 2021; Markova et al., 2019; Nguyen, Flaten, et al., 2023; Wass et al., 2020).

Materials & Methods

The study was preregistered: https://aspredicted.org/WW3_3TF. Deviations are listed in the Supplement.

Infant Experiment

Ninety-eight full-term infants, divided into three different age groups (see Fig. 1C), participated in this experiment. Nineteen infants were excluded due to fussiness (n=9) or technical issues (n=10), leading to a 19,6 % attrition rate (in line with infant EEG studies; Hoehl & Wahl, 2012). The remaining 79 infants were either 3 months (n=26, 14 girls, *mean*=113.04 days, *SD*=5.68 days, *range*=98-120 days), 6 months (n=26, 14 girls, *mean*=195.88 days, *SD*=9.46 days, *range*=182-211 days) or 12-13 months of age (n=27, 9 girls, *mean*=380.44 days, *SD*=14.93 days, *range*=361-413 days). All infants were born with a minimum gestational age of 37 weeks, weighed at least 2500 g at birth, and had no known developmental delays, neurological disorders, or hearing impairments. A primary caregiver gave written informed consent and was present during the experiment. The study was approved by the local Ethics committee of the University of Vienna (no. 00645) in line with the Declaration of Helsinki.

Stimuli

Auditory stimuli were generated by rearranging the refrains of two polyphonic children's songs ("La Vaca Lola" and "Hopp Juliska") using Logic Pro X (Apple, Inc.). Stimuli belonged to four distinct conditions (Fig. 1B). The music condition entailed the presentation of the rearranged refrains. The shuffled music condition entailed a disruption of pitch order and timing regularities of the refrains presented in the music condition. Specifically, the original temporal order of the notes was shuffled (disrupting pitch order), and the IOIs were replaced by a new pool of random values uniformly distributed around the original mean IOI ± 30-50% of the distance between the mean and the minimum original IOI (disrupting timing regularities). The bass and melody notes were treated independently in the shuffling process. In addition, the new pool of IOI was not quantized, so the notes did not fall on the beat, thus disrupting the isochrony of the original playsongs. In the high-pitch condition, the melody was shifted one octave higher than in the music condition. In the low-pitch condition, the bassline was shifted one octave lower than in the music condition. Hence, the two voices composing the high-pitch condition were one octave higher than those composing the low-pitch condition. All stimuli were built using the same musical

instruments: the melody was played by a flute ("VHS Flute" from Yamaha DX7, Arturia) and the bassline was played by an electric bass ("Liverpool" electric bass from Logic Pro), with almost all bassline notes (15 out of 16) falling on the beat. Importantly, all stimuli had the same duration (21 s), tempo (135 BPM), tonality (C-major) and loudness (which stayed within a range of 1.5 LUFS, a measure that considers the sensitivity of the human auditory system across frequencies).

Procedure

Setup and Task

Infants were seated in an infant highchair equipped with an age-appropriate baby seat facing a computer screen (1 m distance, see Fig. 1A). Parents sat next to the infants (on their right and slightly towards the back). The stimuli were presented free field (~60 dB SPL) from two audio speakers (Q Acoustics 3020i) placed on either side of the monitor, facing the infants. During the procedure, infants passively listened to a series of auditory stimuli in the following sequence: First, there were 10 seconds of silence, followed by the shuffled music condition. Then, the music, high-pitch and low-pitch conditions were played in random order. After that, there were another 10 seconds of silence, followed by a second round of the music, high-pitch and low-pitch conditions in random order. Finally, the shuffled music condition was played one last time, followed by 10 seconds of silence. The same song was used in all the different conditions within each presentation block. The order of presentation blocks for each song was counterbalanced. Meanwhile, infants were watching a silent movie of blooming flowers, slowly fading in and out (image duration 10 s). Videos were included to keep infants calm and engaged. Only infants who remained calm for at least two presentation blocks were included in the final sample (n=3 were excluded). Stimulus presentation and the synchronization between EEG and video cameras were controlled in Presentation (Version 23.0, Neuro-behavioral System, Berkeley, CA) utilizing triggers sent at the start of the experiment and the beginning of each trial.

Electroencephalography (EEG). We recorded EEG using a Brain Products BrainAmp DC system with 32 active Ag-AgCl electrodes, mounted on infant-sized EEG caps (Acticap) according to the international 10-10 system. The reference electrode was placed at TP9 (left mastoid), and the ground electrode was placed at Fp1. EEG data were sampled at 1000 Hz. Impedances were below 20 kΩ at the beginning of the experiment.

Video Recording. Infant body movement was recorded using three video cameras (Axis Communication, Lund) sampling at 25 Hz with a resolution of 1920 x 1080 pixels. The

three video recordings, taken from frontal (0°), diagonal (45°) and side view (90°) perspectives, were synchronized in VideoSyncPro (Mangold International). The cameras were positioned at the height of the infants' heads to capture their whole bodies.

Adult Control Experiment.

Twenty-six healthy young adults (11 female, *mean* = 27.55 years of age, *SD*: 7.09 years, *range*=19-47 years) participated in the experiment. The study was approved by the Regional Ethics Committee of Liguria (no. 794/2021). Participants underwent the same experimental procedure as the infants, including the same auditory stimuli and video material. EEG was recorded with 64 Ag-AgCl electrodes, closely matched to the infant configuration (see Supplements for further details) and sampled at 1024 Hz. Data were down-sampled to 1000 Hz to match the pre-processing and data analysis procedure (detailed below) of the infant study. The experiment was conducted to provide a ground truth for the interpretation of the infants' EEG data. Therefore, no video data was recorded to track adults' movements.

Data Processing

EEG pre-processing. Infant EEG data are notoriously noisy as infants may not understand or comply with task instructions. It follows that their behaviour can lead to artefacts in the EEG recordings. To mitigate this issue, we used a combination of open-access denoising algorithms and adopted a fully data-driven pre-processing pipeline that we previously developed to denoise EEG data recorded from awake monkeys (Bianco et al., 2024) and humans dancing (Bigand, Bianco, Abalde, Nguyen, et al., 2024). The pipeline relied on a combination of MATLAB functions developed for toolboxes such as Fieldtrip (Oostenveld et al., 2010) and eeglab (Delorme & Makeig, 2004). Continuous EEG data were band-pass filtered (0.3 Hz - 30 Hz, 3rd order Butterworth filter, zero-phase) and segmented into trials starting 3 seconds before song onset and ending 3 seconds post song offset. Next, noisy and faulty electrodes were identified and rejected by assessing flat-lining for over 5 s (function *clean_flatlines*), correlations between electrodes for $r<.1$ or line noise above 20 SD from the mean of all electrodes (function *clean_channels)*. These initial steps were conducted to find a rough indication of electrodes carrying low-quality signals. In addition, mean, standard deviation, and peak-to-peak distances were calculated over time for each electrode. If any of those variables exceeded 2.5 SD from the mean of the other electrodes, that electrode was (provisionally) discarded. This process was iterated without the outlier electrode(s) until a distribution without outliers was found. Electrodes were then re-

referenced to the averaged mastoids (TP9, TP10) or the left mastoid only in cases when TP10 was noisy (N=26). We then used artifact subspace reconstruction (ASR, threshold 5) to remove artefactual EEG activity (Kothe & Jung, 2015). An automatic independent component analysis (ICA) procedure using ICLabel was implemented, and components that were classified as eye components with a minimum probability of 50 % were rejected (mean=0.76, SD=0.67, 0-2 components per infant). Next, previously excluded electrodes were interpolated by using the neighbouring electrodes (spherical spline; mean number of interpolated channels=6, SD=3.09). We assessed data quality (including trial counts and signal-to-noise ratio [SNR]) across age groups and confirmed that these metrics were comparable (see Supplementary Materials for details).

Video data pre-processing. We extracted infants' movement time series by using DeepLabCut (version 2.2.3; Mathis et al., 2018; Nath et al., 2019) in Python on videos filming the infant from the front (see *Fig. 1*). A trained coder (T.N.) labelled 18 body parts per individual (left eye, right eye, mouth, left shoulder, right shoulder, chest, left elbow, left wrist, left hand, right elbow, right wrist, right hand, left knee, left ankle, left foot, right knee, right ankle, right foot) in 10 frames randomly taken from the videos of each infant in each age group (n=24 [3m], 25 [6m] and 24 [12m]; 80% were used for training). We used a ResNet-50 neural network with default parameters for single subject detection for around 5-17 training iterations (correcting labels of 5 frames per video of age groups) until test error plateaued. The test error was 8.27 (3-month-olds), 10.19 (6-month-olds), and 11.89 (12-month-olds) pixels (image size was 1920 by 1080 pixels). The networks of each age group were used to analyse videos from the same age group. The (x and y) coordinates were then pre-processed using a custom code written in Python. Coordinates that were estimated with likelihood values 2 SD away from the mean probability per age group were set to NaN (i.e., not a number). These gaps in the coordinates of body parts were then recovered from predefined neighbouring body parts in a forward and backward approach (from head to feet, followed by feet to head). For example, if the left wrist was missing at t_0, it was recovered using the position of the left elbow or the left hand at t_0. Next, we assessed the correctness of the estimated body part displacements, i.e., demeaned coordinates. Specifically, we set the coordinates (either x or y) of each body part below or above 3 SD from the mean of all trials of each infant to NaN (outlier procedure). To further verify left- and right-hand displacements, we checked the distance between the wrist and hand, specifically using the

same outlier procedure. The remaining NaN values were replaced using linear interpolation for each body part coordinate in time.

Principal movements. We used Principal Component Analysis (PCA) to reduce dimensionality of the pre-processed movement data and at the same time to extract a set of interpretable Principal Movements (PM) that generalize across trials, conditions, and all infants (Bigand, Bianco, Abalde, & Novembre, 2024). To ensure that PCA captured only common movements across participants, rather than postural or anthropometric differences between them, we demeaned and standardized the body part vectors for each trial. Specifically, we subtracted the mean body part vector (averaged over time) from the body part vectors at each frame. Then, these demeaned body part vectors were divided by a global measure of standard deviation across time, which was computed by combining all body parts to preserve the variance differences between body parts. This allowed us to concatenate all trials from all participants into a single data matrix, with each trial contributing equally to the variance of the pooled matrix. PMs were derived from the eigenvectors, and their corresponding scores obtained from the PCA applied to the data matrix:

$$PM(t) = c_i(t)\, w_i$$

where $c_i(t)$ is the i^{th} PM score at time t and w_i is the i^{th} PM weight vector, or eigenvector (1 × 36). Each of the PMs reflects the trajectory of covarying body parts. We extracted 10 PMs that explained 79.7 % of the whole kinematic variance (Fig. 3). It is important to note that the PCA does not directly extract "real movements" but rather movement dimensions with large variances. PMs' time series (score of each principal component) were low-pass filtered at 10 Hz. Next, we calculated the time series' first-degree derivative to extract velocities and the absolute values of velocities to calculate movement quantities (e.g., Bigand et al., 2024; Górecki & Łuczak, 2013). Velocities include information on movement direction (below or above 0) and are, therefore, particularly useful to analyse correlations with other streams, such as music (detailed in the section on Granger Causality). Movement quantities only consider the amount of movement and can contribute to understanding descriptive differences between conditions, regardless of the movement direction.

Data analyses

EEG: Event-related potentials (ERP) analysis. EEG data were first lowpass filtered at 20 Hz (FIR-type filter) to specifically retain the part of the EEG signal carrying event-related potentials. EEG data were epoched according to the onsets of the bass notes of the

musical stimuli (which fell on the beat of the music in 15 out of 16 notes). Epochs included 600 ms of data (-100 ms to +500 ms relative to tone onset – note that notes occurred every 444 ms, so the last 56 ms of each epoch refers to the next tone). In preparation for further analysis, epochs were automatically examined to check if they were still contaminated by artifacts. Epochs were automatically excluded if the standard deviation of the analysed channels exceeded a voltage threshold of 100 μV within a 200 ms interval. All infants and adults contributed at least 165 clean epochs per condition (3m: *M*=340.09, *SD*=138.96; 6m: *M*=373.69, *SD*=144.07; 12m: *M*=361.47, *SD*=157.46; adults: *M*=399.95, *SD*=8.50). Epochs were baseline-corrected by subtracting the average voltage between -100 and 0 ms relative to the tone onset. Epochs belonging to the same experimental condition were averaged together. Event-related EEG amplitude modulations were compared using cluster-based permutation analyses (n_{Perm}=1000; Maris & Ostenveld, 2007). Clusters were based on both temporal consecutiveness and spatial adjacency of EEG electrodes (3 cm distance). A cluster had to be composed of at least two consecutive time points with a *p-value* <.05 on at least two neighbouring EEG electrodes. The clusters were tested using a *p-value* of <.05 and one-tailed according to the above-detailed hypotheses.

In additional analyses, we extracted the amplitude and latencies of the individual peak of the P1 component. Using two linear mixed-effects models, we tested whether amplitudes (dependent variable) and latencies (dependent variable) were dependent on age and condition (fixed and interaction effect) while assuming a random intercept for each participant.

EEG: Auditory Steady State Response (ASSR) analysis. EEG data were segmented according to each condition (21 seconds long, 3 repetitions of the refrain) and averaged over segments of the same condition to enhance the ASSR (see Cirelli et al., 2016). Each segment was submitted to single-taper frequency transformation using a Hilbert taper to estimate frequencies between 0.5 and 5 Hz, with a frequency resolution of 0.25 Hz. Activity unrelated to the frequency of interest was removed by subtracting the average amplitude measured at neighbouring frequency bins (i.e., −5 to −3 bins and +3 to +5 bins). Performing this subtraction removes residual background noise around the frequency bin of interest, leaving only the activity directly related to the ASSR. Next, we extracted amplitudes at the beat-related (2.25 Hz) frequencies and compared these across conditions.

Movement: Granger Causality analysis. We conducted a Granger Causality analysis in R (function = *causality,* library = *vars)* to measure to what extent the envelope of the auditory stimuli predicted movement velocity over time and across conditions (see also Klein et al., 2022). To do so, we extracted the broadband amplitude envelope of the auditory

stimuli using Hilbert transform. We then conducted Granger Causality analyses on amplitude envelopes predicting the PM velocity time series across different model orders (lags) ranging from 40 ms to 1000 ms. Granger Causality is based on the concept of prediction in the sense that if a time series X_t "Granger-causes" another time series Y_t, then past values of X_t contain information that helps predict Y_t beyond the information contained in past values of Y_t alone. Granger Causality is particularly useful for capturing the temporal relationship between different time series that are characterized by sudden and transient bursts in activity (such as infant rhythmic movement; Thelen, 1979). As a sanity check, we also conducted a reverse analysis predicting music (i.e., the amplitude envelope of the auditory stimuli) from the PM velocity time series and verified that the resulting F-Values (averaged across all lags: 40 – 1000 ms) were statistically lower compared to those yielded by the former (music-to-movement) analysis. Indeed, the sanity check analysis reveals that predictions from audio stimuli to PM velocity time series (M=1.08, SE=0.009) were significantly higher than from PMs to audio stimuli (M=1.03, SE=0.009; $\chi^2(1)=1669.2$, $p<.001$, see *Fig. 5*).

Movement: Event-related analysis. To examine whether the events (i.e., musical notes) comprised within the auditory stimuli would trigger event-related movements, we used an event-related approach (mirroring the ERP analysis discussed above). First, we segmented epochs of 700 ms (between -100 ms and +600 ms relative to peaks in the amplitude). Epochs were then averaged for each condition. Detailed results for this analysis are reported in the Supplements (see *Fig. S2*).

Movement: Periodicity analysis. We analysed the movement periodicity using autocorrelation on the same epochs employed in the event-related movement analysis. This means we estimated the autocorrelation of the movement time series in a running window of 3.5 s (length of 1 bar) with 25% overlap, mirroring the approach by Zenter & Eerola (2010). This approach provides an estimation of the periodicity of the PMs. We extracted the local maxima of the autocorrelation values in each window and fitted a normal distribution (Kernel density estimation) to the collection of these values for each condition. We used a bin width of 0.08 and estimated values between 300 and 600 ms according to the beat-related lag in our auditory stimuli. We then compared the beat-related lag (444 ms lag) across conditions. The results are reported in the Supplements (see *Fig. S3*).

Statistical analyses.

Analyses were conducted in R and included non-parametric tests or linear mixed-effects models (LMM; lme4 package). Each participant was modelled as a random intercept in all

LMMs. Statistical significance was evaluated by likelihood-ratio tests conducted using *Anova* (car package), and post-hoc contrasts were run using *emmeans* (emmeans package). We controlled for increased Type I error from multiple comparisons, when necessary, using the false-discovery rate (Benjamini & Hochberg, 1995). The statistical significance level was set to $\alpha = 0.05$.

References

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother-infant face-to-face communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085

Bianco, R., Tóth, B., Bigand, F., Nguyen, T., Sziller, I., Háden, G. P., Winkler, I., & Novembre, G. (2025). *Human newborns form musical predictions based on rhythmic but not melodic structure* (p. 2025.02.19.639016). bioRxiv. https://doi.org/10.1101/2025.02.19.639016

Bianco, R., Zuk, N. J., Bigand, F., Quarta, E., Grasso, S., Arnese, F., Ravignani, A., Battaglia-Mayer, A., & Novembre, G. (2024). Neural encoding of musical expectations in a non-human primate. *Current Biology*, *34*(2), 444-450.e5. https://doi.org/10.1016/j.cub.2023.12.019

Bigand, F., Bianco, R., Abalde, S. F., Nguyen, T., & Novembre, G. (2024). *Imaging the dancing brain: Decoding sensory, motor and social processes during dyadic dance* (p. 2024.12.17.628913). bioRxiv. https://doi.org/10.1101/2024.12.17.628913

Bigand, F., Bianco, R., Abalde, S. F., & Novembre, G. (2024). The geometry of interpersonal synchrony in human dance. *Current Biology*, *34*(13), 3011-3019.e4. https://doi.org/10.1016/j.cub.2024.05.055

Brown, S. (2022). Group dancing as the evolutionary origin of rhythmic entrainment in humans. *New Ideas in Psychology*, *64*, 100902. https://doi.org/10.1016/j.newideapsych.2021.100902

Cameron, D. J., Dotov, D., Flaten, E., Bosnyak, D., Hove, M. J., & Trainor, L. J. (2022). Undetectable very-low frequency sound increases dancing at a live concert. *Current Biology*, *32*(21), R1222–R1223. https://doi.org/10.1016/j.cub.2022.09.035

Capilla, A., Pazo-Alvarez, P., Darriba, Á., Campo, P., & Gross, J. (2011). Steady-State Visual Evoked Potentials Can Be Explained by Temporal Superposition of Transient Event-Related Responses. *PloS One*, *6*, e14543. https://doi.org/10.1371/journal.pone.0014543

Chen, A., Peter, V., & Burnham, D. (2016). Auditory ERP response to successive stimuli in infancy. *PeerJ*, *4*, e1580. https://doi.org/10.7717/peerj.1580

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2009). The Role of Auditory and Premotor Cortex in Sensorimotor Transformations. *Annals of the New York Academy of Sciences*, *1169*(1), 15–34. https://doi.org/10.1111/j.1749-6632.2009.04556.x

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of Maternal Singing Style on Mother–Infant Arousal and Behavior. *Journal of Cognitive Neuroscience*, *32*(7), 1213–1220. https://doi.org/10.1162/jocn_a_01402

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring Neural Entrainment to Beat and Meter in Infants: Effects of Music Background. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00229

Conrad, N. J., Walsh, J., Allen, J. M., & Tsang, C. D. (2011). Examining infants' preferences for tempo in lullabies and playsongs. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *65*(3), 168–172. https://doi.org/10.1037/a0023296

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488. https://doi.org/10.1016/S0163-6383(97)90037-0

Costa-Giomi, E., & Sun, X. (2016). Infants' Home Soundscape: A Day in the Life of a Family. In *Contemporary Research in Music Learning Across the Lifespan* (pp. 87–96). Routledge.

Damsma, A., Roo, M., Doelling, K., Bazin, P.-L., & Bouwer, F. (2024). *Tempo-dependent selective enhancement of neural responses at the beat frequency can be explained by both an oscillator and an evoked model*. https://doi.org/10.1101/2024.07.11.603023

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Di Liberto, G. M., Pelofi, C., Bianco, R., Patel, P., Mehta, A. D., Herrero, J. L., de Cheveigné, A., Shamma, S., & Mesgarani, N. (2020). Cortical encoding of melodic expectations in human temporal cortex. *eLife*, *9*, e51784. https://doi.org/10.7554/eLife.51784

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., Leroy, F., Hertz-Pannier, L., & Dehaene-Lambertz, G. (2016). Exploring the Early Organization and Maturation of Linguistic Pathways in the Human Infant Brain. *Cerebral Cortex*, *26*(5), 2283–2298. https://doi.org/10.1093/cercor/bhv082

Eckerdal, P., & Merker, B. (2009). 'Music' and the 'action song' in infant development: An interpretation. In *Communicative musicality: Exploring the basis of human companionship* (pp. 241–262). Oxford University Press.

Edalati, M., Wallois, F., Safaie, J., Ghostine, G., Kongolo, G., Trainor, L. J., & Moghimi, S. (2023). Rhythm in the premature neonate brain: Very early processing of auditory beat and meter. *The Journal of Neuroscience*, JN-RM-1100-22. https://doi.org/10.1523/JNEUROSCI.1100-22.2023

Feldman, R. (2007). Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345. https://doi.org/10.1111/j.1467-8721.2007.00532.x

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140091. https://doi.org/10.1098/rstb.2014.0091

Flaten, E., Marshall, S. A., Dittrich, A., & Trainor, L. J. (2022). Evidence for top‐down metre perception in infancy as shown by primed neural responses to an ambiguous rhythm. *European Journal of Neuroscience*, ejn.15671. https://doi.org/10.1111/ejn.15671

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Fujii, S., Watanabe, H., Oohashi, H., Hirashima, M., Nozaki, D., & Taga, G. (2014). Precursors of Dancing and Singing to Music in Three- to Four-Months-Old Infants. *PLOS ONE*, *9*(5), e97680. https://doi.org/10.1371/journal.pone.0097680

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. *The Journal of Neuroscience*, *32*(5), 1791–1802. https://doi.org/10.1523/JNEUROSCI.4107-11.2012

Górecki, T., & Łuczak, M. (2013). Using derivatives in time series classification. *Data Mining and Knowledge Discovery*, *26*, 310–331. https://doi.org/10.1007/s10618-012-0251-4

Grahn, J. A., & Brett, M. (2007). Rhythm and Beat Perception in Motor Areas of the Brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167. https://doi.org/10.3389/fpsyg.2015.01167

Hadders-Algra, M. (2005). Development of Postural Control During the First 18 Months of Life. *Neural Plasticity*, *12*(2–3), 695071. https://doi.org/10.1155/NP.2005.99

Háden, G. P., Bouwer, F. L., Honing, H., & Winkler, I. (2024). Beat processing in newborn infants cannot be explained by statistical learning based on transition probabilities. *Cognition*, *243*, 105670. https://doi.org/10.1016/j.cognition.2023.105670

Háden, G. P., Németh, R., Török, M., & Winkler, I. (2015). Predictive processing of pitch trends in newborn infants. *Brain Research*, *1626*, 14–20. https://doi.org/10.1016/j.brainres.2015.02.048

Háden, G. P., Stefanics, G., Vestergaard, M. D., Denham, S. L., Sziller, I., & Winkler, I. (2009). Timbre-independent extraction of pitch in newborn infants. *Psychophysiology*, *46*(1), 69–74. https://doi.org/10.1111/j.1469-8986.2008.00749.x

Hattori, Y. (2021). Behavioral Coordination and Synchronization in Non-human Primates. In J. R. Anderson & H. Kuroshima (Eds), *Comparative Cognition: Commonalities and Diversity* (pp. 139–151). Springer. https://doi.org/10.1007/978-981-16-2028-7_9

Hattori, Y., & Tomonaga, M. (2020). Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). *Proceedings of the National Academy of Sciences*, *117*(2), 936–942. https://doi.org/10.1073/pnas.1910318116

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01410-x

Hinnekens, E., Barbu-Roth, M., Do, M.-C., Berret, B., & Teulier, C. (2023). Generating variability from motor primitives during infant locomotor development. *eLife*, *12*, e87463. https://doi.org/10.7554/eLife.87463

Hoehl, S., Fairhurst, M., & Schirmer, A. (2021). Interactional synchrony: Signals, mechanisms and benefits. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 5–18. https://doi.org/10.1093/scan/nsaa024

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187–209. https://doi.org/10.1080/87565641.2011.627958

Honing, H. (2018). *The Origins of Musicality*. MIT Press.

Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018). Rhesus Monkeys (Macaca mulatta) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00475

Honing, H., Merchant, H., Háden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus Monkeys (Macaca mulatta) Detect Rhythmic Groups in Music, but Not the Beat. *PLoS ONE*, *7*(12), e51369. https://doi.org/10.1371/journal.pone.0051369

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140088. https://doi.org/10.1098/rstb.2014.0088

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences*, *111*(28), 10383–10388. https://doi.org/10.1073/pnas.1402039111

Hurley, B. K., Martens, P. A., & Janata, P. (2014). Spontaneous sensorimotor coupling with multipart music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1679–1696. https://doi.org/10.1037/a0037154

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54–75. https://doi.org/10.1037/a0024208

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The Faces in Infant-Perspective Scenes Change over the First Year of Life. *PLOS ONE*, *10*(5), e0123780. https://doi.org/10.1371/journal.pone.0123780

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kim, M., & Schachner, A. (2022). The origins of dance: Characterizing the development of infants' earliest dance behavior. *Developmental Psychology*. https://doi.org/10.1037/dev0001436

Kisilevsky, B. s., Hains, S. m. j., Jacquet, A.-Y., Granier-Deferre, C., & Lecanuet, J. p. (2004). Maturation of fetal responses to music. *Developmental Science*, *7*(5), 550–559. https://doi.org/10.1111/j.1467-7687.2004.00379.x

Klein, L., Wood, E. A., Bosnyak, D., & Trainor, L. J. (2022). Follow the sound of my violin: Granger causality reflects information flow in sound. *Frontiers in Human Neuroscience*, *16*, 982177. https://doi.org/10.3389/fnhum.2022.982177

Köster, M., Kayhan, E., Langeloh, M., & Hoehl, S. (2020). Making Sense of the World: Infant Learning From a Predictive Processing Perspective. *Perspectives on Psychological Science*, *15*(3), 562–571. https://doi.org/10.1177/1745691619895071

Kothe, C. A. E., & Jung, T.-P. (2015). *Artifact removal techniques with signal reconstruction* (World Intellectual Property Organization Patent No. 2015/047462 A9). https://patents.google.com/patent/WO2015047462A9/en

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1), 8537. https://doi.org/10.1038/ncomms9537

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.neuron.2010.08.038

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science (New York, N.Y.)*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport*, *13*(1), 47–51. https://doi.org/10.1097/00001756-200201210-00014

Lenc, T., Peter, V., Hooper, C., Keller, P. E., Burnham, D., & Nozaradan, S. (2023). Infants show enhanced neural responses to musical meter frequencies beyond low-level features. *Developmental Science*, *26*(5), e13353. https://doi.org/10.1111/desc.13353

Lense, M. D., Shultz, S., Astésano, C., & Jones, W. (2022). Music of infant-directed singing entrains infants' social visual behavior. *Proceedings of the National Academy of Sciences*, *119*(45), e2116967119. https://doi.org/10.1073/pnas.2116967119

Lijffijt, M., Moeller, F. G., Boutros, N. N., Burroughs, S., Lane, S. D., Steinberg, J. L., & Swann, A. C. (2009). The Role of Age, Gender, Education, and Intelligence in P50, N100, and P200 Auditory Sensory Gating. *Journal of Psychophysiology*, *23*(2), 52–62. https://doi.org/10.1027/0269-8803.23.2.52

Marie, C., & Trainor, L. J. (2013). Development of Simultaneous Pitch Encoding: Infants Show a High Voice Superiority Effect. *Cerebral Cortex*, *23*(3), 660–669. https://doi.org/10.1093/cercor/bhs050

Marie, C., & Trainor, L. J. (2014). Early development of polyphonic sound encoding and the high voice superiority effect. *Neuropsychologia*, *57*, 50–58. https://doi.org/10.1016/j.neuropsychologia.2014.02.023

Markova, G., Nguyen, T., & Hoehl, S. (2019). Neurobehavioral Interpersonal Synchrony in Early Development: The Role of Interactional Rhythms. *Frontiers in Psychology*, *10*, 2078. https://doi.org/10.3389/fpsyg.2019.02078

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), Article 9. https://doi.org/10.1038/s41593-018-0209-y

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, *7*, 274. https://doi.org/10.3389/fnins.2013.00274

Musacchia, G., Ortiz-Mantilla, S., Choudhury, N., Realpe-Bonilla, T., Roesler, C., & Benasich, A. A. (2017). Active auditory experience in infancy promotes brain plasticity in Theta and Gamma oscillations. *Developmental Cognitive Neuroscience*, *26*, 9–19. https://doi.org/10.1016/j.dcn.2017.04.004

Nakata, T., & Trehub, S. E. (2011). Expressive timing and dynamics in infant-directed and non-infant-directed singing. *Psychomusicology: Music, Mind and Brain*, *21*(1–2), 45–53. https://doi.org/10.1037/h0094003

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Nave, K. M., Hannon, E. E., & Snyder, J. S. (2022). Steady state-evoked potentials of subjective beat perception in musical rhythms. *Psychophysiology*, *59*(2), e13963. https://doi.org/10.1111/psyp.13963

Newman, R. S., & Hussain, I. (2006). Changes in Preference for Infant-Directed Speech in Low and Moderate Noise by 4.5- to 13-Month-Olds. *Infancy*, *10*(1), 61–76. https://doi.org/10.1207/s15327078in1001_4

Nguyen, T., Flaten, E., Trainor, L. J., & Novembre, G. (2023). *Early social communication through music: State of the art and future perspectives* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/j5g69

Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313

Nguyen, T., Zimmer, L., & Hoehl, S. (2023). Your turn, my turn. Neural synchrony in mother–infant proto-conversation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), 20210488. https://doi.org/10.1098/rstb.2021.0488

Novembre, G., & Iannetti, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences*, *115*(47), E11002–E11003. https://doi.org/10.1073/pnas.1815311115

Novembre, G., & Keller, P. E. (2018). Music and Action. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology* (pp. 523–537). Springer. https://doi.org/10.1007/978-3-662-55004-5_28

Novembre, G., Pawar, V. M., Bufacchi, R. J., Kilintari, M., Srinivasan, M., Rothwell, J. C., Haggard, P., & Iannetti, G. D. (2018). Saliency Detection as a Reactive Process: Unexpected Sensory Events Evoke Corticomuscular Coupling. *The Journal of Neuroscience*, *38*(9), 2385–2397. https://doi.org/10.1523/JNEUROSCI.2474-17.2017

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnsys.2014.00057

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., & Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, *107*(10), 4758–4763. https://doi.org/10.1073/pnas.0909074107

Phillips-Silver, J., Hartmann, M., Fernández-García, L., Maurno, N. C. G., Toiviainen, P., & González, M. T. D. (2024). Development of full-body rhythmic synchronization in middle childhood. *Scientific Reports*, *14*(1), 15741. https://doi.org/10.1038/s41598-024-66438-7

Phillips-Silver, J., & Keller, P. (2012). Searching for Roots of Entrainment and Joint Action in Early Musical Interactions. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00026

Pratt, H., Starr, A., Michalewski, H. J., Bleich, N., & Mittelman, N. (2008). The auditory P50 component to onset and offset of sound. *Clinical Neurophysiology*, *119*(2), 376–387. https://doi.org/10.1016/j.clinph.2007.10.016

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology*, *5*, 1048.

Remijn, G. B., Hasuo, E., Fujihira, H., & Morimoto, S. (2014). An introduction to the measurement of auditory event-related potentials (ERPs). *Acoustical Science and Technology*, *35*(5), 229–242. https://doi.org/10.1250/ast.35.229

Riva, V., Cantiani, C., Mornati, G., Gallo, M., Villa, L., Mani, E., Saviozzi, I., Marino, C., & Molteni, M. (2018). Distinct ERP profiles for auditory processing in infants at-risk for autism and language impairment. *Scientific Reports*, *8*(1), 715. https://doi.org/10.1038/s41598-017-19009-y

Saadatmehr, B., Edalati, M., Wallois, F., Ghostine, G., Kongolo, G., Flaten, E., Tillmann, B., Trainor, L., & Moghimi, S. (2025). Auditory Rhythm Encoding during the Last Trimester of Human Gestation: From Tracking the Basic Beat to Tracking Hierarchical Nested Temporal

Structures. *Journal of Neuroscience*, *45*(4). https://doi.org/10.1523/JNEUROSCI.0398-24.2024
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4
Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, *19*(10), 831–836. https://doi.org/10.1016/j.cub.2009.03.061
Somervail, R., Zhang, F., Novembre, G., Bufacchi, R. J., Guo, Y., Crepaldi, M., Hu, L., & Iannetti, G. D. (2021). Waves of Change: Brain Sensitivity to Differential, not Absolute, Stimulus Intensity is Conserved Across Humans and Rats. *Cerebral Cortex (New York, N.Y.: 1991)*, *31*(2), 949–960. https://doi.org/10.1093/cercor/bhaa267
Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, *120*(2), 304–308. https://doi.org/10.1016/j.clinph.2008.11.020
Stupacher, J., Hove, M. J., & Janata, P. (2016). Audio Features Underlying Perceived Groove and Sensorimotor Synchronization in Music. *Music Perception*, *33*(5), 571–589. https://doi.org/10.1525/mp.2016.33.5.571
Stupacher, J., Hove, M. J., Novembre, G., Schütz-Bosbach, S., & Keller, P. E. (2013). Musical groove modulates motor cortex excitability: A TMS investigation. *Brain and Cognition*, *82*(2), 127–136. https://doi.org/10.1016/j.bandc.2013.03.003
Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour*, *27*, 699–715. https://doi.org/10.1016/0003-3472(79)90006-X
Thelen, E. (1981). Kicking, rocking, and waving: Contextual analysis of rhythmical stereotypies in normal human infants. *Animal Behaviour*, *29*(1), 3–11. https://doi.org/10.1016/s0003-3472(81)80146-7
Thiessen, E. D., & Saffran, J. R. (2009). How the Melody Facilitates the Message and Vice Versa in Infant Learning and Memory. *Annals of the New York Academy of Sciences*, *1169*(1), 225–233. https://doi.org/10.1111/j.1749-6632.2009.04547.x
Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied Meter: Hierarchical Eigenmodes in Music-Induced Movement. *Music Perception*, *28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59
Trainor, L. J. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development*, *19*(1), 83–92. https://doi.org/10.1016/S0163-6383(96)90046-6
Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140089. https://doi.org/10.1098/rstb.2014.0089
Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. https://doi.org/10.3758/BF03196290
Trainor, L. J., & Zacharias, C. A. (1998). Infants prefer higher-pitched singing. *Infant Behavior and Development*, *21*(4), 799–805. https://doi.org/10.1016/S0163-6383(98)90047-9
Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669–673. https://doi.org/10.1038/nn1084
Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096. https://doi.org/10.1098/rstb.2014.0096
Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae*, *3*(1_suppl), 155–215. https://doi.org/10.1177/10298649000030S109
Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior and Development*, *33*(1), 96–100. https://doi.org/10.1016/j.infbeh.2009.11.006

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The Impact of the Bass Drum on Human Dance Movement. *Music Perception*, *30*(4), 349–359. https://doi.org/10.1525/mp.2013.30.4.349

Vuust, P., & Witek, M. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, 1111. https://doi.org/10.3389/fpsyg.2014.01111

Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal Neural Entrainment during Early Social Interaction. *Trends in Cognitive Sciences*, *24*(4), 329–342. https://doi.org/10.1016/j.tics.2020.01.006

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468–2471. https://doi.org/10.1073/pnas.0809035106

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. K. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*. https://doi.org/10.1016/j.heares.2005.11.010

Zentner, M. R., & Eerola, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, *107*(13), 5768–5773. https://doi.org/10.1073/pnas.1000121107

Supplementary Materials: Development of Auditory and Spontaneous Movement Responses to Music over the First Postnatal Year

Deviations from Pre-Registration

We ran the majority of analyses according to the preregistration. There is, however, one deviation. Namely, we did not assess motor components (such as mu-oscillations) of infants' neural responses to music in the current paper. We deemed these analyses too extensive for the current paper and will follow up on these analyses in future studies.

EEG Electrode Configuration in infants and adults

Infant electrode configuration included the following electrodes:
VEOG, Fp1 (Ground), Fp2, F7, F3, Fz, F4, F8, FC3, FCz, FC4, FT7, FT8, C3, Cz, C4, T7, T8, CP3, CP4, P7, P3, Pz, P4, P8, TP9 (online Ref), TP10 (offline Ref), PO9, PO10, O1, Oz, O2

The adult electrode configuration included the same electrodes as the infant configuration:
VEOG, Fp1, Fp2, F7, F3, Fz, F4, F8, F4, F8, FC3, FCz, FC4, FT7, FT8, C3, Cz, C4, T7, T8, CP3, CP4, P7, P3, Pz, P4, P8, TP9 (offline Ref), TP10 (offline Ref), O1, Oz, O2. CMS, DRL were used for real-time referencing and grounding.

The following electrodes from the infant configuration were not recorded in the adult configuration: PO9, PO10

The following electrodes were recorded in the adult configuration but excluded in the analysis to make infant and adult electrode configurations comparable: Fpz, AF7, AF3, AFz, AF4, AF8, F5, F1, F2, F6, FC5, FC1, FC2, FC6, C5, C1, C2, C6, TP7, CP5, CP1, CPz, CP2, CP6, P9, P5, P1, P2, P6, P10, PO7, PO3, POz, PO4, PO8, Iz.

EEG signal quality checks

We conducted two analyses to compare the EEG data quality across age groups. First, we compared the number of trials that were included in the final analysis per age group. The trial number did not differ significantly across age groups ($p > .361$). Second, we calculated the SNR by dividing the EEG power at the frequency of interest (i.e., 2.25 Hz, matching the musical beat) by the background noise in surrounding bins (3rd to 5th bin, see ASSR methodology for further details; c.f., Christodoulou et al., 2018; Cirelli et al., 2014). This division yields a signal-to-noise ratio that can be averaged across conditions and compared across age groups to assess variations in signal quality. Here, we find that all three age groups show considerable SNR above 1. Further, even though there were descriptive differences, namely 3- and 6-month-old infants exhibited numerically higher SNR (3m: M=2.569, SD=1.104; 6m: M=2.743, SD=1.001) compared to 12-month-old infants (M=1.907, SD=0.749), the three age groups did not differ significantly (three t-tests, controlled for multiple comparison using the

false discovery rate, $p > .134$). Together, these two analyses indicate that signal quality was comparable across age groups.

Table S1. Continued acoustic characteristics of the musical stimuli

Song name (Condition)	*ENVELOPE M±SD (A.U.)*	*ENVELOPE RANGE (A.U.)*
HOPP (MUSIC)	0.162±0.047	-0.014-0.290
HOPP (SHUFFLED MUSIC)	0.161±0.054	-0.012-0.318
HOPP (HIGH PITCH)	0.160±0.046	-0.011-0.290
HOPP (LOW PITCH)	0.180±0.046	-0.007-0.317
LOLA (MUSIC)	0.147±0.067	-0.013-0.265
LOLA (SHUFFLED MUSIC)	0.154±0.059	-0.005-0.305
LOLA (HIGH PITCH)	0.144±0.067	-0.013-0.265
LOLA (LOW PITCH)	0.160±0.069	-0.011-0.262

Note. Hopp indicates the Hungarian playsong ("Hopp Juliska"), Lola indicates the Spanish playsong ("La vaca lola").

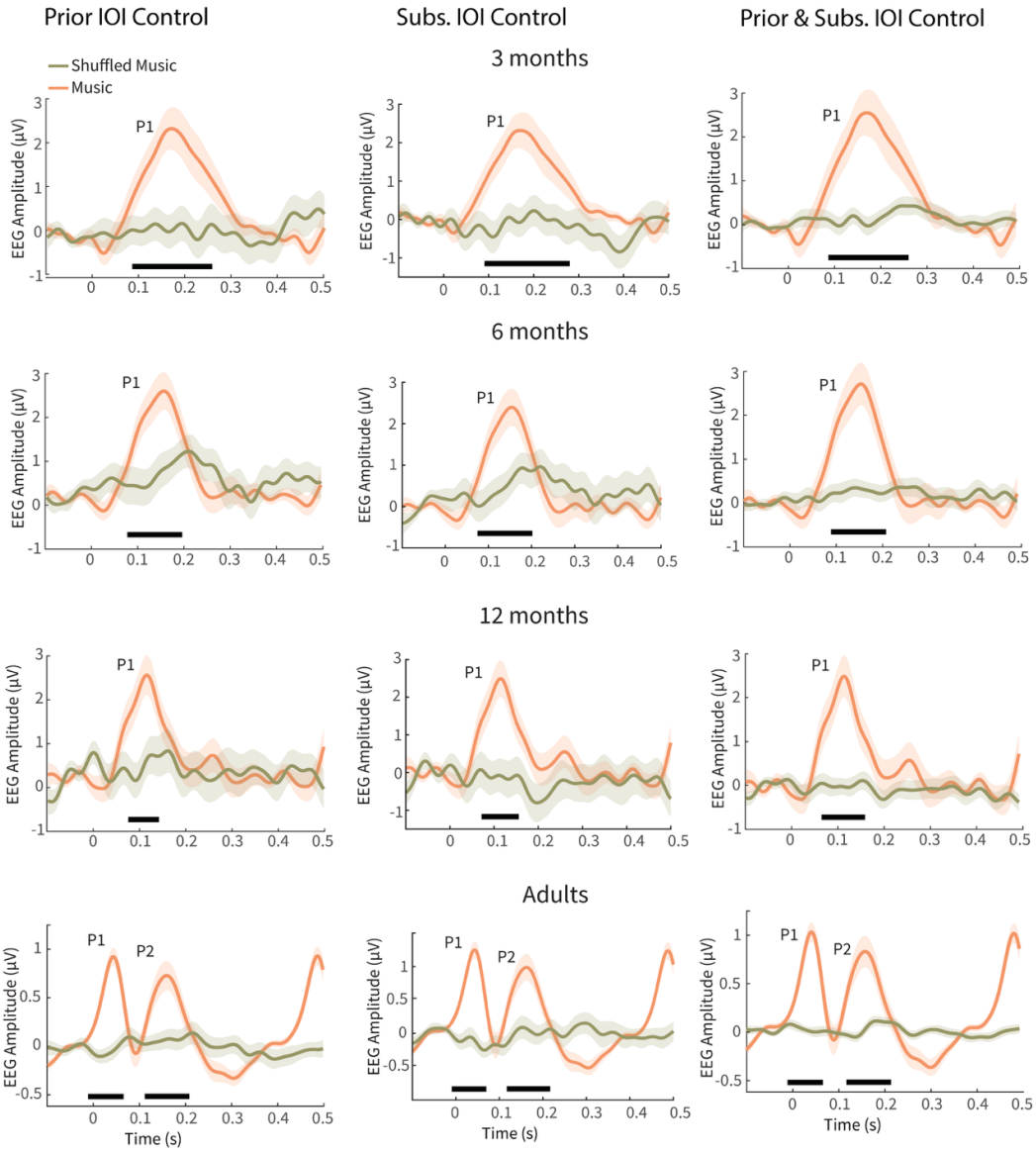


Figure S1. *Event-related potentials (ERPs) elicited by the notes comprised within the music (orange) vs shuffled music (khaki) across four groups of participants from top to bottom: 3-, 6-, 12-month-old infants (N=79) and adults (N=26). These control analyses included only epochs from the shuffled music condition in which the prior (left column), subsequent IOI (middle column) of bassline notes or prior and subsequent IOI of melody and bassline notes (right column) exceeded the median IOI duration (*212 ms*). The number of included epochs in the music condition was matched to the shuffled music condition. Grand-average ERPs were computed by averaging across electrodes within the significant cluster identified for each age group in the music condition based on the primary ERP analysis. Shaded areas indicate the standard error of the mean.*

Variance in Quantity of Movement

To assess how infants' movement patterns varied with musical structure, we analyzed the Variance in Quantity of Movement (QoM) using a mixed-effects model. Model outputs show a significant interaction between age groups and conditions ($\chi^2(2)=8.36$, $p=.015$), age groups and PMs (($\chi^2(18)=242.50$, $p<.001$), and main effects of age ($\chi^2(2)=8.66$, $p=.013$) and PMs ($\chi^2(9)=288.59$, $p<.001$). Post-hoc contrasts indicated that 12-month-old infants showed greater variability in QoM when listening to music compared to shuffled music ($t(69.8)=3.09$, $p=.003$). In contrast, 3- and 6-month-olds' QoM did not vary differently across music and shuffled music ($p>.591$). We also found no significant differences in QoM variance between musical stimuli that differed in pitch ($p>.339$).

Granger Causality Statistics stratified by each PM and age group

To assess whether the Granger Causality results were driven by specific PMs, we averaged the Granger F-values within the 160-200 ms lag window and entered these average values into a linear mixed effect model. This analysis yielded an interaction effect between age, PM, and condition ($\chi^2(18)=31.13$, $p=.028$, *Fig. S1*), while all other effects were not significant ($p>.205$). Post-hoc contrasts showed that in 3-month-old infants, music vs shuffled music drove mostly particular movements such as up-down rocking (PM5), arm pedaling (PM6), whole body wiggling (PM8), feet shuffling (PM9), and feet pedaling (PM10) ($t(583)=2.02$-3.42, $p=.001$-$.044$). For 6-month-olds, music drove mostly front-back rocking (PM1), side swaying (PM2), proto-clapping (PM3), feet kicking (PM4), arm pedaling (PM6), and whole body wiggling (PM8) ($t(573)=2.06$-4.13, $p=.001$-$.047$). In 12-month-olds, music-driven movement mostly entailed front-back rocking (PM1), proto-clapping (PM3), arm pedaling, up-down rocking, feet kicking (PM7), whole body wiggling (PM8), and feet shuffling (PM9) ($t(573)=1.97$-4.87, $p=.001$-$.049$). Conversely, the same analysis, conducted on high- and low-pitch music conditions, yielded a main effect of condition ($\chi^2(1)=47.87$, $p<.001$) indicating that high-pitch music drove movement better than low-pitch music, and that was so across all PMs and age groups ($t(67.6)=6.87$, $p<.001$). Together, these results indicate that the principal movements specifically predicted by music, as opposed to shuffled music, grow in number and change in type with age: in younger infants (3 months), music best predicted lower-body movements, while in 6-month-olds, music best predicted upper-body movements. By 12 months, music best predicted whole-body movements.

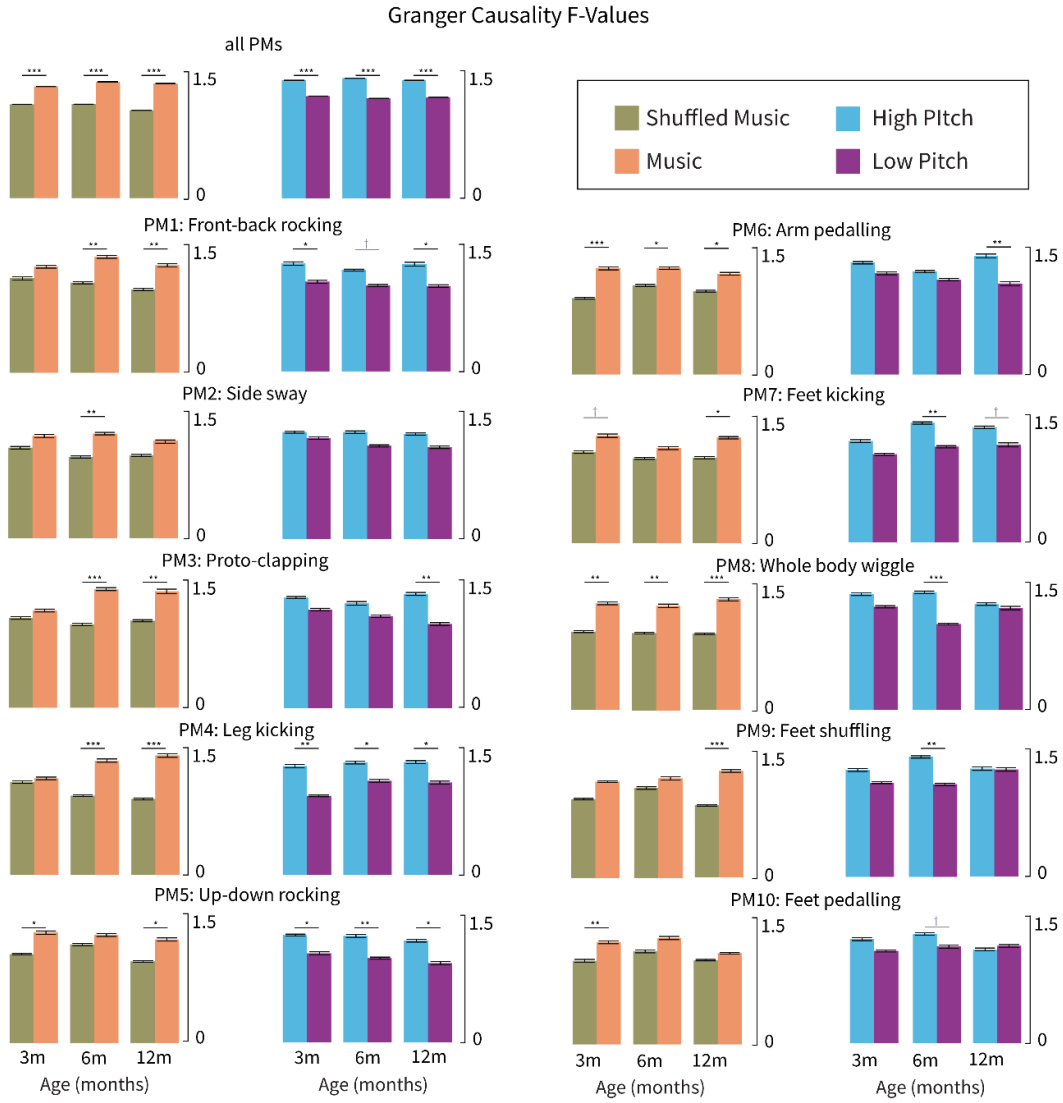


Figure S2. *The figure displays the Granger Causality F-Values (y Axis) averaged across conditions and age groups for all PMs combined (top row left) as well as for individual PMs (1 to 10, left to right). The left columns include music (orange) vs shuffled music (khaki) contrasts for each age group (x-Axis: 3m, 6m, and 12m) while the right columns include high-pitch (blue) vs low-pitch (purple) contrasts. Across all three age groups movement velocity was more strongly predicted by music than shuffled music, and by high-pitch than low-pitch music, p<.001. These effects were particularly pronounced for movement patterns such as front-back rocking (PM1), proto-clapping (PM3), leg kicking (PM4) and whole-body wiggle (PM8), especially so at 6 and 12 months of age, as reflected in the variation of Granger F-values across PMs and ages.*

Movement responses

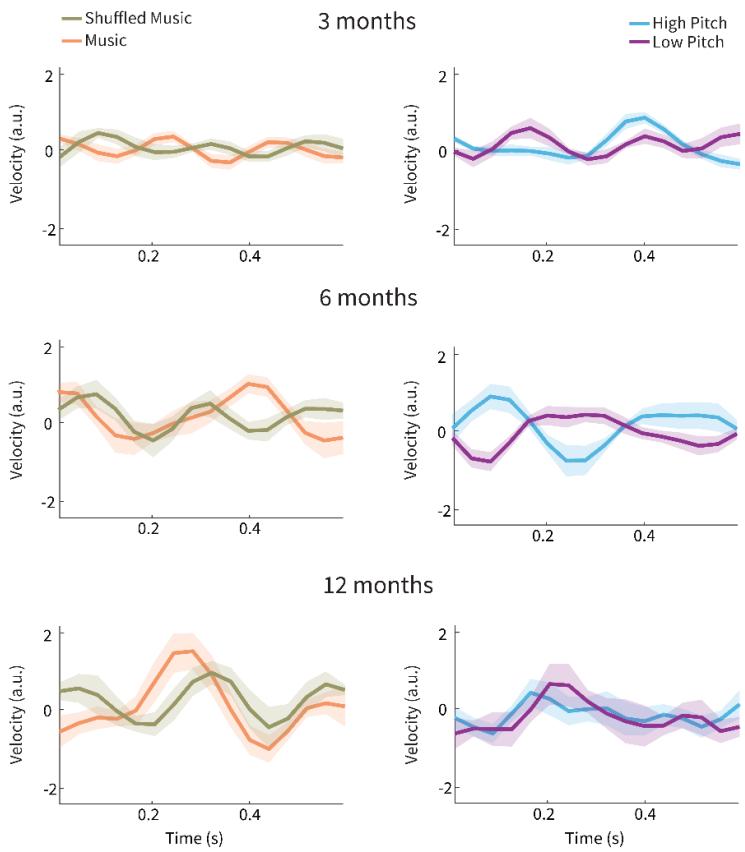


Figure S3*. The figure displays the z-scored velocity (a.u., y-axis) of infants' movement responses over time (in seconds, x-axis) following peaks in the amplitude envelope of the stimuli (at time = 0) across different conditions. These peaks represent moments of increased acoustic intensity in the music, allowing visualization of how infants at three developmental stages (3 months, top row; 6 months, middle row; 12 months, bottom row) synchronized their movements with dynamic changes in the sound. Each subplot compares different conditions, with left plots comparing music (orange) to shuffled music (khaki) and right plots comparing high-pitch (blue) with low-pitch music (purple) conditions. None of the age groups show significant differences in velocities following peaks in the amplitude envelope of the auditory stimuli across conditions. Specifically, the results suggest that changes in sound intensity seem to evoke movement responses that vary in latency and might be inconsistent across trials (with the only exception of the 12-month-old infants, who appear to show a peak around 200 ms across several conditions). In this context, such varying responses are better captured by Granger Causality analysis. Perhaps because such predictive approaches are more powerful in picking up signals that might be either sporadic or variable in amplitude and/or sign across trials (Shojaie & Fox, 2022).*

Movement – Periodicity analyses

Despite infants showing no consistent phase-locked movement responses, we tested further whether these responses were periodic and, if so, whether such periodicity differed across conditions. To do so, we used an auto-correlation approach described in a previous study reporting periodic movements in infants exposed to music (Zentner & Eerola, 2010). The analysis builds upon the distribution of the highest coefficients yielded by autocorrelation analyses (*Fig. S3*). We first examined whether the density distribution of movements was uniform (indicating non-periodic movements) or non-uniform (indicating periodic movements) across PMs for each age group. To do so, we performed a Kolmogorov-Smirnov test, which revealed a highly significant result (KS=0.838-0.869, p<.001), indicating that the movements were periodic across all PMs and age groups. Next, we extracted the density value associated with the beat-related lag (444 ms) and compared these values across conditions averaged over all PMs per age group. Results indicated no significant differences in periodicity between music and shuffled music (ps>.120). All infants, therefore, demonstrated periodic motor responses to auditory stimuli, regardless of whether they listened to music or shuffled music. Additional analyses comparing the pitch conditions revealed no significant differences either (ps>.095).

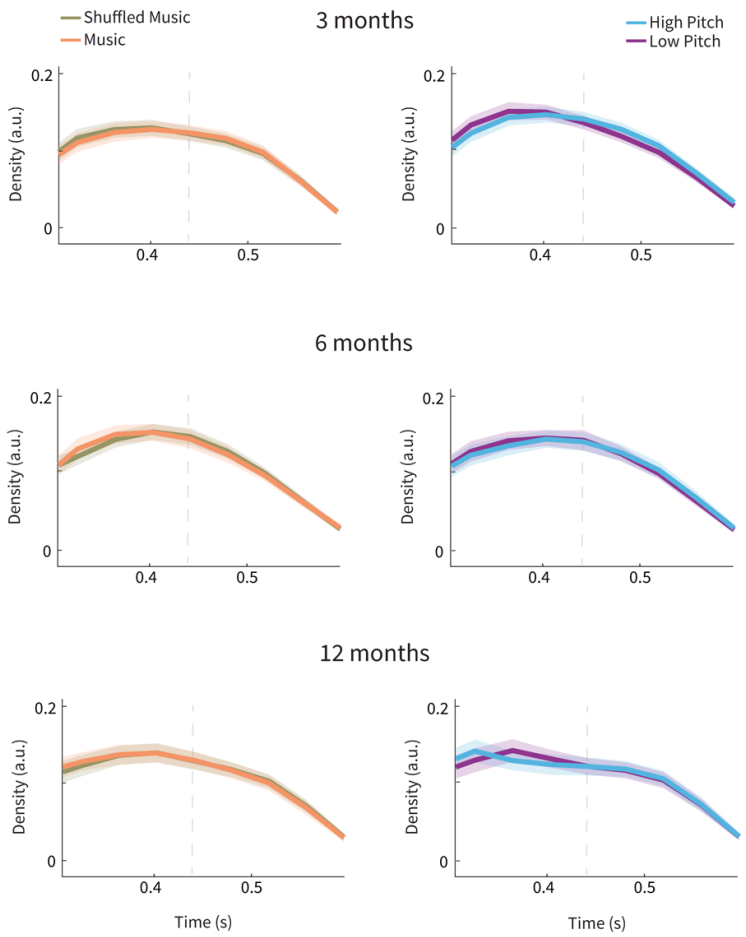


Figure S4. *The graphs depict the density (y Axis) function of infants' periodic movement to music (orange) vs shuffles music (khaki) as well as high-pitch (blue) vs low-pitch music (purple) averaged across all PMs at three developmental stages: 3 months (top row), 6 months (middle row) and 12 months (bottom row). The lag in seconds is depicted on the x Axis and the dotted vertical line indicates the musical beat (444 ms lag). Infants' movements tended to be rhythmic, however, not differently across conditions (i.e., the contrasts of density values across conditions were not significant).*

Links between neural and movement responses

Additional exploratory analyses examined correlations between ERP amplitudes and movement measures using participant-averaged data. Mean ERP amplitudes of the P1 component time window, identified from cluster-based analyses, served as neural measures. Movement measures included (1) total movement quantity (mean velocity across the trial) and (2) Granger causality F-values indicating music-to-movement coupling strength. Analyses contrasted music versus shuffled music conditions and high-pitch versus low-pitch conditions. Linear mixed-effects models were fitted with ERP amplitude as the response variable, either movement quantity or Granger causality F-values as fixed effects, and infants as random intercepts. Results revealed no significant correlations between ERP amplitude and movement quantity, irrespective of conditions (p>.124), and neither when comparing music versus shuffled music (p>0.111) or high versus low pitch (p>0.071). Similarly, no significant correlations were found between ERP amplitude and Granger causality F-values, irrespective of conditions (p>.164) and in either contrast (music vs. shuffled music: p>0.494; high vs. low pitch: p>0.175). These findings might suggest that neural sensitivity to musical structure and motor responsiveness develop independently within the first year of life, consistent with developmental theories positing perceptual sensitivity precedes motor coordination.

2. Manuscript

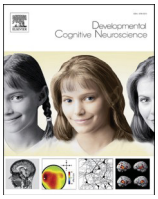
Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313

Developmental Cognitive Neuroscience 64 (2023) 101313

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn



Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing

Trinh Nguyen [a,b,*], Susanne Reisner [a], Anja Lueger [a], Samuel V. Wass [c], Stefanie Hoehl [a], Gabriela Markova [a,d,**]

[a] Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria
[b] Neuroscience of Perception and Action Lab, Italian Institute of Technology, Viale Regina Elena 291, 00161 Rome, Italy
[c] Department of Psychology, University of East London, University Way, London E16 2RD, United Kingdom
[d] Institute for Early Life Care, Paracelsus Medical University, Strubergasse 13, 5020 Salzburg, Austria

ARTICLE INFO

Keywords:
Infant-directed singing
Social interaction
EEG
Neural tracking
Rhythmic movement

ABSTRACT

Infant-directed singing has unique acoustic characteristics that may allow even very young infants to respond to the rhythms carried through the caregiver's voice. The goal of this study was to examine neural and movement responses to live and dynamic maternal singing in 7-month-old infants and their relation to linguistic development. In total, 60 mother-infant dyads were observed during two singing conditions (playsong and lullaby). In Study 1 ($n = 30$), we measured infant EEG and used an encoding approach utilizing ridge regressions to measure neural tracking. In Study 2 ($n = 40$), we coded infant rhythmic movements. In both studies, we assessed children's vocabulary when they were 20 months old. In Study 1, we found above-threshold neural tracking of maternal singing, with superior tracking of lullabies than playsongs. We also found that the acoustic features of infant-directed singing modulated tracking. In Study 2, infants showed more rhythmic movement to playsongs than lullabies. Importantly, neural coordination (Study 1) and rhythmic movement (Study 2) to playsongs were positively related to infants' expressive vocabulary at 20 months. These results highlight the importance of infants' brain and movement coordination to their caregiver's musical presentations, potentially as a function of musical variability.

Musicality, defined as the ability to perceive, process and produce music, is a human universal with deep ontogenetic roots (e.g., Mehr et al., 2021; Savage et al., 2021). Infants experience a variety of different types of music every day (Mendoza and Fausey, 2021; Warlaumont et al., 2022). Most prominently, caregivers sing to them to capture their attention, regulate their arousal, and share emotions with or entertain them (Trehub, 2019). On the other side, children have sophisticated perceptual capabilities when listening to music from early in life: they are sensitive to tempo- and pitch-related organization of musical sequences, such as changes in melodic patterns, harmony, grouping, beat, and meter (e.g., Ferland and Mendelson, 1989; Háden et al., 2015; Krumhansl and Jusczyk, 1990; Trainor and Heinmiller, 1998; Winkler et al., 2009). For example, frequency tagging studies show that infants' brain activity increases relative to rhythmic beats in music, and that these increases are influenced by musical properties such as meter and by musical experience (Cirelli et al., 2016; Lenc et al., 2022). No research to date has, however, examined whether volume fluctuations in sung music directly map onto fluctuations in the brain activity in a listening child. Several recent studies have demonstrated similar correspondence between volume fluctuations in child-directed speech and children's brain activity (Attaheri et al., 2022; Menn et al., 2022), and so we were interested to investigate whether infants also show neural tracking of music.

In addition, we wanted to understand better the mechanisms that might subserve neural tracking of music. Here, we refer to neural tracking such that an exogenous signal and an endogenous activity are temporally aligned, also referred to as neural entrainment in the broad sense (Obleser and Kayser, 2019). In particular, we were interested to study the association between infants' rhythmic movement to music (Sievers et al., 2013) and neural tracking. Fetuses at 35 weeks of gestational age already show more movement to music in comparison to speech (Kisilevsky et al., 2004), and by 2 months of age, they coordinate

* Correspondence to: Viale Regina Elena 291, 00161 Rome, Italy.
** Correspondence to: Liebiggasse 5, 1010 Vienna, Austria.
E-mail addresses: trinh.nguyen@iit.it (T. Nguyen), gabriela.markova@univie.ac.at (G. Markova).

https://doi.org/10.1016/j.dcn.2023.101313
Received 27 February 2023; Received in revised form 29 September 2023; Accepted 12 October 2023
Available online 24 October 2023

their gaze with the beat of infant-directed (ID) singing (Lense et al., 2022). During later infancy, they move their bodies more to music than to other auditory stimuli, and modulate the tempo of their movement contingent on the tempo of the music (Ilari, 2015; Zentner and Eerola, 2010). Because rhythmic movements are known to guide auditory perception (Phillips-Silver and Trainor, 2005, 2007), it has been suggested that this coordination, in particular to the rhythmic structure of the musical input, may modify brain activity and behavior in such a way that they eventually lock onto a common periodicity (Vuilleumier and Trost, 2015). However, no previous research has investigated this hypothesis. Previous discussions of neural tracking of language have concentrated on a range of different factors that drive language entrainment, such as understanding how temporal fluctuations in the critical band envelope of speech directly entrain brain activity (Doelling et al., 2014), and how cognitive factors such as the comprehensibility of speech mediate neural entrainment (Poeppel and Assaneo, 2020). In the present study we examined the association between neural tracking and rhythmic movement to provide a new perspective on the mechanisms that drive neural entrainment to music.

We also wanted to investigate the three-way relationship between neural tracking, rhythmic movements to music, and language development. Recent evidence suggests that neural tracking of nursery rhymes is predictive of infants' later language outcomes (Attaheri et al., 2022; Menn et al., 2022). Brandt and colleagues (2012) propose that the musical aspects of language (i.e., prosody, rhythm, timbral contrast) scaffold the later development of semantic and syntactic aspects of language. In fact, available evidence suggests that an enriched musical environment (i.e., music classes, music at home) during infancy and preschool can promote language development (Papadimitriou et al., 2021; Politimou et al., 2019; Putkinen et al., 2013; Williams et al., 2015; Zhao and Kuhl, 2016). Brain responses of newborns to sung as compared to spoken streams of syllables predict expressive vocabulary at 18 months (François et al., 2017), and ID singing, in particular, has been found to facilitate aspects of phonetic perception and word learning in infants in the second half of the first year of life (Lebedeva and Kuhl, 2010; Thiessen and Saffran, 2009). Rhythmic motor abilities are thought to provide relevant practice of rhythmic, closely timed actions, as they are required for speech (Iverson, 2010; Iverson et al., 2007). Yet, it remains unknown whether infants' physical and neural tracking of music is also related to their language development.

Much of the research examining infant musicality, as reviewed above, used pre-recorded or digitalized stimuli that are easier to control and manipulate. However, pre-recorded music has a phenomenologically different quality compared with live musical performances (Trehub, 2017). Caregivers sing to infants on a daily basis (Steinberg et al., 2021), and ID singing is part of intimate interactions between caregivers and infants with the goal of engaging or soothing (Trainor et al., 1997). ID singing, compared to non-ID singing, is characterized by high pitch, slow tempo, enhanced temporal regularity, and greater dynamic range (Nakata and Trehub, 2011; Trainor et al., 1997). ID songs are often accompanied by idiosyncratic gestures (Eckerdal and Merker, 2009), energetic movement, and caregivers' positive affect (Cirelli et al., 2020; Trehub et al., 2016), all of which are used by caregivers to attract infants' attention, activate, and engage with them (Cirelli et al., 2020; Trehub and Trainor, 1998). Moreover, caregivers adjust their facial expressions and gaze to the timing of the ID songs (Lense et al., 2022). Still, it remains unclear how infants perceive and act upon *live* and *dynamic* ID singing.

The acoustic features of ID songs can be further specified by the context in which they are used, namely, to play or to soothe. Playsongs are characterized by higher rhythmicity, tempo, and pitch level, as well as more dynamic variability and thus increased complexity in comparison to lullabies (Cirelli et al., 2020; Rock et al., 1999; Trehub and Trainor, 1998). On the other hand, lullabies are characterized by lower tempo, pitch, and less variability, intending to soothe and calm an infant. Accordingly, the different acoustic features might affect how infants coordinate their neural and movement responses. Recent research in adults shows that neural tracking was enhanced in slower music (Weineck et al., 2022) and was guided by melodic expectations (Di Liberto et al., 2020). We thus expected enhanced neural tracking in slower and "simpler" ID songs (i.e., lullabies), while we hypothesized that the complexity of playsongs might attenuate infants' neural tracking. Importantly, the specific acoustic features of lullabies and playsongs might have different effects on infants' rhythmic movement than on neural tracking. On the behavioral level, increased rhythmic complexity was related to more movement and reported groove (i.e., the inclination to move) in children and adults (Cameron et al., 2022; Kragness et al., 2022). The enhanced complexity of playsongs might thus be more engaging to children.

Overall, we still know very little about infants' neural tracking of and rhythmic movement to live and dynamic maternal singing. Closing this gap will help us understand how we might be able to facilitate infants' physical, social, and cognitive development through musical communication. In the present study, we used the naturally occurring variance in acoustic features of playsongs and lullabies during live and dynamic ID singing to gain a deeper understanding of the tracking effects on infants' bodies and brains. In Study 1, we measured 7-month-old infants' neural tracking (using EEG) to maternal singing. To allow for more naturalistic and potentially more expressive movements, we observed rhythmic movements in an additional sample of 7-month-old infants while mothers sang to them (Study 2). We hypothesized that infants' neural responses are more coordinated with lullabies than playsongs. In contrast, their rhythmic movements were expected to be more frequent with playsongs than lullabies. At 20 months, we assessed infants' vocabulary to examine whether early musical engagement relates to infants' language development. We hypothesized that higher neural tracking of ID songs, as well as more rhythmic movements, predict children's vocabulary.

1. Methods

1.1. Participants

In Study 1, 30 7-month-old infants (13 females; age: 228.80 ± 8.20 days [$M \pm SD$]) and their mothers (age: 33.96 ± 4.88 years) were included in the final sample. An additional 29 infants were tested but could not be included due to fussiness ($n = 4$), technical problems during the recording ($n = 9$), and bad signal quality due to the naturalistic setup ($n = 16$). In Study 2, 40 7-month-old infants (19 females; age: 227.17 ± 5.8 days) and their mothers (age: 35.10 ± 4.00 years) were included in the final sample. We excluded 7 additional mother-infant dyads from the final analysis due to infant fussiness ($n = 5$), or because the mother did not follow instructions ($n = 2$).

Participants were recruited from a database of families who expressed interest in participating in developmental research and have consented to being contacted by our staff. These families were recruited in neonatal units at local hospitals, in mother-child activity classes, and through social media. All infants were born full-term (38–42 weeks gestational age), had a birth weight of > 2500 g, and had no known developmental delays or neurological or hearing impairments. Infants grew up in predominantly German-speaking households. Mothers were highly educated, with 93.1% (Study 1) and 86.7% (Study 2) of mothers holding a university degree. Fifty-five percent of the mothers in both samples reported playing an instrument, and 20% were singing in a choir or a band. The study was conducted according to the declaration of Helsinki, approved by the university's ethics committee, and parents provided written informed consent.

1.2. Procedure

After arrival in the laboratory, parents and infants were familiarized with the environment. Parents were informed about the study and

signed a consent form. In Study 1, the EEG setup was prepared while the infant sat on the parent's lap. Infants were seated in an infant car seat, and their mothers faced them (Fig. 1A). Study 2 followed the same procedure as Study 1, except infants sat upright in an infant highchair that allowed more movement. In Study 2, we additionally measured infants' electro-cardiac rhythms using a two-lead derivation. These results will be reported elsewhere.

Each dyad was observed during two experimental singing blocks (Fig. 1B), the order of which was randomized between participants. Each singing block was preceded and followed by a baseline (i.e., three in total), during which infants and mothers watched videos of slowly moving shapes for 60 s together. Mothers were instructed to refrain from talking during the baseline, but to reciprocate their infants' communicative attempts (e.g., smiling, pointing towards the screen). During each singing block, mothers continued holding the tablet showing a calm aquarium video to help with infant fussiness. Mothers were instructed to sing one of two types of songs: a playsong or a lullaby. Songs were selected according to their level of familiarity during pilot testing. One example of each type of song was used. These were 'Schlaf Kindlein, schlaf' (lullaby) and 'Es tanzt ein Bibabutzemann' (playsong) (see SM Fig. S1). All mothers sang the same songs and were informed about and provided recordings of the songs at the time of recruitment. Most (87%) of the mothers were familiar with the playsong, and 97% were familiar with the lullaby. All mothers sang four verses of each song and repeated each song, resulting in eight verses for each song. Mothers were prompted with a metronome before the commencement of each song (i.e., playsong = 170 bpm; lullaby = 100 bpm). Three cameras (synchronized in VideoSyncPro by Mangold International) were used to record the experiment.

2. Measures

2.1. Maternal singing

Audio of maternal singing was recorded using a portable microphone (Mangold International) at a 44,100 Hz sampling rate. Audio data were pre-processed using the software Audacity. Excerpts containing infant vocalizations and other noises from the environment were manually removed. To examine whether mothers differentiated between the two singing conditions, we analyzed maternal audio data in terms of mean and standard deviation in tempo (beats per minute from the autocorrelation function of the onset detection curve), pitch (Hz), root mean square (RMS; perceived loudness), and pulse clarity using MIRtoolbox (Lartillot et al., 2008). We also manually checked the derived tempo values by tapping along to the original recordings (random 10% of the sample) and found that our tapping values deviated from the tempo derived from MIRtoolbox by max. 2 BPM.

In order to extract the sound envelopes for further EEG analyses (Jessen et al., 2021), the audio data were then submitted to the NSL toolbox (http://nsl.isr.umd.edu/downloads.html). The resulting matrix contained band-specific envelopes of 128 frequency bands of uniform width on a logarithmic scale with center frequencies logarithmically spaced between 0.1 and 4 kHz. We obtained the broadband temporal envelope of the audio soundtrack by summing up band-specific envelopes across all frequencies to obtain one temporal envelope. The spectrogram of the envelope is included in the supplements (Fig. S1). Lastly, the amplitude envelope data were down sampled to match the 500 Hz sampling rate of the EEG data.

2.2. Infant EEG (Study 1)

We recorded infant EEG using a BrainAmp DC 64-channel system and a 32- channel ActiCap system (Brain Products, Germany) containing active Ag/AgCl electrodes, mounted according to scalp locations of the 10–20 system. Horizontal and vertical electrooculograms were recorded bipolarly to control for eye movements. EEG data were sampled at 500 Hz. Impedances were controlled at the beginning of the experiment and accepted when below 20 kΩ.

EEG data were processed using the *eeglab* (Delorme and Makeig, 2004) and *fieldtrip* toolboxes (Oostenveld et al., 2010) as well as custom MATLAB code. EEG data were band-pass filtered with a noncausal finite impulse response filter (0.3–100 Hz, −6 dB cutoff: 0.15–100.15 Hz). Next, we reduced line noise (50/100 Hz) by using ZapLine from NoiseTools (de Cheveigné, 2020). Noisy channels were identified by assessing the normed joint probability of the average log power from 1 to 100 Hz and rejected if exceeding a threshold of 3 SD from the mean (number of removed channels: $M = 4.61$; *range* = 0 – 10). This step was repeated twice. Next, we used wavelet thresholding to identify and remove artifacts (threshold = hard, wavelet = coif4, level-dependent thresholding; Lopez et al., 2022). The previously rejected channels

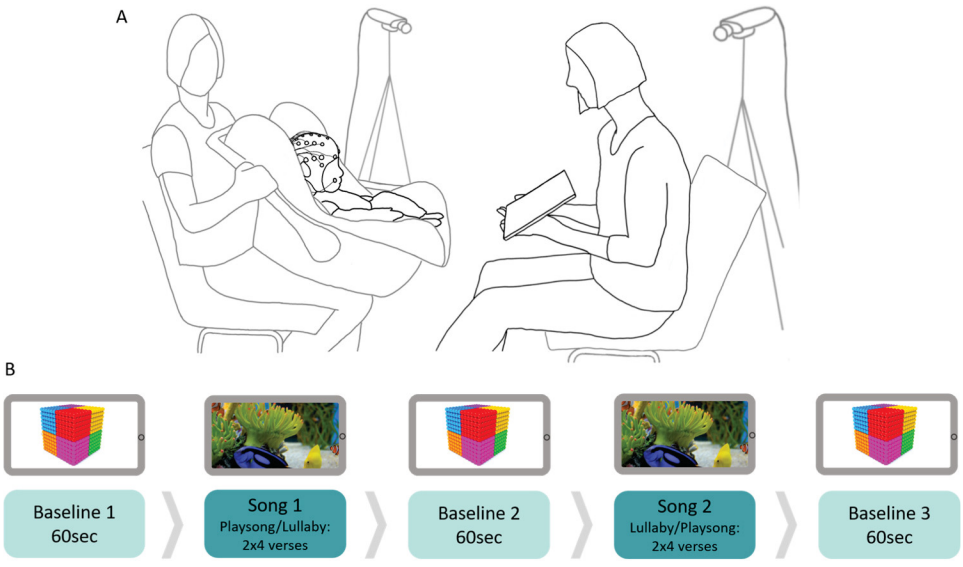


Fig. 1. (A) Illustrated experimental setup. An experimenter held the car seat with the infant facing their mother in Study 1. Video cameras were used to record the mother-infant dyad from two angles. (B) Procedure. Mothers sang two repetitions of four verses for each lullaby and playsong. During both conditions, a tablet held by the mother played a video of fish swimming in an aquarium. A 60-second baseline, during which the mother did not interact with the infant, preceded, and followed each song, during which the tablet showed moving geometric shapes.

were then interpolated using spherical splines. Afterward, all channels were re-referenced to the average of all scalp electrodes (F7, F8, F9, F3, Fz, F4, FT7, FC3, FCz, FC4, FT8, T8 C4, Cz, C3, T7, CP3, CP4, TP10, P8, P4, Pz, P3, P7, PO9, PO10, O1, Oz, O2) and segmented into non-overlapping 1-s-epochs. In preparation for further analysis, epochs were automatically examined to see if they were still contaminated by artifacts. The standard deviation was computed in a sliding window of 200 ms. If the standard deviation exceeded 100 μV at any electrode, the entire epoch was discarded. Participants were excluded from further analysis if less than 90 artifact-free epochs could be retained. Infants' processed EEG data comprised on average 144 epochs from the playsong condition and 126 epochs from the lullaby condition. The spectrogram of the data is provided in Fig. S2. EEG data were then FIR-type bandpass filtered between 1 and 10 Hz (Jessen et al., 2019).

Neural tracking analyses. To quantify the degree to which the 7-month-old infants showed neural tracking of maternal singing, we used encoding models (i.e., temporal response functions [TRF]) to predict the infants' EEG from the amplitude envelope of mothers' singing. The TRF regresses the neural response onto the auditory signal, dividing out the autocovariance structure of the stimulus from the model. Further information on the approach is provided in the supplements.

Following Jessen et al. (2019), we estimated the predictive accuracy of the model by using individual models (computation of individual response function for each infant). We assumed individual models to be suitable for our purposes as ID songs were highly different between dyads, therefore leading to vastly different regressor weights between infants. To that end, 80% of the available data for a given participant was used to train the model. The resulting response function was then correlated with the response observed in the remaining 20% of the data. To accommodate a time lag between changes in the EEG signal and changes in the stimulus signal, predictions were computed over a range of time lags (i.e., 2 ms) between −100 and 500 ms later than the stimulus signal (see Jessen et al., 2021, for further details).

We obtained the optimal regularization parameter λ by training the respective model on the concatenated EEG and audio data of each infant using a variety of λ values between 10^{-7} and 10^{7}. We increased the exponent in steps of 1 and used the resulting models to predict the EEG signal for each participant. By doing so, we obtained a total of fourteen different models (and predictions) based on the different λ parameters. For each of these 14 different models, we computed the mean response function across n-1 participants and used this response function to predict the EEG of the n^{th} fold (i.e., n-fold leave-one-out cross-validation). One fold comprised 1/5 of the data. Finally, we computed the predictive accuracy (i.e., Pearson's correlation coefficient *r* between the predicted EEG and the actual EEG) for each infant and each electrode, resulting in 31 accuracy values per infant and stimulus parameters for each λ value. For each infant, stimulus parameter, and electrode set, we chose the λ value for which the model yielded the highest correlation between the predicted and the actual EEG time-series (Jessen et al., 2021).

For statistical evaluation, we computed two different predictive accuracies per infant and song conditions. First, we computed the correlation between the predicted response generated on a model trained on 80% of the data on the other 20% of the data ("individual model"). Second, a permuted or null predictive accuracy ("circular shift control") was obtained. Before calculating accuracy this way, we reversed and applied a random circular shift to the true EEG time series for participant *n* (to ensure exceeding the potential autoregressive structure of the EEG) and computed the correlation between the shifted EEG and the predicted response from the individual model (for $n = 400$ permutations). The permutations were then averaged, resulting in one so-called "randomized" predictive accuracy value per infant.

To evaluate the resulting response functions, we computed a cluster-based permutation test with 1000 randomizations, testing the obtained response functions against zero. A cluster was defined along with the dimensions of time (lags between the sound envelope and the EEG) and electrode position, with the constraint that a cluster had to extend over at least two adjacent electrodes. A type-1 error probability of less than 0.05 was ensured at the cluster level (Meyer et al., 2021). The cluster-based permutation analysis revealed no significant clusters in the TRF, potentially due to the large individual differences between the infants' evoked responses as well the acoustic features of the ID songs. The detailed results of this evaluation are reported in the supplements (see Fig. S3).

2.3. Infant rhythmic movement

We coded the duration of infant rhythmic movement for both studies during the two singing conditions. Infant rhythmic movement was defined as at least three repetitions of the same movement in a body part or the entire body at regular short intervals (1 s or less; Thelen, 1979). We calculated proportionate movement values to compensate for differences in recording length. To ensure inter-rater reliability, three independent observers coded 30% of all videos yielding high inter-rater reliability with $\kappa = .82$.

2.4. Infant Gaze

To control for infant visual attention in both studies, we coded the following categories of infant gaze from video recordings using INTERACT by Mangold: (1) *gaze at mother's face*; (2) *gaze at mother's body; (3) gaze at tablet*; (4) *gaze away* as infant gaze directed away from their mother and the tablet at something else in their surrounding (Markova and Legerstee, 2006). Coding was conducted without sound, frame-by-frame, and each dwell time had to last at least 1 s. Inter-rater reliability was assessed on 30% of data coded by three pairs of independent observers, yielding moderate to excellent inter-rater reliability within a range of $\kappa = .70–.99$ for all categories. Analyses regarding infant gaze are reported in the supplements (Table S1).

2.5. Questionnaires

To assess language development, we used the Austrian Communicative Development Inventory - Level 2SF (ACDI-2 SF; Marschik et al., 2007) that mothers filled out when children were 20 months of age (age: 20.23 ± 1.35 months [$M \pm SD$]). Further self-reports on mothers' depression symptoms, anxiety levels after singing, and familiarity of infants with the songs are included in the supplements.

3. Statistical analysis

All statistical analyses were conducted in RStudio (RStudio Team, 2022).

First, we ran a linear mixed-effects model comparing the prediction accuracy values between song type (playsong vs lullabies), data type (true (i.e., the predicted infant EEG from original sound envelopes correlated with the original infant EEG) vs shifted (i.e., the predicted infant EEG from the circular shifted sound envelope correlated with the original infant EEG)) in each channel. The prediction accuracy values were Fisher's z transformed. The robustness of effects was evaluated using 95% confidence intervals and whether those excluded 0.

*Prediction accuracy ~ song type * data type * channel +(1+ song type + data type | ID).*

Next, we included the acoustic features of the playsong and lullaby (i.e., tempo, pitch, pulse clarity, and RMS) as fixed and interaction effects to examine whether differences in these features were related to differences in infants' neural tracking of ID singing.

*Prediction accuracy ~ song type *(tempo+pitch+pulse clarity +RMS) + channel +(1 | ID).*

*Prediction accuracy ~ song type *(tempo.sd+pitch.sd+pulse clarity.sd +RMS.sd) + channel +(1 | ID)*

Finally, we included infants' expressive vocabularies at 20 months as

fixed and interaction effects to examine whether variation in their vocabulary was related to differences in infants' neural tracking of ID singing.

*Prediction accuracy ~ song type * vocabulary + channel +(1 | ID).*

Next, we ran a generalized linear mixed-effects model assuming a beta distribution comparing the proportionate duration of rhythmic movements in the lullaby and playsong conditions in both studies.

*Rhythmic movement ~ song type * study +(1 | ID).*

As we found significant differences in amounts of rhythmic movement between studies, we continued to test further relations with neural tracking and acoustic features separately for each study.

First, we tested whether neural coordination to ID songs in infants was related to the amount of rhythmic movement in Study 1.

*Prediction accuracy ~ song type * rhythmic movement + channel +(1 | ID).*

Next, we included the acoustic features of the playsong and lullaby (i.e., tempo, pitch, pulse clarity, and RMS) as fixed and interaction effects to examine whether differences in these features were related to differences in infants' rhythmic movement during ID singing in Study 2.

*Rhythmic movement ~ song type *(tempo+pitch+pulse clarity +RMS) +(1 | ID).*

*Rhythmic movement ~ song type *(tempo.sd+pitch.sd+pulse clarity. sd+RMS.sd) +(1 | ID).*

We included infants' expressive vocabularies at 20 months as fixed and interaction effects to examine whether variation in their vocabulary was related to differences in infants' neural tracking of ID singing during Study 1.

*Prediction accuracy ~ song type * vocabulary + channel +(1 | ID).*

Finally, we included infants' expressive vocabularies at 20 months as fixed and interaction effects to examine whether variation in their vocabulary was related to differences in infants' rhythmic movement during ID singing during Study 2.

*Rhythmic movement ~ song type * vocabulary +(1 | ID).*

4. Results

4.1. Acoustic features of maternal singing (Study 1 & 2)

First, we compared the acoustic features of the playsong and the lullaby (see Table 1 for descriptive statistics). Multiple comparisons were corrected for using the false discovery rate, which is indicated by the *q*-value (Benjamini and Hochberg, 1995). Playsongs were sung louder ($\chi^2(1) = 34.15$, $q < .001$), faster ($\chi^2(1) = 40.77$, $q < .001$), and in a higher pitch ($\chi^2(1) = 8.83$, $q = .004$) than lullabies. Playsongs and lullabies did not differ in pulse clarity ($q = .248$). Playsong varied more in perceived loudness ($\chi^2(1) = 35.97$, $q < .001$), tempo ($\chi^2(1) = 6.79$, $q = .012$), and pitch ($\chi^2(1) = 57.50$, $q < .001$) compared to lullabies. Lullabies varied more in pulse clarity compared to playsongs ($\chi^2(1) = 4.505$, $q = .034$).

4.2. Neural tracking of maternal singing (Study 1)

First, we compared the predictive accuracy obtained by the individual models of both playsong and lullaby. The model results displayed a significant fixed effect of ID song type ($\chi^2(1) = 42.135$, $p < .001$), a fixed effect of data type ($\chi^2(1) = 2430.361$, $p < .001$), and an interaction effect between ID song type and data type, $\chi^2(1) = 40.881$, $p < .001$. The set of effects was further tested in post-hoc contrasts (corrected for multiple comparisons) and showed that infants' neural tracking of ID singing (true pairing) was more accurate (estimates = 0.065, SE = 0.002, 95% CI = [0.062 0.068]) in comparison to the randomized control (shifted pairing; estimates = 0.001, SE = 0.002, 95% CI = [0.004 0.004]). Thus, infants showed significant neural tracking of both lullabies and playsongs. In addition, neural tracking of lullabies (estimates = 0.074, SE = 0.002, 95% CI = [0.070 0.077]) was more accurate than that of playsongs (estimates = 0.057, SE = 0.002, 95% CI = [0.053 0.060], Fig. 2A) in data type true, while neural tracking of the two song types did not differ in data type shifted, $p = .999$. Other effects, such as the fixed effect of channel, the interaction effect between channel and type of song, the interaction effect between channel and data type, and the three-way interaction between channel, type of song, and data type were not significant ($p > .086$). In sum, we found above-threshold neural tracking of lullabies and playsongs, while this was more accurate for lullabies than playsongs across the infant brain.

Neural tracking and acoustic features. Next, we examined the effects of tempo, pitch, pulse clarity, RMS on the predictive accuracies of TRF models tested on the original data. First, the mean of acoustic features was included as fixed and interaction effects in the linear mixed-effects model. The model showed that the fixed effect of ID song type remained when controlling for the acoustic features, $\chi^2(1) = 16.164$, $p < .001$. There were, however, significant interaction effects between ID song type and RMS, $\chi^2(1) = 14.264$, $p < .001$ (Fig. 2B), ID song type and tempo, $\chi^2(1) = 4.925$, $p = .026$ (Fig. 2 C), as well as ID song type and pulse clarity, $\chi^2(1) = 9.796$, $p = .002$ (Fig. 2D). Post-hoc analyses of the interaction effects showed that lower RMS (perceived loudness) was related to more accurate neural tracking of the playsong (*trend* = 0.015, *SE* = 0.004, *95% CI* = [−0.024 −0.008]), while RMS was not related to neural tracking accuracy of lullabies (*trend* = −0.003, *SE* = 0.003, *95% CI* = [−0.009 0.002]). Moreover, slower-sung lullabies were related to more accurate neural tracking (*trend* = −0.008, *SE* = 0.004, *95% CI* = [−0.016 −0.001]), while the tempo of playsongs did not relate to neural tracking (*trend* = 0.000, *SE* = 0.002, *95% CI* = [−0.004 0.005]). Higher pulse clarity in playsongs was further related to enhanced neural tracking (*trend* = 0.010, *SE* = 0.003, *95% CI* = [0.004 0.015]) but did not affect neural tracking of lullabies (*trend* = −0.003, *SE* = 0.002, *95% CI* = [−0.008 0.002]).

Neural tracking and acoustic variability. In the next step, we tested the variability of the acoustic features as fixed and interaction effects on neural tracking. We found significant fixed effects of ID song type ($\chi^2(1) = 8.866$, $p = .003$) and variability of tempo ($\chi^2(1) = 5.972$, $p = .015$), indicating that neural tracking was enhanced when the tempo of ID songs varied less, *estimate* = −0.008, *SE* = 0.003, *95% CI* = [− 0.013 −0.002]. There were also significant interaction effects between ID song type and RMS variability, $\chi^2(1) = 11.762$, $p = .001$, as well as ID song type and pitch variability, $\chi^2(1) = 11.329$, $p = .008$. Post-hoc analyses showed that lower RMS variability was related to

Table 1
Descriptive statistics on acoustic features of live-sung lullabies and playsongs (Study 1 & 2).

	Lullaby				Playsong			
Variables	*M*	*SD*	*min*	*max*	*M*	*SD*	*min*	*max*
RMS (perceived loudness)	0.045	0.035	0.009	0.184	0.056	0.033	0.009	0.210
RMS SD	0.028	0.020	0.006	0.117	0.037	0.021	0.005	0.136
Tempo (BPM)	119.00	8.35	100.28	143.57	128.11	14.27	97.99	158.13
Tempo SD (BPM)	33.143	6.174	21.808	44.532	35.842	6.770	20.869	48.351
Pitch (Fundamental frequency f0, Hz)	260.51	36.29	167.39	368.70	268.64	32.71	168.04	322.56
Pitch SD (Hz)	34.138	6.385	23.297	60.007	42.222	6.229	0.011	0.009
Pulse clarity	0.059	0.030	0.010	0.147	0.065	0.026	0.008	0.145
Pulse clarity SD	0.064	0.011	0.047	0.104	0.061	0.009	0.039	0.081

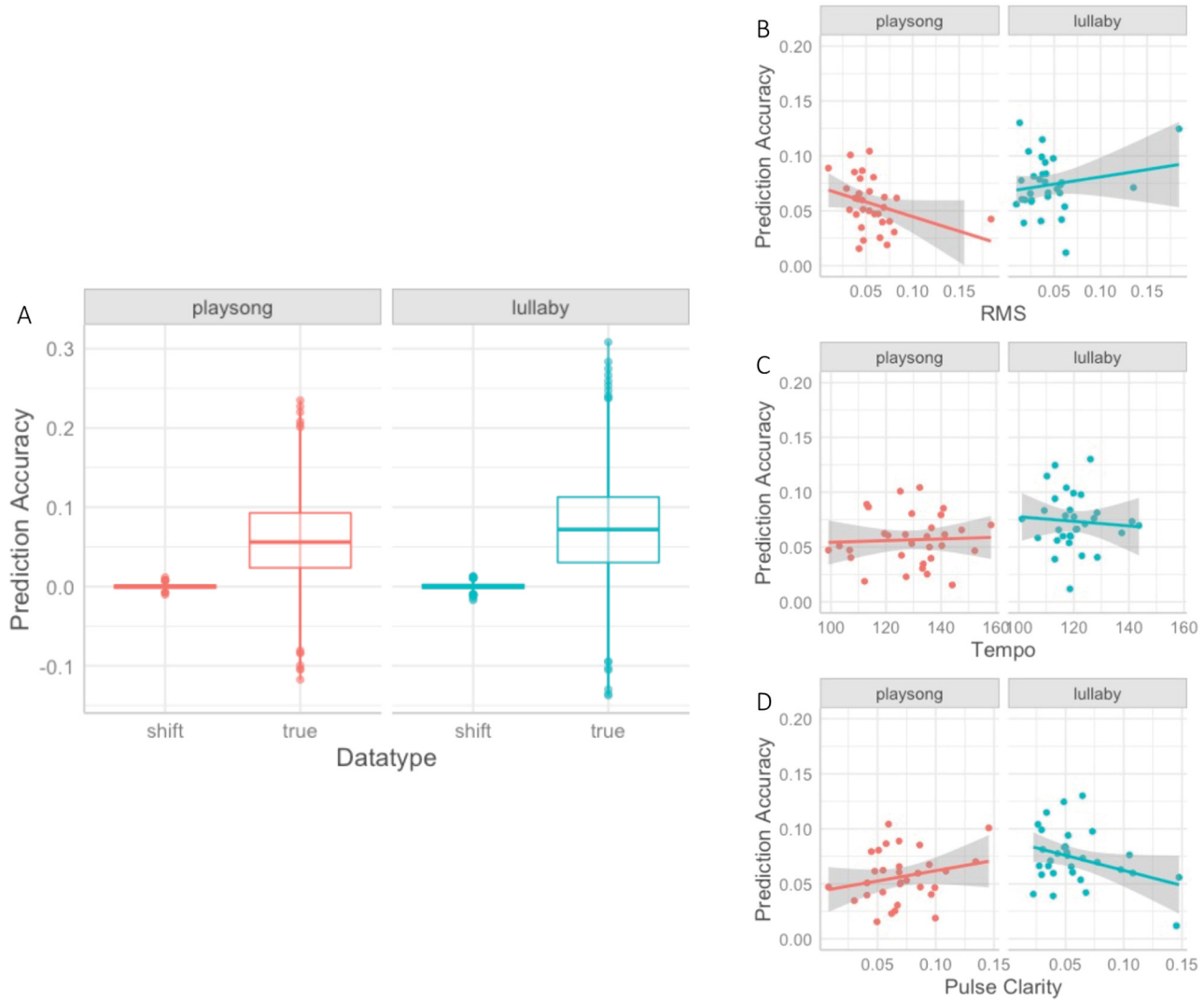


Fig. 2. (A) Graph depicting the interaction effect between true vs. shifted data (x Axis) in the different types of ID songs (columns) on prediction accuracy (y Axis). The difference in prediction accuracy (y Axis) between song types (columns) is modulated by (B) the perceived loudness of playsongs (x Axis), (C) the tempo of lullabies (x Axis), and (D) the pulse clarity of playsongs (x Axis).

more accurate neural tracking of playsongs (*trend* = −0.011, *SE* = 0.004, *95% CI* = [−0.019 −0.004]), while RMS variability was not related to neural tracking of lullabies (*trend* = −0.000, *SE* = 0.003, *95% CI* = [−0.006 0.005]). Lower pitch variability was related to enhanced neural tracking of lullabies (*trend* = −0.006, *SE* = 0.003, *95% CI* = [−0.012 −0.001]), while it was not significantly associated with neural tracking of playsongs (*trend* = 0.006, *SE* = 0.004, *95% CI* = [−0.001 0.013]). All other acoustic features were not related to predictive accuracy ($p > .063$).

4.3. Rhythmic movement to maternal singing (Study 1 & 2)

Based on planned analysis, we conducted rhythmic movement analyses separately for both studies. The overall model, including both studies together, is reported in the Supplements.

Effects of song type on rhythmic movements (Study 1). Generalized linear models reveal that rhythmic movement did not differ significantly between song types in Study 1 ($z = 1.418$, $p = .156$; Fig. 3, left panel).

Neural tracking and rhythmic movement (Study 1). We tested whether rhythmic movements in infants were related to neural tracking of maternal singing. Here, the model outputs showed a significant fixed effect of rhythmic movement, $\chi^2(1) = 10.816$, $p = .001$, indicating that

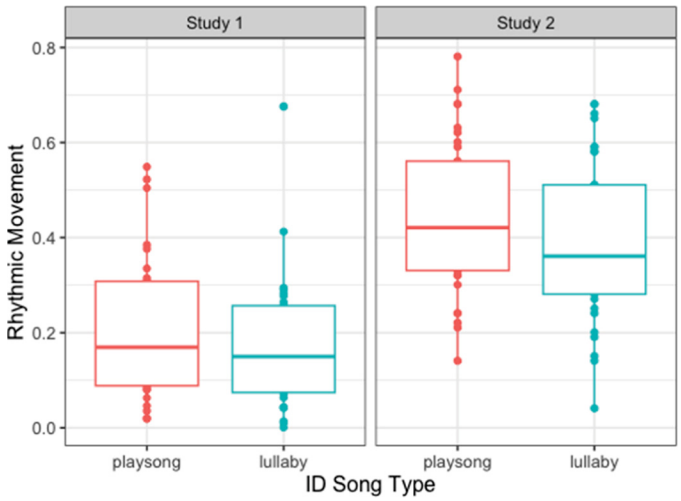


Fig. 3. Graph depicting the fixed effect of the type of ID song (x Axis) on rhythmic movement frequency proportionate to the condition duration (y Axis) in Study 1 (left panel) and Study 2 (right panel).

more frequent rhythmic movements were related to more accurate neural tracking of maternal singing, independent of song type, *estimate* = 0.027, *SE* = 0.011, *95% CI* = [0.007 0.049]. Importantly, the fixed effect of song types remained, $\chi^2(1) = 34.594$, $p < .001$ (lullabies were better tracked than playsongs), while the interaction effect between rhythmic movement and song type was not significant, $p = .941$.

Effects of song type on rhythmic movements (Study 2). Next, we ran the planned and more detailed rhythmic movement analysis (including acoustic features and language) for Study 2. Infants' rhythmic movement differed significantly between song types in Study 2 ($z = 2.172$, $p = .029$, Fig. 3, right panel).

Acoustic features and rhythmic movement (Study 2). Next, we tested the role of the acoustic features of maternal ID singing on infants' rhythmic movement. We found no significant association between mean values and standard deviations of acoustic features and the amount of rhythmic movement ($p > .070$).

4.4. Relations between neural tracking (Study 1), rhythmic movement (Study 2), and children's language development

Finally, we examined whether infants' neural tracking and rhythmic movement in response to maternal ID singing are related to their expressive vocabulary at 20 months. The model outputs for Study 1 (N = 27) revealed a significant interaction effect of neural tracking and ID song type on vocabulary size at 20 months, $\chi^2(1) = 5.236$, $p = .022$. The post-hoc analysis (see Fig. 4A) revealed a significant, but not robust, trend for enhanced neural tracking of playsongs in relation to larger vocabulary size at 20 months (*trend* = 0.002, *SE* = 0.002, *95% CI* = [−0.002 0.006]), and no significant relation between neural tracking of lullabies and vocabulary size (*trend* = −0.002, *SE* = 0.002, *95% CI* = [−0.006 0.002]). Results for Study 2 (N = 33) showed that infants' rhythmic movement was positively related to infants' expressive vocabulary depending on the song type, $\chi^2(1) = 4.933$, $p = .026$. Specifically, post-hoc trend analysis showed that infants' rhythmic movement during the playsong, but not the lullaby, was positively related to their vocabulary (see Fig. 4B). The trend was, however, not robust (*trend* = 0.072, *SE* = 0.068, *95% CI* = [−0.062 0.205]).

5. General discussion

In the two studies reported here, we investigated 7-month-old infants' neural tracking of and rhythmic movement to live and dynamic maternal ID singing. In addition, we tested whether and how these forms of coordination are related to children's language development. We found that 7-month-old infants coordinate both at the neural and behavioral levels with their mothers' live ID singing. In line with our hypotheses, infants showed better neural tracking of lullabies than playsongs (in Study 1), while they displayed more rhythmic movements during playsongs than during lullabies (in Study 2). Interestingly, only infants' neural and behavioral coordination with their mothers' playsongs, but not lullabies, predicted their vocabulary size at 20 months.

Infants showed above-threshold neural tracking of live-sung ID songs in comparison to a shuffled control. This finding highlights that infants are able to neurally track live musical communication beyond rhythmic pure tones (Cirelli et al., 2016), recordings of nursery rhymes (Attaheri et al., 2022), or infant-directed speech (Menn et al., 2022). Neural tracking reflects the degree to which brain activity regresses onto a continuous stimulus (such as volume fluctuations in song; e.g., Jessen et al., 2019; Kalashnikova et al., 2018). Specifically, our encoding approach, based on TRF, quantifies the degree to which infant ongoing brain activity can be "predicted" by the sound envelope of maternal singing of a lullaby vs. playsong. The closer the infant brain tracks the stimulus, the higher the derived encoding accuracy, that is, the better we can estimate the individual baby's brain response based on the sound envelope of their mother's singing. In adults, directing top-down selective attention towards a stimulus associates with increased neural tracking (e.g., Zion Golumbic et al., 2013). These results indicate that infants' neural tracking of live-sung ID songs could reflect attentional processes as well. In addition, neural tracking depends on stimulus features that can facilitate or obstruct neural processing. For instance, maternal infant-directed speech contains amplitude modulations that make it particularly easy to track for infants (Menn et al., 2022).

Interestingly, our results indicate that infants were more accurate at tracking lullabies than playsongs even when controlling for concurrent rhythmic movements. Lullabies are sung with a regular and repetitive pattern (Cirelli et al., 2020; Trainor et al., 1997) that could help infants to neurally track this type of song. This interpretation is supported by acoustic analyses of maternal singing in both studies showing that lullabies, compared to playsongs, were sung slower, lower in pitch, and less variable, possibly making them more predictable. These individual acoustic features of lullabies were further related to enhanced neural tracking, such that individual lullabies sung at a slower tempo also showed enhanced neural tracking (Fig. 2C). This is consistent with previous findings with adults which show enhanced neural tracking of slower music (Weineck et al., 2022). Our results thus corroborate the relation between a clear and predictable structure in tempo and pitch

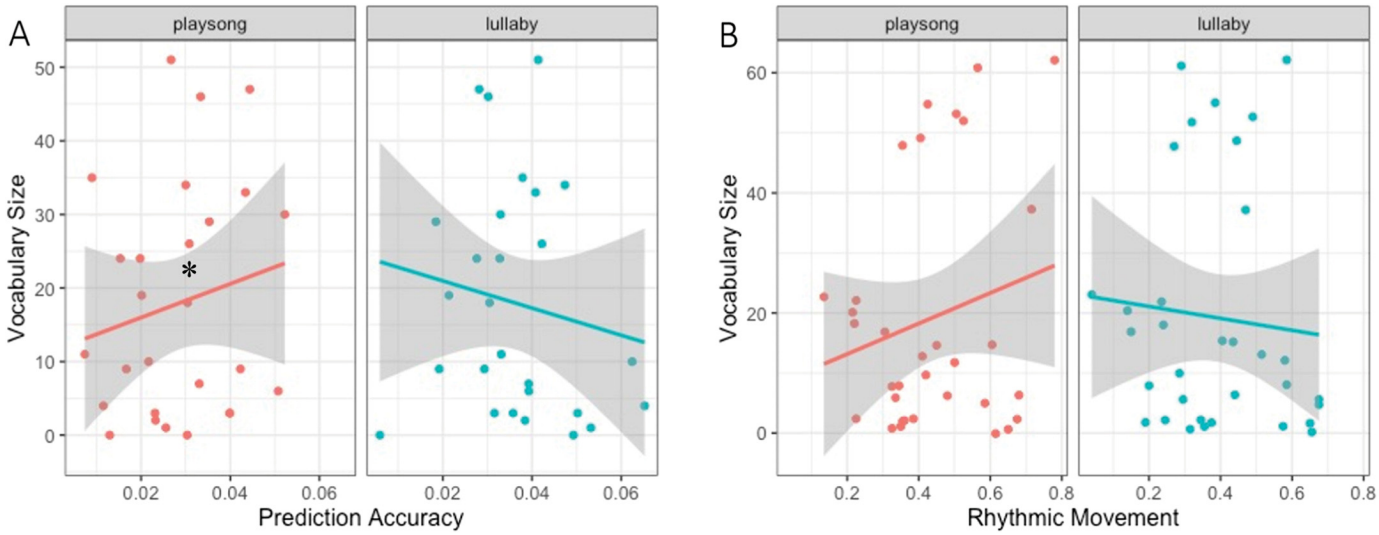


Fig. 4. (A) Graph depicting the relation between neural tracking (x Axis) and (B) rhythmic movement at 7 months (x Axis) in playsongs (red) and lullaby (blue) on estimated means of vocabulary size at 20 months (y Axis). Shaded area indicates the 95% confidence interval.

and potential ease of neural coordination. This increased clarity in tempo- and pitch-related structures in lullabies might also be particularly associated with their regulatory functions, such as modulating emotions (Trehub et al., 2015) and maintaining infant attention and composure (Shenfield et al., 2003). Despite their more variable and faster structure, infants were also able to neurally track playsongs (Study 1) and actually showed more rhythmic movement during playsongs than lullabies (Study 2). In line with our argumentation, neural coordination was enhanced when caregivers sang with higher pulse clarity, quieter, and less variability in loudness. We, therefore, suggest that caregivers might counterbalance the complexity of playsongs by enhancing their rhythmic structure (Zentner and Eerola, 2010) as well as using loudness to guide infants' attention (Calignano et al., 2021). In previous research, pulse clarity was also related to infants' amounts of rhythmic movement (Zentner and Eerola, 2010). Generally, our results suggest that variability in acoustic features, potentially up to a certain threshold, does not hinder infants' neural tracking of ID songs. Instead, increased rhythmic variability might increase engagement (Cameron et al., 2022). The discrepancy in rhythmic movement frequency between playsongs and lullabies is likely further pronounced by infants' familiarity with the songs (see Supplements) and potentially their expectation to engage in an activating social context. Alternatively, rhythmic movement might arise to help infants process these acoustically more variable ID songs, as we found that more accurate neural tracking in infants was related to more rhythmic movements irrespective of song type. Even though in our studies we could not distinguish between musical and non-musical rhythmic behavior, previous research suggests that the perception of rhythm while listening to music often involves spontaneous movement on these periodicities (Hurley et al., 2014; Toiviainen et al., 2010). In adults, rhythmic motor activity facilitates processing of on-beat tones through temporal alignment of attention fluctuations (Morillon et al., 2014; Zalta et al., 2020). Music stimulates movement and activates motor pathways (Grahn and Brett, 2007; Vuilleumier and Trost, 2015), and already 5-month-olds move rhythmically significantly more to music than speech, suggesting that there may be a predisposition for rhythmic movement in response to music (Zentner and Eerola, 2010). In the current study, the types of rhythmic movements included were 1) whole-body movements where infants used the footstep of the highchair to move their body upwards, and 2) kicking/tapping their feet or flapping their arms and moving their fingers. Interestingly, passive body movement was also found to influence 7-month-old infants' encoding of musical rhythms (Phillips-Silver and Trainor, 2005), indicating that their own motor experiences may contribute to young infants' processing of more variable acoustic input.

Importantly, we found that infants' diverging coordination to live and dynamic lullaby and playsong was also reflected in the association with language outcomes at 20 months. Results showed that infants' neural tracking of and rhythmic movement to playsongs, but not lullabies, were positively related to their later vocabulary size, even though the relation was not robust. This finding supports the proposition that musicality evolved as a common ability for infants to decipher music and language (Brandt et al., 2012). Preverbal infants, who are not yet familiar with the rules of language, classify speech according to pitch, melody, and rhythm in order to understand its meaning (Goswami, 2012). It is precisely prosodic information (i.e., slow speech rhythms) that is crucial for language acquisition (Goswami, 2012; Langus et al., 2017). Neural tracking of the more variable playsong might, for example, relate to infants' ability to extract prosodic information from a vocal stream and thus possibly support their word segmentation abilities (Menn et al., 2022), marked by larger expressive vocabulary size at age 20 months. ID singing may be particularly helpful for language acquisition, in addition to ID speech, due to its metric structure that makes prosodic stress more pronounced (Nakata and Trehub, 2011; Suppanen et al., 2019; Trainor et al., 1997). In addition, moving rhythmically could help infants process the prosodic information from the song, a notion supported by arguments that rhythmic abilities are a central part of language acquisition (Thomson and Goswami, 2008). An alternative explanation we cannot rule out is that infants' movements affected the EEG response; movement-related artifacts in the EEG might have increased infant neural tracking in the playsong and thus the relation between neural tracking and infants' later language development. Nonetheless, our study provides promising evidence for the connection between early musical experiences and language development.

The paradigm has some limitations. Mothers' singing was highly varied on the individual level, which did not allow us to make stronger claims about the relation of specific acoustic features (such as pitch and tempo) in relation to neural tracking. Hence, future studies might consider more controlled stimuli that vary less across infants (e.g., Lense et al., 2022; Nguyen et al., 2023; Weineck et al., 2022) or use methodological approaches that can take into account how the variance in stimuli relates to infants' tracking/attention over time. In addition, we need to consider the fit of the musical piece to the particular situation in which we observed infants. The context of the laboratory setting likely afforded the playsong more than the lullaby, as the mothers and the experimenter intended to keep infants attentive and entertained. Thus, the congruence between the observational situation and the playsong may have contributed to our pattern of results. Moreover, it is possible that mothers themselves moved differently during the two types of songs, and infants simply tracked and imitated their mothers' movement type and frequency. Playsongs, in particular, are not only expressed through voice but also the body and are often intrinsically multimodal (Eckerdal and Merker, 2009; Lense et al., 2022). In fact, our supplementary analyses showed that infants looked significantly longer at their mothers during the playsong than during the lullaby. However, infant gaze behavior during the playsong was neither associated with their neural tracking nor rhythmic movements in that condition. Instead, we found that neural tracking of lullabies was weaker when infants looked away from mothers' faces and bodies, suggesting that neural tracking is related to infants' attention towards the face (Lense et al., 2022). To fully disentangle the factors underlying infants' neural tracking and rhythmic movements during ID songs, future studies need to include more in-depth analyses of the accompanying sensorimotor synchronization processes, including caregivers' movements (e.g., rocking, playful gestures, etc.).

6. Conclusion

Singing represents a flexible tool for caregivers to meet the various needs of their infants. However, it remains a question of great theoretical controversy why ID singing is such a powerful communication method with very young infants (Mehr et al., 2021; Savage et al., 2021). Markova et al. (2019, 2020) argued that ID singing has inherent acoustic characteristics that support infants' coordination to a particular interactive rhythm purported by the caregiver, and this process could have important developmental implications. Findings of the present study corroborate these hypotheses by showing that lullabies may be used to help infants track a slower and regular rhythmic pattern, while playsongs, containing more variability, could be perceived as more engaging and thus make infants move rhythmically. The acoustic variability of playsongs also seems to promote language acquisition in that rhythmic variability may draw attention to the songs' prosodic information (Hannon and Johnson, 2005). These findings highlight potential avenues for future studies to test the neural and behavioral basis of the communicative function of mother-infant musical interaction by experimentally manipulating the relevant acoustic features. With its semi-naturalistic design, the present study provides a better understanding of the dynamics, mechanisms, and developmental outcomes of early musical interactions.

Funding

This research was partly funded by a doctoral stipend from the

Studienstiftung des Deutschen Volkes granted to T.N. and by the Austrian Science Fund (FWF) DK Grant "Cognition & Communication" (W1262-B29).

CRediT authorship contribution statement

Trinh Nguyen: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration **Susanne Reisner**: Data curation, Writing – original draft, Writing – review & editing, Visualization **Anja Lueger**: Investigation, Data – curation, Writing – original draft **Sam V. Wass**: Methodology, Formal analysis, Writing – review & editing **Stefanie Hoehl**: Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition **Gabriela Markova**: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to Liesbeth Forsthuber, Christina Schaetz, Martina de Eccher, Sarah Paul, Marlies Schermann, Anna Matyk, Rahel Loos, Madlene Radosavljevic, Pierre Labendzki, Jakob Weickmann and Elisa Roberti for their support in data acquisition, processing, and video coding. Additionally, we thank all families who participated in the study and the Department of Obstetrics and Gynecology of the Vienna General Hospital for supporting our participant recruitment.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2023.101313.

References

Attaheri, A., Choisdealbha, Á.N., Di Liberto, G.M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., Goswami, U., 2022. Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. NeuroImage 247, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698.

Attaheri, A., Choisdealbha, Á.N., Rocha, S., Brusini, P., Liberto, G.M.D., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Oliveira, M.A. e, Brough, C., Flanagan, S., Goswami, U. 2022. Infant low-frequency EEG cortical power, cortical tracking and phase-amplitude coupling predicts language a year later (p. 2022.11.02.514963). bioRxiv. https://doi.org/10.1101/2022.11.02.514963.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc.: Ser. B (Methodol.) 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Brandt, A., Slevc, R., Gebrian, M., 2012. Music and early language acquisition. Front. Psychol. 3, 327. https://doi.org/10.3389/fpsyg.2012.00327.

Calignano, G., Dispaldro, M., Russo, S., Valenza, E., 2021. Attentional engagement during syllable discrimination: the role of salient prosodic cues in 6- to 8-month-old infants. Infant Behav. Dev. 62, 101504 https://doi.org/10.1016/j.infbeh.2020.101504.

Cameron, D.J., Caldarone, N., Psaris, M., Carrillo, C., Trainor, L.J., 2022. The complexity-aesthetics relationship for musical rhythm is more fixed than flexible: evidence from children and expert dancers. Dev. Sci., e13360 https://doi.org/10.1111/desc.13360.

de Cheveigné, A., 2020. ZapLine: a simple and effective method to remove power line artifacts. NeuroImage 207, 116356. https://doi.org/10.1016/j.neuroimage.2019.116356.

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

Cirelli, L.K., Jurewicz, Z.B., Trehub, S.E., 2020. Effects of maternal singing style on mother–infant arousal and behavior. J. Cogn. Neurosci. 32 (7), 1213–1220. https://doi.org/10.1162/jocn_a_01402.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Di Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Mehta, A.D., Herrero, J.L., de Cheveigné, A., Shamma, S., Mesgarani, N., 2020. Cortical encoding of melodic expectations in human temporal cortex. ELife 9, e51784. https://doi.org/10.7554/eLife.51784.

Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85, 761–768.

Eckerdal, P., Merker, B., 2009. "Music" and the "action song" in infant development: An interpretation. In *Communicative musicality: Exploring the basis of human companionship*. Oxford University Press,, pp. 241–262.

Ferland, M.B., Mendelson, M.J., 1989. Infants' categorization of melodic contour. Infant Behav. Dev. 12 (3), 341–355. https://doi.org/10.1016/0163-6383(89)90043-X.

François, C., Teixidó, M., Takerkart, S., Agut, T., Bosch, L., Rodriguez-Fornells, A., 2017. Enhanced neonatal brain responses to sung streams predict vocabulary outcomes by age 18 months. Sci. Rep. 7 (1), 12451 https://doi.org/10.1038/s41598-017-12798-2.

Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. J. Cogn. Neurosci. 19 (5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893.

Háden, G.P., Honing, H., Török, M., Winkler, I., 2015. Detecting the temporal structure of sound sequences in newborn infants. Int. J. Psychophysiol. 96 (1), 23–28. https://doi.org/10.1016/j.ijpsycho.2015.02.024.

Hannon, E.E., Johnson, S.P., 2005. Infants use meter to categorize rhythms and melodies: implications for musical structure learning. Cogn. Psychol. 50 (4), 354–377. https://doi.org/10.1016/j.cogpsych.2004.09.003.

Hurley, B.K., Martens, P.A., Janata, P., 2014. Spontaneous sensorimotor coupling with multipart music. J. Exp. Psychol.: Hum. Percept. Perform. 40, 1679–1696. https://doi.org/10.1037/a0037154.

Ilari, B., 2015. Rhythmic engagement with music in early childhood: a replication and extension. J. Res. Music Educ. 62 (4), 332–343. https://doi.org/10.1177/0022429414555984.

Iverson, J.M., 2010. Developing language in a developing body: the relationship between motor development and language development. J. Child Lang. 37 (2), 229. https://doi.org/10.1017/S0305000909990432.

Iverson, J.M., Hall, A.J., Nickel, L., Wozniak, R.H., 2007. The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. Brain Lang. 101 (3), 198. https://doi.org/10.1016/j.bandl.2006.11.004.

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

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

Kalashnikova, M., Peter, V., Di Liberto, G.M., Lalor, E.C., Burnham, D., 2018. Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. Article 1 Sci. Rep. 8 (1). https://doi.org/10.1038/s41598-018-32150-6.

Kisilevsky, B. s, Hains, S. m j, Jacquet, A.-Y., Granier-Deferre, C., Lecanuet, J., 2004. Maturation of fetal responses to music (p). Dev. Sci. 7 (5), 550–559. https://doi.org/10.1111/j.1467-7687.2004.00379.x.

Kragness, H.E., Anderson, L., Chow, E., Schmuckler, M., Cirelli, L.K., 2022. Musical groove shapes children's free dancing. Dev. Sci. https://doi.org/10.1111/desc.13249.

Krumhansl, C.L., Jusczyk, P.W., 1990. Infants' perception of phrase structure in music. Psychol. Sci. 1 (1), 70–73. https://doi.org/10.1111/j.1467-9280.1990.tb00070.x.

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

Lartillot, O., Toiviainen, P., Eerola, T., 2008. A Matlab Toolbox for Music Information Retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (Eds.), Data Analysis, Machine Learning and Applications. Springer, pp. 261–268. https://doi.org/10.1007/978-3-540-78246-9_31.

Lebedeva, G.C., Kuhl, P.K., 2010. Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. Infant Behav. Dev. 33, 419–430. https://doi.org/10.1016/j.infbeh.2010.04.006.

Lenc, T., Peter, V., Hooper, C., Keller, P.E., Burnham, D., Nozaradan, S., 2022. Infants show enhanced neural responses to musical meter frequencies beyond low-level features (n/a(n/a)). Dev. Sci., e13353. https://doi.org/10.1111/desc.13353.

Lense, M.D., Shultz, S., Astésano, C., Jones, W., 2022. Music of infant-directed singing entrains infants' social visual behavior. Proc. Natl. Acad. Sci. 119 (45), e2116967119 https://doi.org/10.1073/pnas.2116967119.

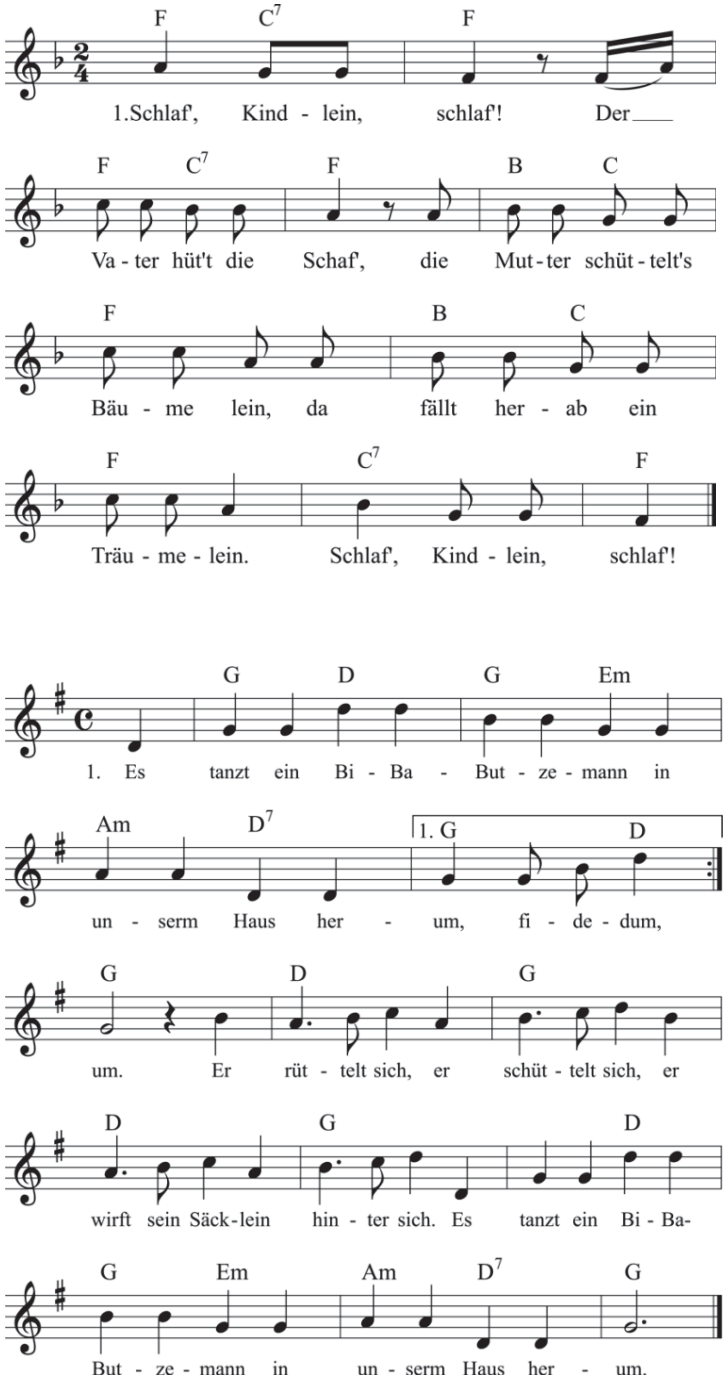
Lopez, K.L., Monachino, A.D., Morales, S., Leach, S.C., Bowers, M.E., Gabard-Durnam, L. J., 2022. HAPPILEE: HAPPE in low electrode electroencephalography a standardized pre-processing software for lower density recordings. NeuroImage 260, 119390. https://doi.org/10.1016/j.neuroimage.2022.119390.

Markova, G., Legerstee, M., 2006. Contingency imitation and affect sharing: foundations of infants' social awareness. Dev. Psychol. 42 (1), 132–141. https://doi.org/10.1037/0012-1649.42.1.132.

Markova, G., Nguyen, T., Hoehl, S., 2019. Neurobehavioral interpersonal synchrony in early development: the role of interactional rhythms. Front. Psychol. 10, 2078 https://doi.org/10.3389/fpsyg.2019.02078.
Markova, G., Nguyen, T., Schätz, C., de Eccher, M., 2020. Singing in tune – being in tune: relationship between maternal playful singing and interpersonal synchrony (Cairn.info). Enfance 1 (1), 89–107. https://doi.org/10.3917/enf2.201.0089.
Marschik, P.B., Einspieler, C., Garzarolli, B., Prechtl, H.F.R., 2007. Events at early development: are they associated with early word production and neurodevelopmental abilities at the preschool age. Early Hum. Dev. 83 (2), 107–114. https://doi.org/10.1016/j.earlhumdev.2006.05.009.
Mehr, S.A., Krasnow, M.M., Bryant, G.A., Hagen, E.H., 2021. Origins of music in credible signaling. Behav. Brain Sci. 44, e60 https://doi.org/10.1017/S0140525×20000345.
Mendoza, J.K., Fausey, C.M., 2021. Everyday music in infancy. Dev. Sci. 24 (6) https://doi.org/10.1111/desc.13122.
Menn, K.H., Michel, C., Meyer, L., Hoehl, S., Männel, C., 2022. Natural infant-directed speech facilitates neural tracking of prosody. NeuroImage 251, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991.
Menn, K.H., Ward, E.K., Braukmann, R., van den Boomen, C., Buitelaar, J., Hunnius, S., Snijders, T.M., 2022. Neural tracking in infancy predicts language development in children with and without family history of Autism. Neurobiol. Lang. 3 (3), 495–514. https://doi.org/10.1162/nol_a_00074.
Meyer, M., Lamers, D., Kayhan, E., Hunnius, S., Oostenveld, R., 2021. Enhancing reproducibility in developmental EEG research: BIDS, cluster-based permutation tests and effect sizes. Dev. Cogn. Neurosci. 52, 101036 https://doi.org/10.1016/j.dcn.2021.101036.
Morillon, B., Schroeder, C.E., Wyart, V., 2014. Motor contributions to the temporal precision of auditory attention. Nat. Commun 5, 1. https://doi.org/10.1038/ncomms6255.
Nakata, T., Trehub, S.E., 2011. Expressive timing and dynamics in infant-directed and non-infant-directed singing. Psychomusicology: Music Mind Brain 21 (1–2), 45–53. https://doi.org/10.1037/h0094003.
Nguyen, T., Flaten, E., Trainor, L.J., Novembre, G., 2023. Early social communication through music: State of the art and future perspectives. Developmental cognitive neuroscience 63, 101279. https://doi.org/10.1016/j.dcn.2023.101279.
Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. 23 (11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2010. FieldTrip: open source software for advanced analysis of MEG, EEG, and Invasive electrophysiological data. Comput. Intell. Neurosci. 2011, e156869 https://doi.org/10.1155/2011/156869.
Papadimitriou, A., Smyth, C., Politimou, N., Franco, F., Stewart, L., 2021. The impact of the home musical environment on infants' language development. Infant Behav. Dev. 65, 101651 https://doi.org/10.1016/j.infbeh.2021.101651.
Phillips-Silver, J., Trainor, L.J., 2005. Feeling the beat: movement influences infant rhythm perception, 1430–1430 Science 308 (5727). https://doi.org/10.1126/science.1110922.
Phillips-Silver, J., Trainor, L.J., 2007. Hearing what the body feels: Auditory encoding of rhythmic movement. Cognition 105 (3), 533–546.
Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. Nat. Rev. Neurosci. 21 (6), 322–334. https://doi.org/10.1038/s41583-020-0304-4.
Politimou, N., Dalla Bella, S., Farrugia, N., Franco, F., 2019. Born to speak and sing: musical predictors of language development in Pre-schoolers. Front. Psychol. 10 https://doi.org/10.3389/fpsyg.2019.00948.
Putkinen, V., Saarikivi, K., Tervaniemi, M., 2013. Do informal musical activities shape auditory skill development in preschool-age children? Frontiers in psychology 4, 572. https://doi.org/10.3389/fpsyg.2013.00572.
Rock, A.M.L., Trainor, L.J., Addison, T.L., 1999. Distinctive messages in infant-directed lullabies and play songs. Dev. Psychol. 35 (2), 527–534. https://doi.org/10.1037/0012-1649.35.2.527.
Savage, P.E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., Fitch, W.T., 2021. Music as a coevolved system for social bonding. Behav. Brain Sci. 44, e59 https://doi.org/10.1017/S0140525×20000333.
Shenfield, T., Trehub, S.E., Nakata, T., 2003. Maternal singing modulates infant arousal. Psychol. Music 31 (4), 365–375. https://doi.org/10.1177/03057356030314002.
Sievers, B., Polansky, L., Casey, M., Wheatley, T., 2013. Music and movement share a dynamic structure that supports universal expressions of emotion. Proc. Natl. Acad. Sci. 110 (1), 70–75. https://doi.org/10.1073/pnas.1209023110.
Steinberg, S., Liu, T., Lense, M.D., 2021. Musical engagement and parent-child attachment in families with young children during the Covid-19 pandemic. Front. Psychol. 12, 641733 https://doi.org/10.3389/fpsyg.2021.641733.
Suppanen, E., Huotilainen, M., Ylinen, S., 2019. Rhythmic structure facilitates learning from auditory input in newborn infants. Infant Behav. Dev. 57, 101346 https://doi.org/10.1016/j.infbeh.2019.101346.
Thelen, E., 1979. Rhythmical stereotypies in normal human infants. Anim. Behav. 27, 699–715. https://doi.org/10.1016/0003-3472(79)90006-X.
Thiessen, E.D., Saffran, J.R., 2009. How the melody facilitates the message and vice versa in infant learning and memory. Ann. N. Y. Acad. Sci. 1169 (1), 225–233. https://doi.org/10.1111/j.1749-6632.2009.04547.x.
Thomson, J.M., Goswami, U., 2008. Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. J. Physiol. 102 (1), 120–129. https://doi.org/10.1016/j.jphysparis.2008.03.007.
Toiviainen, P., Luck, G., Thompson, M.R., 2010. Embodied meter: hierarchical eigenmodes in music-induced movement. Music Percept. 28 (1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59.
Trainor, L.J., Heinmiller, B.M., 1998. The development of evaluative responses to music. Infant Behav. Dev. 21 (1), 77–88. https://doi.org/10.1016/S0163-6383(98)90055-8.
Trainor, L.J., Clark, E.D., Huntley, A., Adams, B.A., 1997. The acoustic basis of preferences for infant-directed singing. Infant Behav. Dev. 20 (3), 383–396. https://doi.org/10.1016/S0163-6383(97)90009-6.
Trehub, S.E., 2017. The Maternal Voice as a Special Signal for Infants. In: Filippa, M., Kuhn, P., Westrup, B. (Eds.), *Early Vocal Contact and Preterm Infant Brain Development: Bridging the Gaps Between Research and Practice* (pp. Springer International Publishing, pp. 39–54. https://doi.org/10.1007/978-3-319-65077-7_3.
Trehub, S.E., 2019. Nurturing infants with music. Int. J. Music Early Child. 14 (1), 9–15. https://doi.org/10.1386/ijmec.14.1.9_1.
Trehub, S.E., Trainor, L., 1998. Singing to infants: lullabies and play songs. Adv. Infancy Res. 12, 43–77.
Trehub, S.E., Ghazban, N., Corbeil, M., 2015. Musical affect regulation in infancy. Ann. N. Y. Acad. Sci. 1337 (1), 186–192. https://doi.org/10.1111/nyas.12622.
Trehub, S.E., Plantinga, J., Russo, F.A., 2016. Maternal Vocal interactions with infants: reciprocal visual influences: maternal speech and singing. Soc. Dev. 25 (3), 665–683. https://doi.org/10.1111/sode.12164.
U. Goswami Entraining the Brain: Applications to Language Research and Links to Musical Entrainment 2012 doi: 10.18061/1811/52980.
Vuilleumier, P., Trost, W., 2015. Music and emotions: from enchantment to entrainment. Ann. N. Y. Acad. Sci. 1337 (1), 212–222. https://doi.org/10.1111/nyas.12676.
Warlaumont, A.S., Sobowale, K., Fausey, C.M., 2022. Daylong mobile audio recordings reveal multitimescale dynamics in infants' vocal productions and auditory experiences. Curr. Dir. Psychol. Sci. 31 (1), 12–19. https://doi.org/10.1177/09637214211058166.
Weineck, K., Wen, O.X., Henry, M.J., 2022. Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. ELife 11, e75515. https://doi.org/10.7554/eLife.75515.
Williams, K.E., Barrett, M.S., Welch, G.F., Abad, V., Broughton, M., 2015. Associations between early shared music activities in the home and later child outcomes: findings from the Longitudinal Study of Australian Children. Early Child. Res. Q. 31, 113–124. https://doi.org/10.1016/j.ecresq.2015.01.004.
Winkler, I., Háden, G.P., Ladinig, O., Sziller, I., Honing, H., 2009. Newborn infants detect the beat in music. Proc. Natl. Acad. Sci. 106 (7), 2468–2471. https://doi.org/10.1073/pnas.0809035106.
Zalta, A., Petkoski, S., Morillon, B., 2020. Natural rhythms of periodic temporal attention. Nat. Commun. 11, 1. https://doi.org/10.1038/s41467-020-14888-8.
Zentner, M.R., Eerola, T., 2010. Rhythmic engagement with music in infancy. Proc. Natl. Acad. Sci. 107 (13), 5768–5773. https://doi.org/10.1073/pnas.1000121107.
Zhao, T.C., Kuhl, P.K., 2016. Musical intervention enhances infants' neural processing of temporal structure in music and speech. Proc. Natl. Acad. Sci. 113 (19), 5212–5217. https://doi.org/10.1073/pnas.1603984113.
Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a cocktail party. Neuron 77 (5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037.

Supplementary material to Sing to me, baby:
Infants show neural tracking and rhythmic movement to live and dynamic maternal singing

Figure S1. Notations and lyrics of the lullaby ("Schlaf, Kindlein, schlaf") and the playsong ("Es tanzt ein Bibabutzemann").



Note. The lullaby is a German children's song in F-major, and we primed participants to sing the song at 100 BPM using a metronome. The playsong is a German children's song in G-Major, and participants were primed to sing the song at 170 BPM. Each verse in the songs lasted on average 14.6 s (lullaby) to 22.3 s (playsong), and mothers repeated the verse eight times in total, amounting to average audio lengths of 116.42 s (lullaby; range = 96-207 s) and 178.25 s (playsong; range = 113-232 s).

Maternal depression symptoms and anxiety levels after singing

All participating mothers completed the Edinburgh Postnatal Depression Scale (EPDS; Bergant et al., 1998), to control for depressive mood (M = 4.83, SD = 3.64, *range* = 0-13). As mothers showed subrange (≥13) depression symptoms, we were able to include all dyads in subsequent analyses. Mothers also completed the State-Trait Anxiety Inventory (STAI; Marteau & Bekker, 1992) upon the completion of the experiment, which captured their low to medium anxiety levels after singing (M = 22.51, SD = 11.86, *range* = 1.43 - 47.13).

Removal of background noise including infant vocalizations

We manually removed excerpts containing environmental noises from the audio recording using the software Audacity (https://www.audacityteam.org/). These excerpts included noises the infant made with their body (e.g., hitting against the highchair), vegetative noises (e.g., coughs and burps), infant vocalizations, and other background noise. The removed contaminations did not differ significantly between the lullaby and playsong conditions, both in frequency (V = 201, p = .52) and relative duration (V = 225, p = .89). Additionally, infant vocalizations did not differ significantly in frequency (V = 63.5, p = .07) or relative duration (V = 102, p = .17) between the lullaby and playsong conditions. To ensure inter-rater reliability, two independent observers coded manually removed audio excerpts on whether they contained infant vocalizations in 30% of all audios, yielding high inter-reliability with κ = .93.

Assessment of infant exposure to music

We collected self-reported information on infant exposure to music: Families listened to music for an average of 14.38 hours/week (*SD* = 15.37, *range* = 2-70); mothers sang to infants for an average of 5.86 hours/week (*SD* = 4.79, *range* = 0-20); mothers made music to infants for an average of 3.72 hours/week (*SD* = 3.80, *range* = 0.5-21).

Familiarity with ID songs

We then tested whether infants' familiarity with the songs was associated with how well they tracked or rhythmically moved to the ID songs. Familiarity (based on maternal self-report) was not significantly related to the ID songs, $p > .057$. However, familiarity was related to infants' rhythmic movement to the different types of ID songs. Results revealed a significant interaction effect between the type of ID song and infants' familiarity with the playsong, $X^2(1) = 7.407$, $p = .006$. Post-hoc analysis revealed that infants' difference in rhythmic movement to the two types of ID songs was more pronounced when they were more familiar with the playsong, *contrast estimate* = 0.096, SE = 0.035, 95% CI = [0.027 0.165].

Additional information on TRF

The response model is defined as:

$$r(t,n) = \sum_l \quad w(l,n)s(t-l) + \varepsilon(t,n)$$

where $r(t,n)$ is the neural response, across each channel included in the model *(n)*, across all time points *(t); w* is the TRF that describes the linear transformation of the stimulus *(s)* to the ongoing neural response *(t)*, and ε is the residual error at each channel not explained by the model.

The TRF (w) is estimated using the following equation:

$$w = (S^T S + \lambda I)^{-1} S^T r$$

where $S^T r$ is the result of convolution between the zero-lagged neural response *(r)*, and the speech stimulus at different time lags (S), $S^T S$ is the autocovariance matrix of the stimulus at each time lag which is divided out from the linear relationship modeled in $S^T r$. I is the identity matrix, and λ is the ridge parameter, which enforces a smoothness constraint on the output of the model (w). The output (w) is a time-windows by EEG channels matrix, containing the resulting TRF weights for each channel at each time window (see Crosse et al., 2016 for further explanations). A constant term is included in the model by concatenating a column of ones to the left of *S*.

A method of ridge regression is used to improve the reliability of the estimated coefficients and prevent over-fitting. Ridge regression works by introducing a bias term to the model, which penalizes TRF values as a function of their distance from 0 (Crosse et al., 2016). This reduces over-estimation problems in *w* and reduces the likelihood of high frequency artifacts in the estimated TRF model (Fiedler et al., 2017; Haufe et al., 2014) .

Gaze behavior

Table S1 shows infants' gaze behaviors towards their mother, the tablet, and away from both as proportions according to the respective lengths of the songs. We tested whether infant gaze behavior differed between the two conditions.

Paired t-tests revealed that infants looked significantly longer at their mothers in the playsong condition in comparison to the lullaby condition, $t(70) = 2.56$, $p = .013$. Infants' gaze at either the tablet or away did not significantly differ between conditions, $p = .680$, $p = .452$.

Table S1. Descriptive statistics of infant's gaze behavior

	Lullaby				Playsong			
Variables	*M*	*SD*	*min*	*max*	*M*	*SD*	*min*	*max*
Gaze towards mother's face (%)	12.38	12.10	0.00	70.39	15.73	15.42	0.00	76.08
Gaze towards mother's body (%)	0.08	0.40	0.00	2.40	0.05	0.37	0.00	3.02
Gaze towards tablet (%)	58.29	20.77	0.00	97.92	55.17	22.10	10.40	97.42
Gaze away (%)	28.08	18.22	0.00	70.10	28.31	18.68	0	89.60

Gaze and neural tracking

Next, we examined whether neural tracking in both conditions was associated with the infants' gaze behaviors. We, therefore, separately included the infants' gaze behaviors during the playsong and lullaby (i.e., gaze towards mother, gaze at the tablet, and gaze away) as fixed and interaction effects, to examine whether differences in these features were related to differences in infants' neural tracking of ID singing. Infants, who looked away for longer durations during the lullaby condition, also showed lower predictive accuracy, namely weaker neural tracking, *estimate* = -0.038, SE = 0.018, 95% CI = [-0.074 -0.004], $X^2(1) = 11.94$, $p < .001$. Infants' other-looking behavior was not related to how well they neurally tracked their mother's singing, $p > .090$.

Gaze and rhythmic movement

We tested whether infants' gaze behavior affected their rhythmic movement. We, however, found no significant associations between infant gaze and their rhythmic movement, $p > .115$.

Trial duration and neural tracking

Next, we analyzed whether infants' neural tracking was associated with the length of the conditions. We, therefore, added trial durations as a fixed effect to the linear mixed effects model:

*Prediction accuracy ~ song type * trial duration + channel + (1 | ID)*

The model output revealed that trial duration was not significantly related to neural tracking, p=.604. While the interaction effect between song type and trial durations was significant ($X^2(1) = 5.85$, $p = .016$, the estimated slopes were not robust (lullaby: emtrends = 0.002, SE = 0.002, 95% CI = [-0.001 0.004]; playsong: emtrends = -0.003, SE = 0.002, 95% CI = [-0.006 0.000].

Trial duration and gaze behavior

We also analyzed whether infants' gaze behavior was associated with the length of the conditions. We, therefore, added trial durations as a fixed effect to the linear mixed effects model:

*Trial duration ~ song type * (gaze at mother + gaze at body + gaze away) + (1 | ID)*

The model output revealed that trial duration was not significantly related to infants' gaze behavior to the mother, $p = .959$, but that trial duration was shorter when infants looked away for longer ($X^2(1) = 4.45$, $p = .035$, *emtrends* = -51.812, SE = 28.832, 95% CI = -106.65 2.784]. The effect was, however, not robust."

Additional Rhythmic Movement Analyses

We also assessed rhythmic movements in Study 1 and Study 2 together and tested whether infants showed more rhythmic movement during the lullaby or the playsong condition in both studies. The generalized linear mixed-effects model output displayed a significant effect of song type, $\chi^2(1) = 6.474$, $p = .011$. Model estimates revealed that infants showed less rhythmic movement in the lullaby condition (*estimate* = 0.250, *SE* = 0.020, *95% CI* = [0.212 0.292]) than in the playsong condition (*estimate* = 0.296, *SE* = 0.022, *95% CI* = [0.254 0.341]). Given the methodological differences between studies (i.e., that infants were restrained in Study 1 but not in Study 2), we also find significant differences between studies, $\chi^2(1) = 40.893$, $p < .001$. As predicted, infants moved more in Study 2 (*estimate* = 0.410, *SE* = 0.031, *95% CI* = [0.350 0.472] than in Study 1 (*estimate* = 0.168, *SE* = 0.020, *95% CI* = [0.132 0.211]). The interaction effect was not significant, p=.860.

Frequency information of EEG and sound envelopes

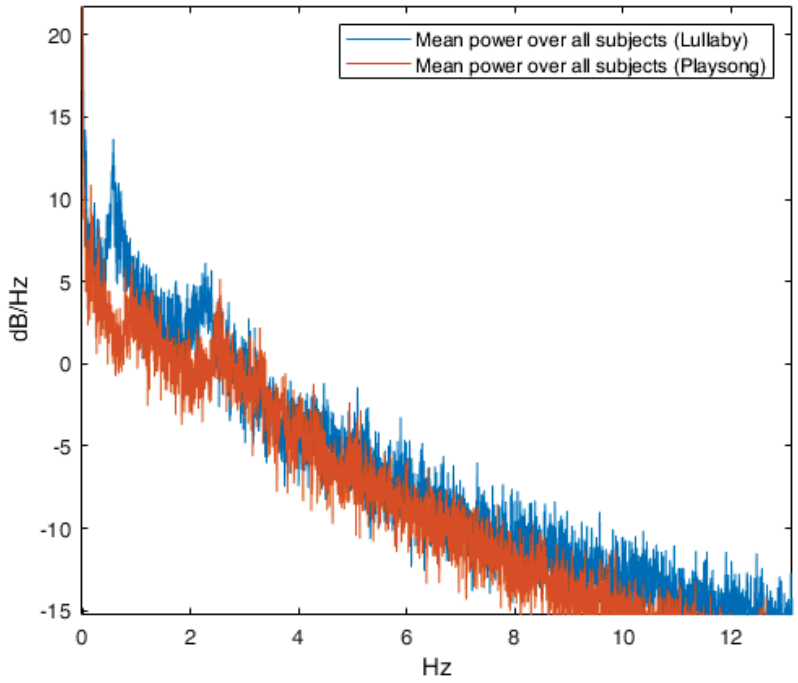


Figure S2. Frequency plot of amplitude-modulated sound envelopes for ID songs averaged over each condition (lullaby = blue, playsong = red). Frequency is plotted on the x Axis, and power in dB/Hz is on the y Axis. The songs show beat- (1.9-3.4 Hz) and meter-related (0.5-1.2 Hz) peaks relevant to the tempo of each condition in the spectrogram.

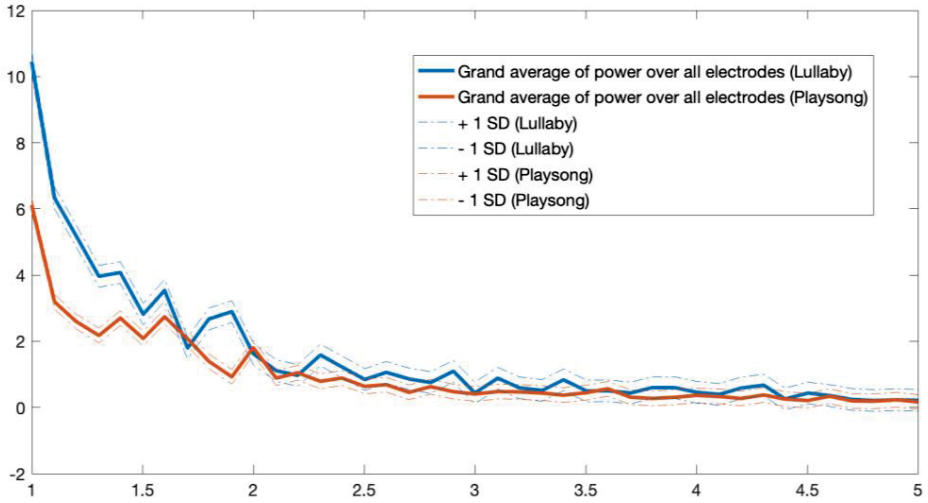


Figure S3. Frequency plot of infants' EEG power averaged over each condition (lullaby = blue, playsong = red) and all channels. Frequency is plotted on the x Axis and logged absolute power is on the y Axis. EEG power in the lullaby conditions shows a slight peak over the beat-related frequencies (averaged at 2.3 Hz), individual visualization of infants' EEG power, however, shows vast individual differences mirroring the individual differences in the acoustic features of the ID songs.

Temporal response functions from "forward" models

We computed the individual temporal response functions (TRF) for the maternal singing envelope in the *playsong* and *lullaby* conditions in *forward* models (see *Figure S3*). TRF were compared against zero. However, we found no significant clusters in the TRF. We assume that either longer and cleaner recordings or more infants are needed to extract a clear temporal response function (Crosse et al., 2021).

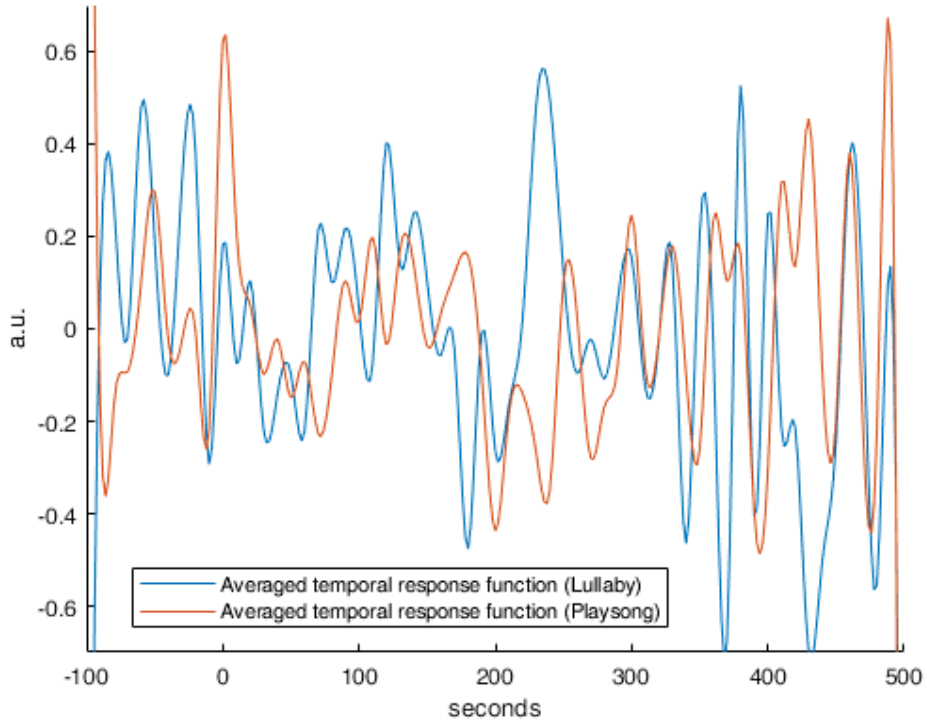


Figure S4. Temporal response functions of infants' EEG signal in frontal channels in the lullaby (blue) and playsong (red) condition. Time is depicted on the x Axis, and the weights are depicted on the y Axis. The weights were neither significantly different from zero nor between each other, $p > .30$.

References

Bergant, A. M., Nguyen, T., Heim, K., Ulmer, H., & Dapunt, O. (1998). Deutschsprachige Fassung und Validierung der» Edinburgh postnatal depression scale «. *DMW-Deutsche Medizinische Wochenschrift*, *123*(03), 35-40.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience, 10*. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00604

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, *14*(3), 036020. https://doi.org/10.1088/1741-2552/aa66dd

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British journal of clinical Psychology, 31*(3), 301-306.

3. Manuscript

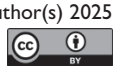
Reisner, S., Nguyen, T., Labendzki, P., Hoehl, S., & Markova, G. (2025). The reciprocal relationship between maternal infant-directed singing and infant gaze. *Musicae Scientiae, 0*(0). https://doi.org/10.1177/10298649251385676

Article

MUSICÆ SCIENTIÆ

The reciprocal relationship between maternal infant-directed singing and infant gaze

Musicae Scientiae
1–14



DOI: 10.1177/10298649251385676
journals.sagepub.com/home/msx
Sage

Susanne Reisner[1,2], Trinh Nguyen[1,3,4], Pierre Labendzki[5], Stefanie Hoehl[1] and Gabriela Markova[1,6]

Abstract

Infant-directed (ID) playsongs and lullabies have distinct acoustic properties connected to their functions to elicit and diffuse infant attention, respectively. In the performative context of ID singing, it is crucial that infants and caregivers adjust to each other for the songs' function to be reached. In this study, we observed face-to-face ID singing between mothers and their 7-month-old infants and measured variability in maternal singing (i.e., spectral flux) around the onset of infant social gaze toward the mother. Results showed that maternal acoustic variability and infant attention were increased in playsongs over lullabies. Furthermore, mothers increased their acoustic variability both before and after the onset of infant social gaze, especially in playsongs. These findings suggest that mothers increase acoustic variability both to modulate and respond to infant attention, and infants respond to more variable singing by paying more attention to the singing caregiver. Thus, we propose that ID singing interactions are reciprocal, linking infant attentional displays and maternal acoustic responses.

Keywords

infant-directed singing, spectral flux, acoustic variability, infant attention, infant–caregiver interaction, cross-modal reciprocity

[1]Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Austria
[2]Vienna Doctoral School Cognition, Behavior and Neuroscience, University of Vienna, Austria
[3]Neuroscience of Perception and Action Lab, Italian Institute of Technology, Italy
[4]Department of Developmental and Biological Psychology, Heidelberg University, Germany
[5]Department of Psychology, University of East London, UK
[6]Institute for Early Life Care, Paracelsus Medical University, Austria

Corresponding authors:
Susanne Reisner, Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, Vienna 1010, Austria.
Email: susanne.reisner@univie.ac.at

Gabriela Markova, Institute for Early Life Care, Paracelsus Medical University, Strubergasse 21, 5020 Salzburg, Austria.
Email: gabriela.markova@pmu.ac.at

Infant-directed (ID) singing occurs in cultures all over the world (Mehr et al., 2019; Mehr & Krasnow, 2017; Trehub, 2019; Trehub et al., 1993), and caregivers routinely engage in ID singing with their infants (Steinberg et al., 2021; Trehub et al., 1997). A growing number of studies have investigated the acoustic qualities of ID singing and infant reactions to it separately. Yet, it is likely that there is a reciprocal relationship between fine-grained acoustic changes in maternal singing and infant attention to achieve effective communication through different modalities. Here, we investigate the reciprocity between infant attention and maternal ID singing in a face-to-face interaction.

ID singing is especially attention-grabbing for infants (for a review, see Nguyen, Flaten, et al., 2023), even more so than ID or adult-directed speech (Fernald et al., 1989; Nakata & Trehub, 2004; Outters et al., 2020; Tsang et al., 2017). Infants look longer at mothers during ID singing than during speech (Nakata & Trehub, 2004), and they pay above-chance levels of attention to caregivers' eyes around the beat of ID singing (Lense et al., 2022). Likewise, caregivers make ID singing engaging by modifying their facial expressions and eye movements (Español & Shifres, 2015) to highlight salient information to the beat of their singing (Lense et al., 2022). Moreover, playful singing was found to be positively associated with dyadic gaze coordination during early face-to-face interactions (Markova et al., 2020). Thus, infant–caregiver musical interactions exhibit an intricate interplay of attentional displays and mutual adjustment through multiple modalities.

Caregivers sing to their infants in different caretaking contexts. Depending on their communicative intent, they choose between two broad categories of ID songs: Lullabies to soothe infants and divert attention, and playsongs to engage and direct infants' attention (Cirelli et al., 2020; Rock et al., 1999; Trehub, 2018; Trehub & Trainor, 1998). These functionalities are linked to the ID songs' acoustic qualities. Playsongs are usually sung faster, higher and more variable in pitch, and louder than lullabies (e.g., Cirelli et al., 2020; Rock et al., 1999; Trehub & Trainor, 1998).

For both song types, caregivers and infants need to reciprocally adjust to one another for the songs' communicative function to be reached. In early social interactions, caregivers and infants coordinate their actions to engage and influence each other's behaviour (Caulfield, 1995; Jaffe et al., 1973). Infants direct caregiver attention through a variety of vocal, affect, and gestural displays and, thus, play an active role in their social interactions from birth. Caregivers respond age-appropriately to infant displays (Caulfield, 1995). In the more performative context of ID singing, caregivers may vary their songs over time to achieve this adjustment (Nguyen, Reisner, et al., 2023). These variations can be captured by time-variant methods, such as spectral flux. Spectral flux measures both frequency and amplitude changes over time (Müller, 2015). Unlike methods focusing solely on amplitude changes, spectral flux thus captures spectro-temporal information in the audio stream (Weineck et al., 2022). This allows for detecting changes, such as pitch variations, that might be missed by examining the amplitude envelope alone. As a result, spectral flux provides a more accurate and holistic representation of musical events since frequency and amplitude modulations can occur independently (Zeng et al., 2005). Interestingly, a previous study found that caregivers adjust the level of spectral flux contained in ID speech to their preterm infant's wakefulness state (Saliba et al., 2020). These findings suggest that increased spectral flux may prompt infants to interact with caregivers, and in contrast, lower spectral flux may help avoid overstimulating infants and create a calmer environment.

In this study, we examined the reciprocity between live maternal ID singing and 7-month-old infants' attention. More specifically, we assessed the spectral flux of maternal ID singing of a playsong and a lullaby during a semi-naturalistic musical interaction around the onsets of

infant social gaze toward the mother. We aimed to test whether changes in ID singing drive changes in infant attention, changes in infant attention prompt changes in maternal singing, or changes in maternal singing and infant attention co-occur. Accordingly, we hypothesized that maternal singing would increase in variability (i.e., show increased spectral flux) around times of increased infant attention (i.e., around the onset of infant social gaze toward their mother), either as a tool to grab infant attention (if occurring before infant social gaze onset) or as a reaction to infant attention (if occurring after infant social gaze onset). We also hypothesized that this interplay of infant and caregiver adjustment is especially important in playsongs because of their proposed function to attract attention (Cirelli et al., 2020). Therefore, we expected playsongs to contain higher spectral flux than lullabies and to induce more instances of infant social gaze.

Methods

Participants

We included 74 7-month-old infants (33 girls; age: 227.54 ± 6.96 days [$M \pm SD$]) and their mothers (age: 34.58 ± 4.53 years) in the final sample. We decided on this age group because 6 and 7-month-old infants are especially attracted to ID singing (Tsang et al., 2017). We excluded 29 additional mother–infant dyads from the final analysis due to infant fussiness ($n = 12$), equipment failure ($n = 5$), no infant social gaze occurrence in both singing conditions ($n = 2$), maternal failure to follow instructions ($n = 5$), incomplete experiment ($n = 4$), and excessive background noise ($n = 1$).

We recruited participants from a database of families who had expressed interest in and consented to be contacted about partaking in developmental research. These families were recruited in neonatal units at local hospitals, in mother–child activity classes, and through social media. All infants were born full-term (gestational age of 36–42 weeks), had a birth weight of >2,500 g, and had no known developmental delays or neurological or hearing impairments. Infants grew up in predominantly German-speaking households. Mothers were highly educated, with 87.8% holding a university degree. Overall, 70.3% of mothers were first-time mothers and had no other children, 17.6% had one older child, one mother had a same-aged twin, one mother had three older children, and 9.3% provided no information on birth order. The primary caregiver was the mother in 68.9% of participants, both mother and father in 13.5% of participants, the mother and sibling in one participant, and 16.2% provided no information on the primary caregiver. 97.3% of mothers reported playing and singing at least one hour every week in the presence of their infant ($M = 11.7 \pm 11$ h listened to, $M = 6.7 \pm 5.9$ h sung). 56.8% of mothers reported playing an instrument, and 20.3% sang in a choir or a band. The study was conducted according to the Declaration of Helsinki, approved by the University of Vienna's ethics committee, and caregivers provided written informed consent.

Procedure

During the experiment, infants sat in a car seat or highchair while their mothers faced them, holding a tablet (Figure 1[a]). Infants were seated in a car seat or highchair, depending on electroencephalogram (EEG) or infant motion measurements detailed in another paper (Nguyen, Reisner, et al., 2023), or according to their preference. Each dyad was observed during two experimental singing blocks (Figure 1[b]), the order of which was randomized between

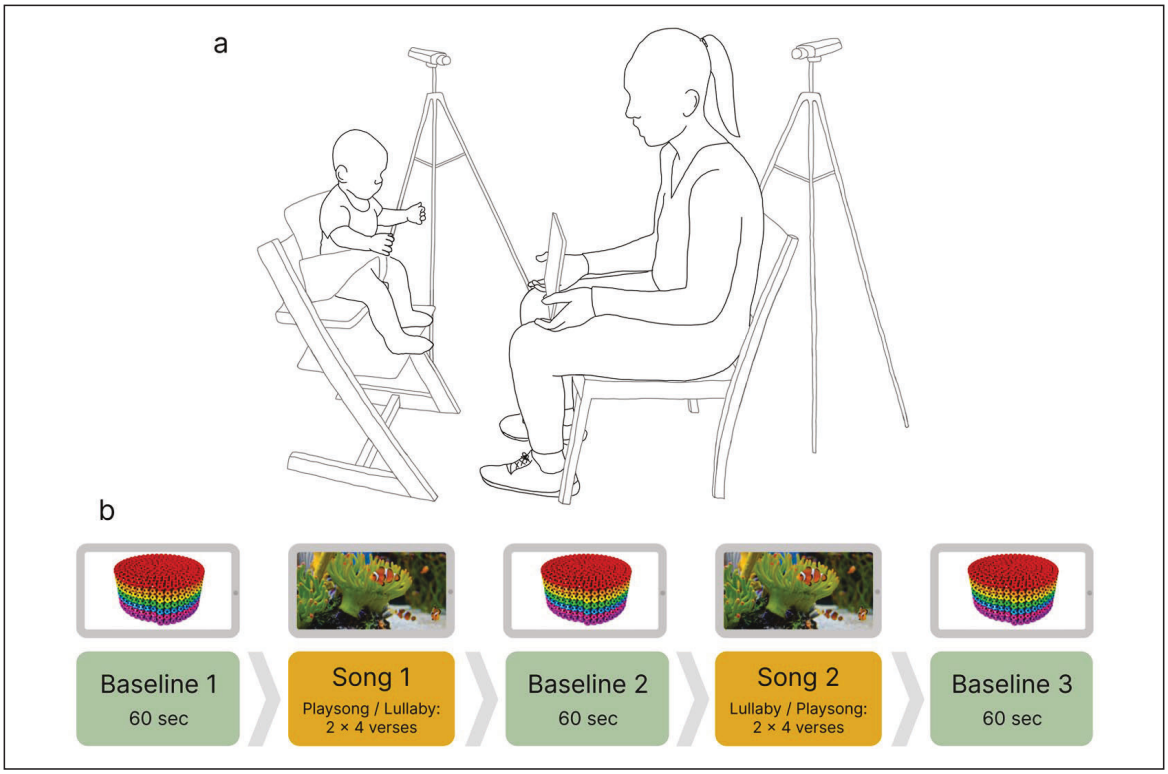


Figure 1. (a) Experimental setup illustration. The infant sat facing their mother in a highchair or a car seat held by an experimenter (not depicted). (b) Procedure. Mothers sang two repetitions of four verses for each lullaby and playsong. During both conditions, a tablet held by the mother played a video of fish swimming in an aquarium. A 60 s baseline, during which the mother did not interact with the infant, preceded and followed each song, during which the tablet showed moving geometric shapes without sound.

participants. A 60 s baseline preceded and followed each singing block (i.e., three baseline blocks in total), during which infants and mothers watched silent videos of slowly moving shapes. We included this break to give dyads some rest between the two singing conditions. Mothers were asked not to talk during baseline blocks but to reciprocate their infants' communicative attempts via facial expressions and gestures (e.g., smiling, pointing at the tablet). Per singing block, mothers were instructed to sing either a playsong or a lullaby. We selected well-known songs per category, "Schlaf Kindlein, schlaf" (lullaby) and "Es tanzt ein Bibabutzemann" (playsong) (see Figure 1S in the Supplemental material). While mothers were given song recordings to prepare before the experiment, most of them were familiar with the playsong (91%) and the lullaby (96%). On average, mothers sang both the playsong ($M = 2.48$, on a Likert scale from 1 = *never* to 5 = *very often*) and the lullaby ($M = 2.77$ on the Likert scale) moderately often to their infants at home. To ensure that mothers differentiated between the two songs, they were prompted with a metronome before the start of each singing block (i.e., playsong = 170 bpm; lullaby = 100 bpm). The metronome was stopped before the mothers began to sing. Mothers were asked to sing to their infants as they naturally would, and thus they did not have to follow the particular score for each song. All mothers sang two repetitions of four verses of each song. During the singing conditions, the tablet played a calm aquarium video to help with infant fussiness. Three cameras (synchronized in VideoSyncPro, Mangold International) recorded the experiment at 25 Hz. This experiment was part of a larger study, parts of which have already been published in a previous publication (Nguyen, Reisner, et al., 2023).

Measures

Maternal singing. Maternal singing was recorded using a microphone (Mangold International), which was placed on a table in the testing room approximately 1 m away from the mother and recorded at a 44,100 Hz sampling rate and pre-processed in Audacity. Excerpts containing audio clipping, infant vocalizations, vegetative noises, and other environmental noises were manually removed. We calculated the following acoustic characteristics of maternal singing per singing condition: spectral flux, amplitude, and pitch using a custom Python script (Python Software Foundation, 2023) and tempo using MIRtoolbox (Lartillot et al., 2008) in Matlab. We opted for tempo extraction via bpm in the MIRtoolbox over Python because of greater accuracy via manual checks. Amplitude was calculated via the amplitude envelope with the numpy package (Harris et al., 2020), pitch via F0 extraction with the parselmouth package (Jadoul et al., 2018). Spectral flux calculations are explained in the Spectrul flux section. We calculated mean spectral flux, amplitude, and pitch as a manipulation check. The main analysis comprised fine-grained acoustic changes in spectral flux relating to infant social gaze onset. As an additional check, we calculated the same fine-grained acoustic changes for amplitude (sound envelope) and pitch (fundamental frequency), which can be found in the Supplemental material (Sections 2.4.S and 2.5.S).

Infant gaze. We coded infant gaze as a proxy for infant attention. We distinguished between social gaze (i.e., infant gaze toward the mother's face) and non-social gaze (i.e., infant gaze oriented away from the mother's face or body). Only gaze behaviour with a duration of at least 1 s was coded. All videos were coded frame-by-frame and without sound by two principal coders who coded $N = 30$ and $N = 34$ videos, respectively. For assessing inter-rater reliability, two additional independent coders coded $N = 10$ and $N = 12$ videos, respectively, resulting in double coding of 30% of all videos. High inter-rater reliability of a mean of $\kappa = .94$ was obtained. Descriptive statistics of where infants looked (social vs non-social gaze locations) can be found in Table 2S in the Supplemental material.

Spectral flux. Using custom Python code (see data availability below), we extracted spectral flux time series from root-mean-square-normalized audio in maternal singing 5 s before and after infant social gaze onsets. Spectral flux was computed as the sum over frequency (from 0 to 3000 Hz) of the absolute difference in amplitude between consecutive spectra, and in that regard, it measures both overall amplitude and frequency changes over time (Müller, 2015).

$$Spectral\,Flux(t) = \sum_{f=0Hz}^{3000Hz} \left| X(t;f) - X(t+0.01;f) \right|$$

$X(t; f)$ is the spectral amplitude at time t and frequency f computed over a .03 s window with .02 s overlap, resulting in spectral flux with a temporal resolution of 10 ms. The spectrum was limited to frequencies between 0 and 3000 Hz, as that band contains most of the energy in speech and singing (Lindblom & Sundberg, 2007). We chose spectral flux as an index for acoustic saliency because (1) it measures acoustic variability over time, thus modelling acoustic novelty (Müller & Chiu, 2024), and (2) it has been found to regress better onto brain signals than simple amplitude modulations, as it also measures changes over the frequency domain (Weineck et al., 2022). Especially regarding attention, we think that spectral flux offers a more holistic approach when measuring complex, multi-layered, musical stimuli (Bianco et al., 2025). We

excluded instances of gaze onsets where more than 1 s were silent either before or after the gaze onset. This is a methodological consideration to exclude instances in which the spectral flux time series had to be silenced because of other noise sources (environmental sounds, infant vocalizations). Changes in spectral flux were time-locked around gaze onset, synced, and averaged per song type (lullaby or playsong). Surrogate data were simulated by generating 1,000 random onset points per participant and song, excluding time points within 5 s of a social look onset. In the permutation analyses, we compared 10 s windows around infant social gaze onset to surrogate 10 s windows from the same condition not containing infant social gaze. Similar analyses around the infant social gaze offsets can be found in the Supplemental material.

Statistical analysis

All statistical analysis was done in R Studio (Posit Team, 2024) and Python v3.11.5 (Python Software Foundation, 2023).

We conducted linear or generalized mixed-effects models (LME or GLME, respectively, depending on data distribution) to test for differences in social gaze frequency, absolute duration over the whole song, relative duration (i.e., infant social gaze duration relative to song duration), and individual social look length between the two singing conditions. We calculated the relative duration of infant social gaze because playsongs were significantly longer than lullabies, LME: $\chi^2(1) = 138.41$, $p < .001$, and square-root-transformed the relative duration of looks to aid model convergence. In the Supplemental material (Section 2.2.S), we repeated all analyses with added seat type as a fixed effect to test for differences between infants seated in a car seat vs a highchair.

$$\textit{LME: Social gaze frequency} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{GLME: Social gaze absolute duration} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{LME: Square-root-transformed social gaze relative duration} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{LME: Log10-transformed individual look length of social gaze} \sim \textit{song type} + (1 \mid ID)$$

We also conducted linear mixed-effects models to look for differences in song length and differences in mean spectral flux, amplitude, pitch, and tempo over the whole playsong and lullaby. In the Supplemental material (Section 2.2.S), we again added seat type as a fixed effect to compare infants seated in a car seat versus a highchair.

$$\textit{LME: Mean spectral flux} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{LME: Mean amplitude} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{LME: Mean pitch,} \sim \textit{song type} + (1 \mid ID)$$

$$\textit{LME: Mean tempo} \sim \textit{song type} + (1 \mid ID)$$

Next, we calculated Spearman correlations to test the correlation between infant social gaze frequency, absolute and relative duration, and mean spectral flux in playsongs and lullabies.

We conducted a permutation analysis with independent *t*-tests with custom Python code (see data availability below) to identify time points in the spectral flux 5 s before and after infant social gaze onsets (1,000 permutations, *p*-values below the 5th/2 percentile) in playsongs and lullabies. Similar analyses for amplitude (sound envelope) and pitch (fundamental frequency), as well as for social gaze offsets, can be found in the Supplemental material. For each permutation ($N = 1{,}000$), a continuous surrogate *p*-value was obtained by randomly re-shuffling the labels of the observed and surrogate look-related changes, and by then performing a series of *t*-tests for the spectral flux over the duration of the look-related changes, resulting in a time series of one truly observed *p*-value and 1,000 re-shuffled *p*-values. We then checked for each time point in the look-related changes whether the truly observed *p*-value was below the Bonferroni-corrected 5th percentile of the re-shuffled *p*-values. We Bonferroni-corrected the 5th percentile for two testing categories (playsong and lullaby). We report significance for time points where the truly observed *p*-value is below the Bonferroni-corrected re-shuffled 5th percentile.

Results

Gaze

We compared the frequency, absolute, and relative duration of infant social gaze between playsongs and lullabies to see if infant attention differed between the two conditions (see Table 1). Infant social gaze differed significantly between conditions, with infants showing significantly more instances, $\chi^2(1) = 12.80$, $p < .001$, longer absolute durations, $\chi^2(1) = 169.98$, $p < .001$, longer relative durations, $\chi^2(1) = 3.94$, $p = .04$, and longer individual looks, $\chi^2(1) = 11.27$, $p < .001$, of social gaze toward the mother during playsongs than during lullabies.

Maternal singing

We tested differences in the acoustic qualities of maternal singing over the whole song (see Table 1) to check for mean acoustic differences between the two singing conditions as a manipulation check. Overall, playsongs were significantly longer than lullabies, $\chi^2(1) = 138.41$, $p < .001$, and contained significantly higher mean spectral flux, $\chi^2(1) = 118.17$, $p < .001$, higher mean amplitude, $\chi^2(1) = 40.69$, $p < .001$, and higher tempo, $\chi^2(1) = 54.76$, $p < .001$, than lullabies. There was no significant difference between songs in mean pitch, $\chi^2(1) = .001$, $p = .981$.

Next, we examined the association between spectral flux and infant social gaze. Mean spectral flux positively correlated with absolute duration of infant social gaze ($S = 400{,}441$, $p = .033$, $\rho = .178$). There was a trend for mean spectral flux to positively correlate with the frequency of infant social gaze ($S = 412{,}579$, $p = .067$, $\rho = .153$), and with its relative duration ($S = 418{,}996$, $p = .095$, $\rho = .14$).

Spectral flux dynamics around infant social gaze

To detect fine time scale acoustic changes in maternal singing and examine whether maternal singing increases in spectral flux around times of increased infant attention, we tested

Table 1. Descriptive statistics on infant social gaze during playsongs and lullabies and acoustic qualities of maternal playsongs and lullabies. Absolute gaze frequency, relative and absolute gaze duration refer to means over the whole condition, duration of individual looks refers to the descriptive statistics of single looking events. Significant condition differences are highlighted in bold ($N = 74$).

	Playsong				Lullaby			
	M	*SD*	min	max	*M*	*SD*	min	max
	Infant social gaze							
Absolute frequency	**6.93**	4.00	1.00	17.00	**5.50**	3.25	1.00	17.00
Relative duration (%)	**15.85**	14.63	0.01	68.36	**12.71**	11.69	0.01	71.38
Absolute duration (s)	**27.55**	27.37	1.20	129.00	**17.27**	15.66	1.08	89.00
Duration of individual looks (s)	**4.00**	5.37	1.00	73.36	**3.21**	3.52	1.00	36.34
	Maternal singing							
Song duration (s)	**183.89**	25.11	134.00	238.00	**146.91**	24.94	96.00	207.00
Mean spectral flux (a.u.)	**395.97**	109.81	137.41	672.91	**205.09**	103.71	38.24	668.55
Mean amplitude (a.u.)	**138.84**	28.10	40.91	187.33	**108.61**	29.54	26.34	231.51
Mean pitch (Hz)	249.86	35.44	154.34	337.02	249.77	39.60	153.88	430.14
Mean tempo (bpm)	**162.82**	22.03	109.95	199.65	**136.13**	25.38	95.98	199.63

differences between original and surrogate maternal spectral flux 5 s before and after infant social gaze onset.

In playsongs, spectral flux was significantly higher than chance level ($p <$ 5th/2 percentile) at time points from T −5.0 to T −2.0 s before, from T −1.0 s before to T + 2.0 s after, and from T + 3.0 to T + 5.0 s after infant social gaze onset (see Figure 2[a] and Table 4S in the Supplemental material). In lullabies, spectral flux was significantly lower than chance level ($p <$ 5th/2 percentile) at T −1.5 and T −1.0 s before infant social gaze onset, and significantly above-chance level ($p <$ 5th/2 percentile) from T −3.5 to T −2.5 and T −0.5 s before, and at T + 0.5, T + 1.0, T + 2.0, and T + 3.0 s after infant social gaze onset (see Figure 2[b] and Table 4S in the Supplemental material). Similar analyses regarding amplitude (sound envelope) and pitch (fundamental frequency) can be found in the Supplemental material.

Discussion

This study investigated how mothers and their 7-month-old infants adjusted to each other while the mothers sang a playsong and a lullaby. We aimed to determine whether variability in maternal singing (i.e., spectral flux) changes to modulate or respond to infant attention (i.e., social gaze). Our study revealed two main results. First, infants showed more gaze toward their mothers during playsongs than during lullabies; correspondingly, playsongs contained higher spectral flux than lullabies. Second, mothers showed changes in spectral flux before and after the onsets of infant social gaze in both types of songs. Interestingly, mothers' peaks in spectral flux as a response to infant attention were especially pronounced in playsongs. These results suggest a reciprocal relationship between mother and infant behaviour in a singing interaction.

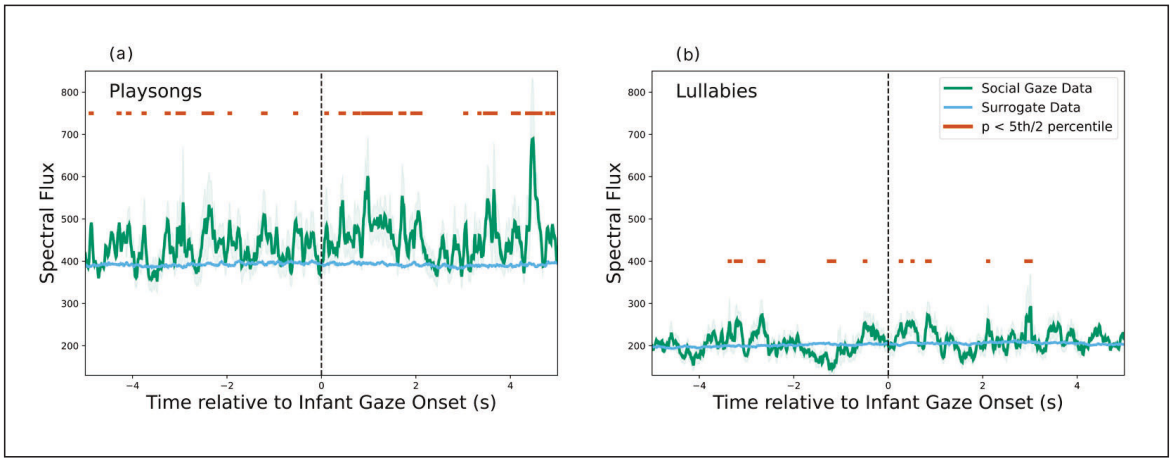


Figure 2. Mean spectral flux playsongs (a) and lullabies (b) 5 s before and after infant social gaze onset (dashed line), depicting instances of infant social gaze toward the mother (green) vs surrogate time points, which excluded time points within 5 s of an infant social gaze onset (blue). Horizontal orange lines indicate time points where the spectral flux around infant social gaze was significantly above or below the spectral flux of non-social surrogate looks.

Our finding that maternal spectral flux increased before the onset of infant social gaze, especially during playsongs, corroborates previous research showing that caregivers adjust their ID speech by producing rising pitch contours to gain infant eye contact (Fernald et al., 1989; Stern et al., 1982). Thus, mothers may have tried to gain their infants' attention by increasing the variability in their singing during playsongs more so than during lullabies. This interpretation is further supported by our findings that infants looked longer and more frequently toward their mothers during playsongs than lullabies. Moreover, higher spectral flux across conditions correlated with higher absolute looking duration and marginally correlated with higher relative looking duration and frequency. Indeed, playsongs were found to have higher spectral flux than lullabies, as indicated by more acoustic variability and changes over time. This increased acoustic variability induces auditory uncertainty, which in turn might attract infant attention (Kidd et al., 2014; Poli et al., 2020). Therefore, the more variable playsong was arguably more captivating for the infants, both gaining and maintaining their attention. On the other hand, lullabies could have been acoustically more predictable to fulfil their soothing function (Nguyen, Reisner, et al., 2023; Schwartz, 2004) and thus did not elicit as much attention from the infants. This interpretation is further supported by previous research finding that infants focus their attention more inward during lullabies than during playsongs (Rock et al., 1999). Perhaps even below-baseline arousal responses could further play into this pattern (Cirelli et al., 2020).

Maternal acoustic responses after the onset of infant attention also depended on song type. In playsongs, mothers reciprocated infant attention by increasing spectral flux almost continuously from around 0 to 2 s and 3 to 5 s after infant social gaze onset, possibly to maintain infant attention. Presumably, infant social attention motivated mothers to keep infants engaged with an entertaining and captivating performance, in line with the presumed function of playsongs (Cirelli et al., 2020). In line with these results, previous research findings indicate that caregivers variably adjust their pitch contours in ID speech to maintain infant attention (Stern et al., 1982) and show more frequent and adaptive ID speech depending on whether they can observe their infant's reciprocal response (Braarud & Stormark, 2008; Lam & Kitamura, 2012; Nencheva & Lew-Williams, 2022; Snow, 1972). In our study, we also found more reciprocal responses in the playsong condition. However, ID singing might be more constrained regarding variable pitch changes than ID speech is

(Menn et al., 2022). This constraint on pitch is shown in our supplementary analyses. Amplitude might have been the greater driving factor for variable modulation. In lullabies, on the other hand, mothers did not alter their variability as much as during playsongs after infant social gaze onset, showing very few peaks in spectral flux and even dips (i.e., less variability than in the surrogate data). Mothers might have performed lullabies with lower variability to conform to the song's intended soothing function. This pattern of results is aligned with a study examining sleeping newborns, showing that caregivers reduce their spectral flux during ID speech to reduce stimulation and increase spectral flux to animate and interact with their newborn infants (Saliba et al., 2020). Further studies could probe into this functionality of ID singing in a more systematic way.

Findings from Bainbridge et al. (2020) suggest that infants show functional differences in their responses even to pre-recorded songs, indicating that reciprocity may not be strictly necessary for some aspects of song function. However, social interaction may serve as an additional driving force in shaping these responses. Specifically, live interactions introduce a feedback loop in which mothers dynamically adapt their singing based on infant cues, amplifying the salience of these responses through social contingency. Prior research on audience effects has demonstrated that live interactions often elicit stronger and more engaged responses compared to recorded stimuli, emphasizing the importance of social presence in communication (Kragness et al., 2023; Rochat, 2001). In this framework, infants are not merely passive recipients of maternal singing but active participants in a reciprocal exchange (Beebe et al., 2016). Furthermore, the extent to which reciprocity influences song function may depend on the type of song—different song types could elicit distinct patterns of adaptation from both mother and child. Thus, while reciprocity may not be an absolute requirement for song function to be realized, it likely enhances its communicative and regulatory effects, reinforcing the interactive nature of caregiver–infant singing.

Limitations

This study was not without limitations. First, the setup with the tablet was not ideal. While it made it easier to keep infants calm, it could have been distracting to the interaction, and mothers might have had to compete with the tablet for their infants' attention. However, infants were also looking at the surrounding environment for about a third of the song duration (see Table 2S in the Supplemental material). The tablet was a methodological consideration to reduce infant motion, as we collected infant EEG data in a subset of infants, the results of which can be found in a previous publication (Nguyen, Reisner, et al., 2023). To keep the study design consistent, we opted to keep the tablet in all test sessions. For future studies, we would discourage the use of a tablet if compatible with infant compliance.

Second, technical limitations did not allow us to delve into acoustic variability around beats. We needed a sliding window for our analysis, which was not possible with the MIR toolbox beat extraction, which we chose over Python for greater accuracy. Recent work has shown that the beats of ID singing play an important role both in caregiver behavior (via highlighting salient information) and in infant attentional responses (Lense et al., 2022). Investigating maternal acoustic variability around beats would be an interesting future direction, but we advise to manually code the beats of live singing to ensure maximum accuracy.

Third, this article does not explore the multimodality of maternal signals, such as gestures, facial expressions, or touch. Previous research has shown the importance of eye movements and facial expressions in ID singing (Español & Shifres, 2015; Lense et al., 2022). Future research could expand our study on the interactive nature of ID singing by combining acoustic measures with infant and caregiver behavioral measures to get more insight into the multimodality and possible reciprocity of ID singing.

Fourth, we did not explore trial effects due to not having enough looks per verse. This could be an interesting future direction to uncover where in the song infants paid the most attention.

Furthermore, while mothers sang live to their infants in this study, they were still asked to sing on cue. Especially the lullaby might not have fit the mood of awake infants. Even though there are still significant differences in both maternal acoustic qualities and infant attention, future studies could prioritize even more naturalistic settings, for example, by letting caregivers choose songs that fit their infant's mood. Recruiting a more diverse sample of caregivers and infants would further enhance the generalizability of results.

Conclusion

In social interactions, infants are not merely passive observers but are, in fact, actively shaping their interactions (Begus & Southgate, 2018), especially when caregivers reciprocate their social bids (Phillips et al., 2023). This study is the first to show this reciprocity between infants and caregivers in a singing context. Infants not only responded to maternal singing by paying more attention to their variable singing, but mothers also adjusted acoustically to their infants' social gaze by changing the variability of their singing. This was particularly the case during playsongs. These findings substantiate the distinct functions of ID songs—exciting and attention-grabbing playsongs and calming lullabies. Cross-modal reciprocity between caregivers' and infants' social behaviours supports engaging infants and might amplify song function, thus contributing to infant communicative development.

Acknowledgments

The authors are grateful to Liesbeth Forsthuber, Christina Schaetz, Martina de Eccher, Sarah Paul, Marlies Schermann, Anna Matyk, Rahel Loos, Jakob Weickmann, Elisa Roberti, Marlies Jarosch, Anja Lueger, Anna Matyk, Tina Duric, Carole Lambert, Natalie Geißler, and Eluisa Nimpf for their support in data acquisition, processing, and video coding. Additionally, the authors thank all families who participated in the study and the Department of Obstetrics and Gynecology of the Vienna General Hospital for supporting our participant recruitment.

ORCID iDs

Susanne Reisner https://orcid.org/0000-0003-1455-9297
Trinh Nguyen https://orcid.org/0000-0003-0420-2147
Pierre Labendzki https://orcid.org/0000-0002-8688-6935
Stefanie Hoehl https://orcid.org/0000-0003-0472-0374
Gabriela Markova https://orcid.org/0000-0003-1092-0137

Author contributions

Susanne Reisner: Conceptualization, Methodology, Formal analysis, Data curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Project administration **Trinh Nguyen:** Conceptualization, Methodology, Software, Resources, Writing—Original Draft, Writing—Review & Editing **Pierre Labendzki:** Methodology, Software, Writing—Original Draft, Writing—Review & Editing, Visualization **Stefanie Hoehl:** Conceptualization, Methodology, Resources, Writing—Review & Editing, Supervision, Funding acquisition **Gabriela Markova:** Conceptualization, Methodology, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing, Supervision, Project administration.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by an Austrian Science Fund (FWF) "DK Cognition and Communication 2": W1262-B29 [10.55776/W1262]. This research was also funded by the University of Vienna. T.N. acknowledges support by Horizon Europe (MSCA, SYNCON, 101105726). For open access purposes, the author has applied a CC BY public copyright licence to any author-accepted manuscript version arising from this submission.

Conflict of Interest

The authors declare no conflict of interest.

Data availability statement

Data for this publication can be found at: https://osf.io/dzvhb/overview.

Code availability

The code used in this study is available to the public under a Creative Commons licence at: https://github.com/susannereisner/SINGGaze_GIT_repository

Supplemental material

Supplemental material for this article is available online.

References

Bainbridge, C. M., Bertolo, M., Youngers, J., Atwood, S., Yurdum, L., Simson, J., Lopez, K., Xing, F., Martin, A., & Mehr, S. A. (2020). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behaviour*, *5*(2), 256–264. https://doi.org/10.1038/s41562-020-00963-z

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother–infant face-to-face communication. *Developmental Psychology*, *52*(4), 556–571. https://doi.org/10.1037/a0040085

Begus, K., & Southgate, V. (2018). Curious learners: How infants' motivation to learn shapes and is shaped by infants' interactions with the social world. In M. M. Saylor & P. A. Ganea (Eds.), *Active learning from infancy to childhood: Social motivation, cognition, and linguistic mechanisms* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-77182-3_2

Bianco, R., Tóth, B., Bigand, F., Nguyen, T., Sziller, I., Háden, G. P., Winkler, I., & Novembre, G. (2025). Human newborns form musical predictions based on rhythmic but not melodic structure [Preprint]. *bioRxiv*. https://doi.org/10.1101/2025.02.19.639016

Braarud, H. C., & Stormark, K. M. (2008). Prosodic modification and vocal adjustments in mothers' speech during face-to-face interaction with their two- to four-month-old infants: A double video study. *Social Development*, *17*(4), 1074–1084. https://doi.org/10.1111/j.1467-9507.2007.00455.x

Caulfield, R. (1995). Reciprocity between infants and caregivers during the first year of life. *Early Childhood Education Journal*, *23*(1), 3–8. https://doi.org/10.1007/BF02353372

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of maternal singing style on mother–infant arousal and behavior. *Journal of Cognitive Neuroscience*, *32*(7), 1213–1220. https://doi.org/10.1162/jocn_a_01402

Español, S., & Shifres, F. (2015). The artistic infant directed performance: A microanalysis of the adult's movements and sounds. *Integrative Psychological and Behavioral Science*, *49*(3), 371–397. https://doi.org/10.1007/s12124-015-9308-4

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Jadoul, Y., Thompson, B., & De Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Jaffe, J., Stern, D. N., & Peery, J. C. (1973). 'Conversational' coupling of gaze behavior in prelinguistic human development. *Journal of Psycholinguistic Research*, *2*(4), 321–329. https://doi.org/10.1007/BF01067054

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kragness, H. E., Eitel, M. J., Anantharajan, F., Gaudette-Leblanc, A., Berezowska, B., & Cirelli, L. K. (2023). An itsy bitsy audience: Live performance facilitates infants' attention and heart rate synchronization. *Psychology of Aesthetics, Creativity, and the Arts*, *19*(5), 1124–1133. https://doi.org/10.1037/aca0000597

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, *15*(2), 212–221. https://doi.org/10.1111/j.1467-7687.2011.01118.x

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A Matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer. https://doi.org/10.1007/978-3-540-78246-9_31

Lense, M. D., Shultz, S., Astésano, C., & Jones, W. (2022). Music of infant-directed singing entrains infants' social visual behavior. *Proceedings of the National Academy of Sciences*, *119*(45), e2116967119. https://doi.org/10.1073/pnas.2116967119

Lindblom, B., & Sundberg, J. (2007). The human voice in speech and singing. In *Springer Handbook of Acoustics* (pp. 669–712). https://doi.org/10.1007/978-0-387-30425-0_16

Markova, G., Nguyen, T., Schätz, C., & de Eccher, M. (2020). Singing in tune – being in tune: Relationship between maternal playful singing and interpersonal synchrony. *Enfance, N°*, *1*(1), 89–107. https://doi.org/10.3917/enf2.201.0089

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684. https://doi.org/10.1016/j.evolhumbehav.2016.12.005

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Müller, M. (2015). *Fundamentals of music processing*. Springer eBooks. https://doi.org/10.1007/978-3-319-21945-5

Müller, M., & Chiu, C.-Y. (2024). A basic tutorial on novelty and activation functions for music signal processing. *Transactions of the International Society for Music Information Retrieval*, *7*(1), 179–194. https://doi.org/10.5334/tismir.202

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455–464. https://doi.org/10.1016/j.infbeh.2004.03.002

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, 66, 101047. https://doi.org/10.1016/j.dr.2022.101047

Nguyen, T., Flaten, E., Trainor, L., & Novembre, G. (2023). Early social communication through music: State of the art and future perspectives [Preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/j5g69

Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., & Markova, G. (2023). Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Developmental Cognitive Neuroscience*, *64*, 101313. https://doi.org/10.1016/j.dcn.2023.101313

Outters, V., Schreiner, M. S., Behne, T., & Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy*, *25*(4), 478–499. https://doi.org/10.1111/infa.12334

Phillips, E. A. M., Goupil, L., Whitehorn, M., Bruce-Gardyne, E., Csolsim, F. A., Marriott-Haresign, I., & Wass, S. V. (2023). Proactive or reactive? Neural oscillatory insight into the leader–follower dynamics of early infant–caregiver interaction. *Proceedings of the National Academy of Sciences*, *120*(15), e2122481120. https://doi.org/10.1073/pnas.2122481120

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Posit Team. (2024). *RStudio: Integrated development environment for R* (Version 2023.12.1.402)[Computer software]. Posit Software, PBC. https://www.posit.co/

Python Software Foundation. (2023). *Python* (Version 3.11.5)[Computer software].

Rochat, P. R. (2001). Social contingency detection and infant development. *Bulletin of the Menninger Clinic*, *65*(3), 347–360. https://doi.org/10.1521/bumc.65.3.347.19847

Rock, A. M. L., Trainor, L. J., & Addison, T. L. (1999). Distinctive messages in infant-directed lullabies and play songs. *Developmental Psychology*, *35*(2), 527–534. https://doi.org/10.1037/0012-1649.35.2.527

Saliba, S., Esseily, R., Filippa, M., Gratier, M., & Grandjean, D. (2020). Changes in the vocal qualities of mothers and fathers are related to preterm infant's behavioural states. *Acta Paediatrica*, *109*(11), 2271–2277. https://doi.org/10.1111/apa.15238

Schwartz, F. J. (2004). Medical music therapy for the premature baby – Research review. In M. Nöcker-Ribaupierre (Ed.), *Music therapy for premature and newborn infants* (pp. 85–96). Barcelona Publishers.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565. https://doi.org/10.2307/1127555

Steinberg, S., Liu, T., & Lense, M. D. (2021). Musical engagement and parent-child attachment in families with young children during the COVID-19 pandemic. *Frontiers in Psychology*, *12*, Article 641733. https://doi.org/10.3389/fpsyg.2021.641733

Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, *18*(5), 727–735. https://doi.org/10.1037/0012-1649.18.5.727

Trehub, S. E. (2018). Music lessons from infants. In *Oxford handbook of music psychology*. Oxford University Press.

Trehub, S. E. (2019). Nurturing infants with music. *International Journal of Music in Early Childhood*, *14*(1), 9–15. https://doi.org/10.1386/ijmec.14.1.9_1

Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43–77.

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental Psychology*, *33*(3), 500–507. https://doi.org/10.1037/0012-1649.33.3.500

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behavior and Development*, *16*(3), 285–295. https://doi.org/10.1016/0163-6383(93)80036-8

Tsang, C. D., Falk, S., & Hessel, A. (2017). Infants prefer infant-directed song over speech. *Child Development*, *88*(4), 1207–1215. https://doi.org/10.1111/cdev.12647

Weineck, K., Wen, O. X., & Henry, M. J. (2022). Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *eLife*, *11*, e75515. https://doi.org/10.7554/eLife.75515

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences*, *102*(7), 2293–2298. https://doi.org/10.1073/pnas.0406460102

SUPPLEMENTARY MATERIAL

1. S. Methods

1.1. S. Infant-directed songs

We selected the German lullaby “Schlaf, Kindlein, schlaf” and the playsong “Es tanzt ein Bibabutzemann”. Each verse was repeated eight times with lyrical variations, resulting in average audio lengths of 148.19 s for the lullaby (range = 96-207 s) and 183.92 s for the playsong (range = 134-238 s).

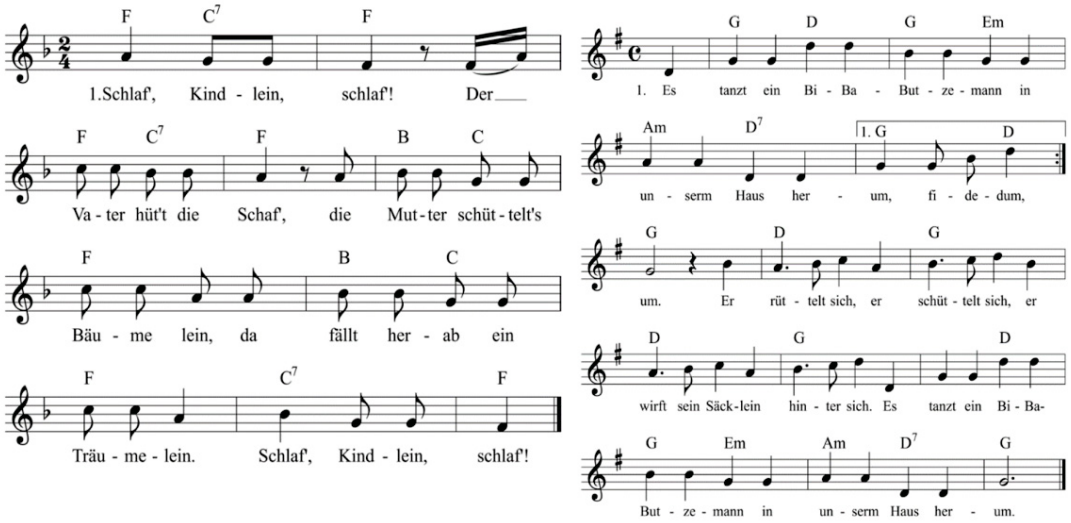


Figure 1S. Notations and lyrics of the lullaby (“Schlaf, Kindlein, schlaf”) and the playsong (“Es tanzt ein Bibabutzemann”). Note that mothers were not asked to follow the notations exactly.

1.2. S. Audio Preparation

We manually removed infant vocalisations, vegetative noises (e.g., coughs and burps), background noise (i.e., infants hitting against the highchair, metronome ticks), and audio clipping from the audio files using the software Audacity (https://www.audacityteam.org/). To remove audio clipping, i.e., distorted audio due to audio input of too high amplitude, we first detected audio signals of intensities above the 0.95 percentile with a custom Matlab script. We then manually checked these sections for distorted waveforms, which we silenced. For descriptive statistics on the silenced sections, see Table 1S. Two independent observers coded manually removed audio excerpts in 30% of

all audios on whether they contained infant vocalisations, yielding high inter-rater reliability with $\kappa = 0.93$. The removed contaminations did not differ significantly between the lullaby and playsong conditions, both in frequency ($V = 1055$, $p = .147$) and relative duration ($V = 1323$, $p = .470$).

Table 1S. Descriptive statistics on silenced sections in maternal playsongs and lullabies ($N = 74$).

	Silenced Sections							
	Playsong				Lullaby			
	M	SD	min	max	M	SD	min	max
Frequency	20.12	17.16	0	89	17.3	14.55	0	65
Relative Duration (%)	9.1	9.1	0	38.1	8.7	8	0	35
Absolute Duration (s)	16.46	16.57	0	71.6	12.89	12.59	0	60.07

1.3. S. Spectral Flux

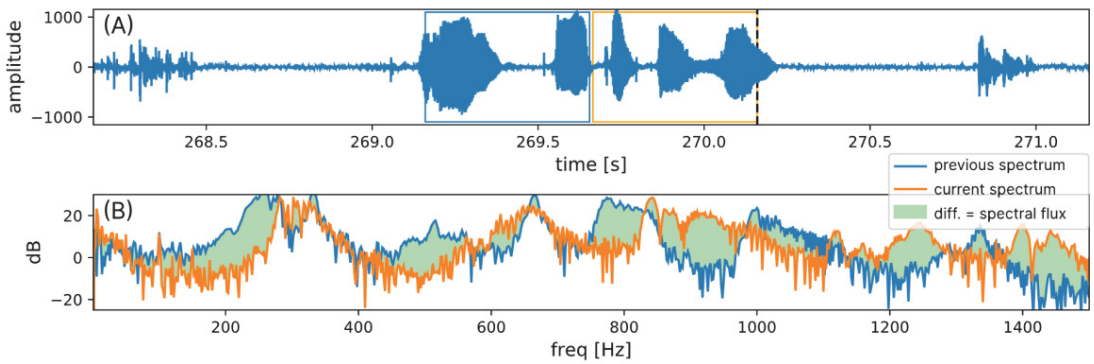


*Figure 2*S. Illustration of spectral flux. (A) shows the acoustic waveform over time, with the orange and blue windows indicating respectively the current time window and the previous time window. (B) Current and previous spectra are computed during the time window from (A) using an FFT. Spectral flux is visualised as the green area between the two curves and is formally computed as the frequency-wise difference between consecutive spectra.

2. S. Results

2.1. S. Infant Gaze

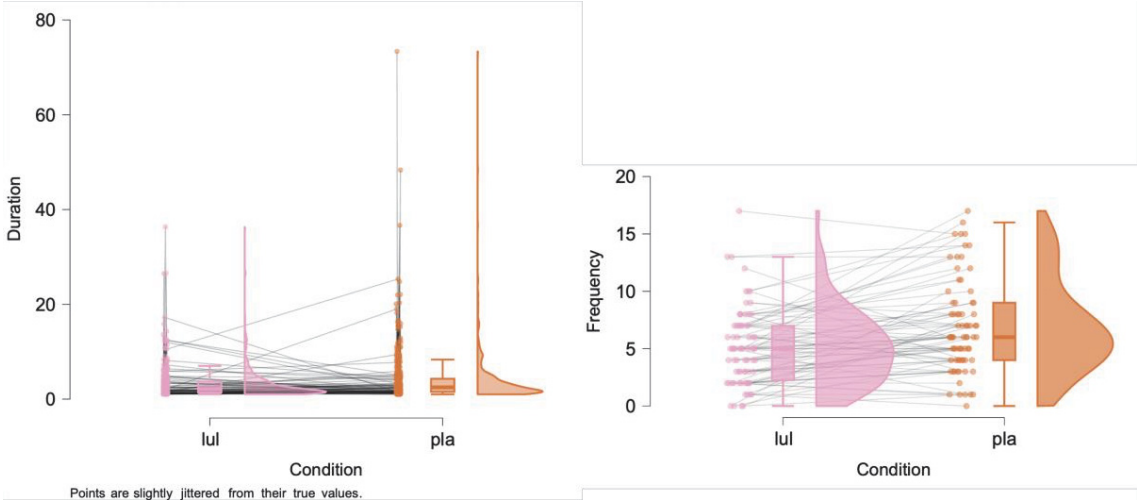


Figure 3S. Infant social gaze durations and frequency in the two singing conditions ($N = 74$).

Table 2S. Descriptive statistics on infant proportional looking duration. Looking behaviour was distinguished between social gaze (towards the mother's face), non-social gaze (towards the mother's body, the tablet, away from the mother's face or body), and not codable ($N = 74$).

	Infant Gaze							
	Playsong				Lullaby			
	M	SD	min	max	M	SD	min	max
Mother's face (%)	14.9	14.4	0.0	69.1	11.7	12.4	0.0	70.4
Mother's body (%)	0.1	0.5	0.0	3.4	0.1	0.3	0.5	2.2
Tablet (%)	56.2	20.8	12.3	97.4	59.9	19.9	16.2	97.9
Away (%)	28.1	16.9	0.0	64.4	27.5	18.3	0.0	70.1
Not Codable (%)	0.2	1.3	0.0	10.8	0.3	1.9	0.0	16.0

Latencies between individual occurrences of infant social looks were variable (see Table 3S). Over both conditions, 187 looks occurred within 5 seconds before the next look onset, and 550 looks occurred after 5 seconds before the next infant social gaze onset.

Table 3S. Descriptive statistics on latencies (in seconds) between individual infant social looks ($N = 74$).

	Infant Gaze							
	Playsong				Lullaby			
	M	SD	min	max	M	SD	min	max
Latency (s)	20.48	50.31	0.76	700.96	21.71	36.01	1.1	455.72

2.2. S. Effects of Seat Type

To test the effect of seat type (car seat vs highchair), we added seat type as a fixed effect to our models:

LME: Social gaze frequency ~ song type + seat type + (1|ID)

GLME: Social gaze absolute duration ~ song type + seat type + (1|ID)

LME: Square-root-transformed social gaze relative duration ~ song type + seat type + (1|ID)

LME: Log10-transformed individual look length of social gaze ~ song type + seat type + (1|ID)

LME: Mean spectral flux ~ song type + seat type + (1|ID)

LME: Mean amplitude ~ song type + seat type + (1|ID)

LME: Mean pitch, ~ song type + seat type + (1|ID)

LME: Mean tempo ~ song type + seat type + (1|ID)

Seat type (car seat vs highchair) did not have a significant effect on infant gaze frequency ($\chi^2(1) = .743$, $p = .389$), absolute duration (χ^2 (1) = .691, $p = .406$), relative duration (χ^2 (1) = 2.63, $p = .105$), or duration of individual social looks (χ^2 (1) = 2.736 , p = .1). Seat type also did not have a significant effect on song length (χ^2 (1) = . 443, $p = .506$), mean spectral flux (χ^2 (1) = .677, $p = .411$), mean amplitude (χ^2 (1) = .022, $p = .882$), mean tempo (χ^2 (1) = 1.242, $p = .265$), or mean pitch (χ^2 (1) = .15, $p = .7$).

79

80 **2.3. S. Permutation Analyses - Spectral Flux**

81 **2.3.1. S. Spectral Flux Dynamics Around Infant Social Gaze Onset**

82 *Table 4S*. Exact time points of above-chance changes in spectral flux in playsongs and
83 lullabies from five seconds before (T -5.0) to five seconds after (T +5.0) the onset of infant
84 social gaze, chunked into half-second intervals. Time points where the observed spectral flux
85 was significantly below the mean surrogate spectral flux are marked with an asterisk (*). The
86 total length (in seconds) of above- or below-threshold changes of every half-second interval
87 is noted in bold and in curly brackets ($N = 74$).

Time (s)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -5.0	[-4.88 – -4.85] **{0.04}**	
T -4.5	[-4.29 – -4.26] **{0.09}**	
T -4.0	[-4.09 – -4.05], [-3.76 – -3.75] **{0.02}**	
T -3.5	[-3.74 – -3.73] [-3.27 – -3.25] **{0.05}**	[-3.35 – -3.34] **{0.02}**
T -3.0	[-3.24 – -3.22], [-3.04 – -3.03], [-2.99 – -2.98], [-2.94 – -2.90] **{0.12}**	[-3.22 – -3.19], [-3.17 – -3.11] **{0.11}**
T -2.5	[-2.49 – -2.44],[-2.42 – -2.3], [-2.26] **{0.20}**	[-2.72 – -2.63] **{0.10}**
T -2.0	[-1.94 – -1.92] **{0.03}**	
T -1.5		[-1.25]* **{0.01}***
T -1.0	[-1-23 – -1.17], [-0.94] **{0.08}**	[-1.24 – -1.20]* , [-1.18 – -1.13]* **{0.11}***
T -0.5	[-0.55 – 0.52], [-0.42] **{0.05}**	[-0.49 – -0.47] **{0.03}**

T 0	[0.11 – 0.13] **{0.03}**	
T +0.5	[0.41 – 0.49], [0.71 – 0.73] **{0.12}**	[0.27 – 0.29], [0.42], [0.52 – 0.53] **{0.06}**
T +1.0	[0.75 – 0.8], [0.89 – 1.06], [1.08 – 1.24] **{0.41}**	[0.83 – 0.89] **{0.07}**
T +1.5	[1.25 – 1.29], [1.31 – 1.48], [1.68 – 1.74] **{0.30}**	
T +2.0	[1.75 – 1.77], [1.91], [1.93 – 1.94], [1.99 – 2.11] **{0.19}**	[2.12 – 2.13] **{0.02}**
T +3.0	[2.83], [3.05 – 3.08] **{0.05}**	[2.93 –3.03] **{0.11}**
T +3.5	[3.35 – 3.36], [3.47 – 3.6], [3.64 – 3.70], [3.73] **{0.24}**	
T +4.0	[3.76], [4.06 – 4.13], [4.17 – 4.19] **{0.12}**	
T +4.5	[4.36 – 4.39], [4.41 – 4.65] **{0.29}**	
T +5.0	[4.79 – 4.80], [4.89 – 4.92] **{0.06}**	

2.3.2. S. Spectral Flux Dynamics Around Infant Social Gaze Offset

In playsongs, spectral flux was significantly higher than chance level (p <5th/2 percentile), at T -5.0 seconds before and from T -4.0 seconds before to T +5.0 seconds after infant social gaze offset, with only very short periods of above-chance level spectral flux form T 0 to T + 2.0 seconds after infant social gaze offset (see Fig. 4SA and Tab. 5S). In lullabies, only very short periods of significantly higher or lower than chance level (p <5th/2 percentile) spectral flux occurred. Spectral flux was lower than chance level at T -5.0 and T -3.0 seconds before infant social gaze offset, and higher than chance level at T -4.0 and T -2.5

seconds before, and at T +1.0, T +2.5, and T +4.0 seconds after infant social gaze offset (see Fig. 4SB and Tab. 5SB).

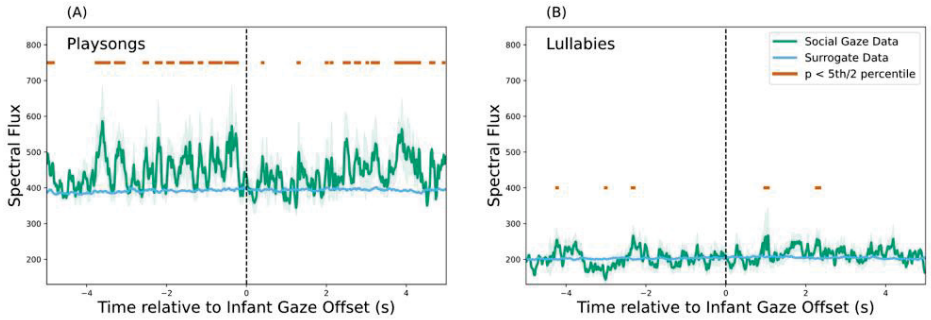


Figure 4S. Mean spectral flux playsongs (A) and lullabies (B) 5 s before and after infant social gaze offset (dashed line), depicting instances of infant social gaze towards the mother (green) vs surrogate time points which excluded time points within 5 s of an infant social gaze offset (blue). Horizontal orange lines indicate time points where the spectral flux around infant social gaze was significantly above or below the spectral flux of non-social surrogate looks ($N = 74$).

Table 5S. Exact time points of above-chance changes in spectral flux in playsongs and lullabies from five seconds before (T -5.0) to five seconds after (T +5.0) the offset of infant social gaze, chunked into half-second intervals. Time points where the observed spectral flux was significantly below the mean surrogate spectral flux are marked with an asterisk (*). The total length (in seconds) of above- or below-threshold changes of every half-second interval is noted in bold and in curly brackets ($N = 74$).

Time (s)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -5.0	[-4.99 – -4.93], [-4.85 – -4.83] **{0.10}**	[-4.79]* **{0.01}***
T -4.0	[-3.84], [-3.75] **{0.02}**	[-4.22 – -4.21], [-3.8] **{0.03}**
T -3.5	[-3.74 – -3.47], [-3.44 – -3.43], [-3.27 – -3.25] **{0.33}**	

T -3.0	[-3.24 – -3.06] **{0.19}**	[-3.0 – -2.99]* **{0.02}***
T -2.5	[-2.55 – -2.47] **{0.09}**	[-2.33 – -2.29] **{0.05}**
T -2.0	[-2.24 – -2.13], [-1.96 – -1.88], [-1.85 – 1.83] **{0.24}**	
T -1.5	[-1.63 – -1.49], [-1.47 – -1.39], [-1.36 – -1.35] **{0.26}**	
T -1.0	[-1.17 – -1.11], [-1.03], [-0.9 – -0.84], [-0.82 – -0.79] **{0.19}**	
T -0.5	[-0.74 – -0.63], [-0.5 – -0.39], [-0.37 – -0.25] **{0.37}**	
T 0	[-0.24 – -0.22] **{0.03}**	
T +0.5	[0.41 – 0.42], [0.52] **{0.03}**	
T +1.0	[0.79] **{0.01}**	[0.99 – 1.01], [1.05 – 1.07] **{0.06}**
T +1.5	[1.3 – 1.33], [1.38] **{0.05}**	
T +2.0	[2.01 – 2.03], [2.05], [2.13 – 2.15] **{0.07}**	
T +2.5	[2.46 – 2.58], [2.74] **{0.14}**	[2.28 – 2.29], [2.32 – 2.33], [2.35 – 2.36] **{0.06}**
T +3.0	[2.75 – 2.81], [2.83 – 2.84], [3.03 – 3.04], [3.16 – 3.24] **{0.20}**	
T +3.5	[3.25 – 3.31], [3.59] **{0.08}**	

T +4.0	[3.75 – 4.21] **{0.47}**	[4.20] **{0.01}**
T +4.5	[4.25 – 4.27], [4.32 – 4.34], [4.62 – 4.66], [4.69 – 4.70] **{0.13}**	
T +5.0	[4.94 – 4.95], [4.99] **{0.03}**	

2.4. S. Permutation analyses - Amplitude (Sound Envelope)

2.4.1. S. Amplitude Dynamics Around Infant Social Gaze Onset

In playsongs, amplitude was significantly higher than chance level (p <5th/2 percentile) from T -4.5 to T -3.0, from T -2.0 to T -0.5 seconds before, from T +0.5 to T +2.0, and at T +3.5, T +4.5 and T +5.0 seconds after infant social gaze onset (see Tab. 6S and Fig. 5SA). In lullabies, amplitude was significantly lower than chance level (p <5th/2 percentile) at T -1.5 and T -1.0 seconds before infant social gaze onset and significantly higher than chance level (p <5th/2 percentile) at T -3.5, T -2.5 seconds before, and at T +0.5, T +1.0, and T +4.5 seconds after infant social gaze onset (see Tab. 6S and Fig. 5SB).

In comparison to spectral flux, there seem to be some overlaps, like above-chance increases in both spectral flux and amplitude envelope around T + 1.0. However, spectral flux seems to have more above-chance changes. This reflects that spectral flux is comprised of more acoustic qualities than just the amplitude envelope.

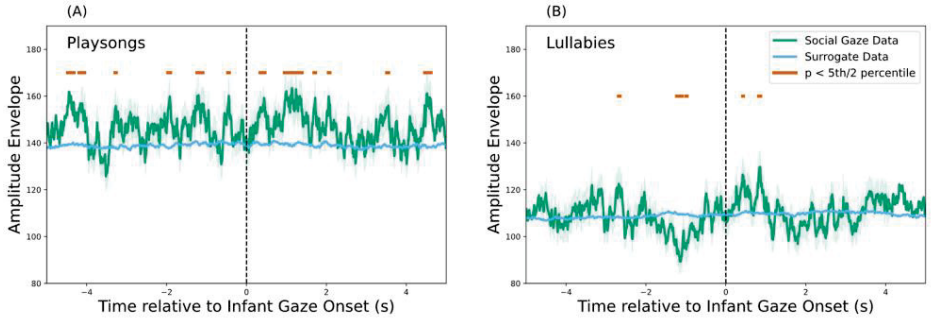


Figure 5S. Mean amplitude of playsongs (A) and lullabies (B) five seconds before and after infant social gaze onset (dashed line), depicting instances of infant social gaze towards the mother (green) vs surrogate time points which excluded time points within 5 s of an infant social gaze onset (blue). Horizontal orange lines indicate time points where the amplitude

during infant social gaze was significantly above or below the amplitude of surrogate looks ($N = 74$).

Table 6S. Exact time points of above-chance changes in amplitude in playsongs and lullabies from five seconds before (T -5.0) to five seconds after (T +5.0) the onset of infant social gaze, chunked into half-second intervals. Time points where the observed spectral flux was significantly below the mean surrogate amplitude are marked with an asterisk (*). The total length (in seconds) of above- or below-threshold changes of every half-second interval is noted in bold and in curly brackets ($N = 74$).

Time (beats)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -4.5	[-4.47 – -4.35], [-4.32 – -4.31] **{0.15}**	
T -4.0	[-4.18 – -4.13], [-4.11 – -4.09], [-4.05 – -4.04] **{0.11}**	
T -3.5	[-3.28 – -3.25] **{0.04}**	[-3.34] **{0.01}**
T -3.0	[-3.21], [-2.90] **{0.02}**	
T -2.5		[-2.68 – -2.64] **{0.05}**
T -2.0	[-1.96 – -1.89] **{0.08}**	
T -1.5	[-1.31], [-1.27], [-1.25] **{0.03}**	[-1.44]*, [-1.27]*, [-1.25]* **{0.03}***
T -1.0	[-1.23 – -1.15], [-1.12 – -1.11], [-1.09 – -1.08], [-1.06], [-0.94] **{0.15}**	[-1.22 – -1.08]*, [-1.03]*, [-0.99 – -0.96]* **{0.20}***
T -0.5	[-0.46 – -0.42] **{0.05}**	
T +0.5	[0.35 – 0.40], [0.42 – 0.47] **{0.12}**	[0.43 – 0.45] **{0.03}**

T +1.0	[0.96 – 1.24] **{0.28}**	[0.81], [0.83 – 0.88] **{0.07}**
T +1.5	[1.25 – 1.29], [1.31 – 1.40], [1.70 – 1.74] **{0.20}**	
T +2.0	[2.04], [2.06 – 2.10] **{0.06}**	
T +3.5	[3.37], [3.52 – 3.57] **{0.07}**	
T +4.5	[4.47 – 4.48], [4.50 – 4.63] **{0.16}**	[4.35] **{0.01}**
T +5.0	[4.90] **{0.01}**	

2.4.2. S. Amplitude Dynamics Around Infant Social Gaze Offset

In playsongs, amplitude was significantly higher than chance level (p <5th/2 percentile) from T -4.0 seconds to T 0 seconds before, and from T +2.5 to T +4.5 seconds after infant social gaze offset (see Tab. 7S and Fig. 6SA). In lullabies, amplitude was significantly lower than chance level (p <5th/2 percentile) at T -3.5, T -3.0, T -0.5 seconds before, and at T +0.5 seconds after the onset of infant social gaze offset, and significantly higher than chance level at T -2.5 seconds before, at T +2.0 to T +3.0, and at T +4.0 seconds after infant social gaze offset. (see Tab. 7S and Fig. 6SB)

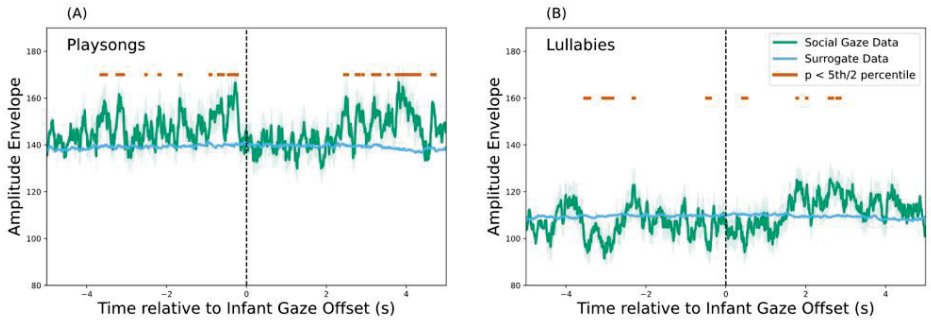


Figure 6S. Mean amplitude (sound envelope) in playsongs (A) and lullabies (B) five seconds before and after infant social gaze offset (dashed line), depicting instances of infant social gaze towards the mother (green) vs surrogate time points which excluded time points within 5 s of an infant social gaze offset (blue). Horizontal orange lines indicate time points where the pitch during infant social gaze was significantly above or below the pitch of non-social surrogate looks (N = 74).

Table 7S. Exact time points of above-chance changes in amplitude in playsongs and lullabies from five seconds before (T -5.0) to five seconds after (T +5.0) the offset of infant social gaze, chunked into half-second intervals. Time points where the observed spectral flux was significantly below the mean surrogate amplitude are marked with an asterisk (*). The total length (in seconds) of above- or below-threshold changes of every half-second interval is noted in bold and in curly brackets ($N = 74$).

Time (beats)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -4.0	[-4.04] **{0.01}**	
T -3.5	[-3.72], [-3.70], [-3.66], [-3.63 – -3.60], [-3.58 – -3.50], [-3.44] **{0.17}**	[-3.52 – -3.43]*, [-3.41 – -3.40]* **{0.12}***
T -3.0	[-3.23 – -3.07] **{0.17}**	[-3.09]*, [-3.07 – -3.01]*, [-2.98 – -2.95]*, [-2.89 – -2.82]* **{0.20}***
T -2.5	[-2.51 – -2.50] **{0.02}**	[-2.31 – -2.28] **{0.04}**
T -2.0	[-2.22], [-2.18 – -2.15] **{0.05}**	
T -1.5	[-1.67 – -1.63], [-1.60], [-1.53] **{0.07}**	
T -1.0	[-0.9 – -0.88] **{0.03}**	
T -0.5	[-0.72], [-0.69 – -0.68], [-0.66 – -0.57], [-0.44 – -0.34] [-0.31 – -0.25] **{0.31}**	[-0.47 – -0.46]*, [-0.44 – -0.41]*, [-0.39 – -0.38]* **{0.08}***
T 0	[-0.24 – -0.22] **{0.03}**	
T +0.5		[0.43 – 0.44]*, [0.47 – 0.50]*, [0.52 – 0.53]* **{0.08}***

T +1.0		[1.18]* **{0.01}***
T +1.5		
T +2.0		[1.77], [1.79 – 1.80], [1.83], [1.85], [1.89], [2.03 – 2.04] **{0.08}**
T +2.5	[2.47 – 2.51], [2.53 – 2.54] **{0.07}**	[2.33], [2.60 – 2.68], [2.70] **{0.11}**
T +3.0	[2.76 – 2.83], [2.87], [2.92 – 2.93], [3.17 – 3.22] **{0.17}**	[2.79 – 2.87], [3.18] **{0.10}**
T +3.5	[3.25 – 3.26], [3.28 – 3.35], [3.51], [3.56 – 3.57], [3.64], [3.66] **{0.15}**	
T +4.0	[3.76 – 4.05], [4.07 – 4.19], [4.21 – 4.22] **{0.45}**	[4.17], [4.23] **{0.02}**
T +4.5	[4.27 – 4.28], [4.33 – 4.35], [4.65 – 4.73] **{0.14}**	

2.5. S. Permutation analyses – Pitch (Fundamental Frequency)

2.5.1. S. Pitch Dynamics Around Infant Social Gaze Onset

In playsongs, pitch was significantly lower than chance level (p <5th/2 percentile) at T -1.5 seconds before and T +4.0 seconds after infant social gaze onset (see Tab. 8S and Fig. 7SA). In lullabies, amplitude was significantly lower than chance level (p <5th/2 percentile) at T -4.5 and T -0.5 seconds before infant social gaze onset (see Tab. 8S and Fig. 7SB). The sporadicity of robust pitch changes could be due to songs having a distinct pitch that mothers could not vary as much as amplitude or spectral flux in response to infant attention.

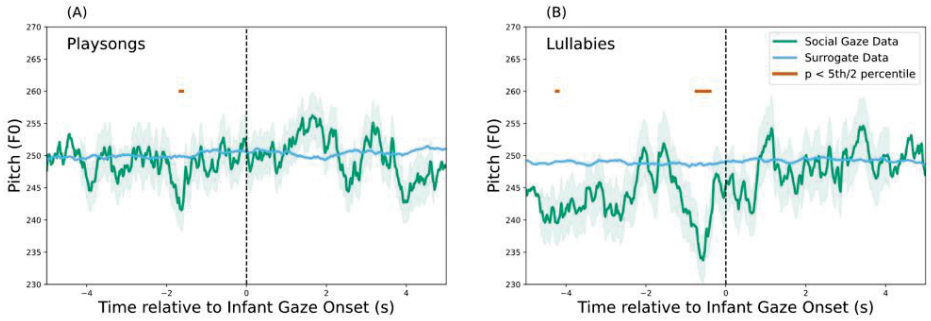


Figure 7S. Mean pitch (fundamental frequency) in playsongs (A) and lullabies (B) five seconds before and after infant social gaze onset (dashed line), depicting instances of infant social gaze towards the mother (green) vs surrogate time points which excluded time points within 5 s of an infant social gaze onset (blue). Horizontal orange lines indicate time points where the pitch during infant social gaze was significantly below the pitch of non-social surrogate looks ($N = 74$).

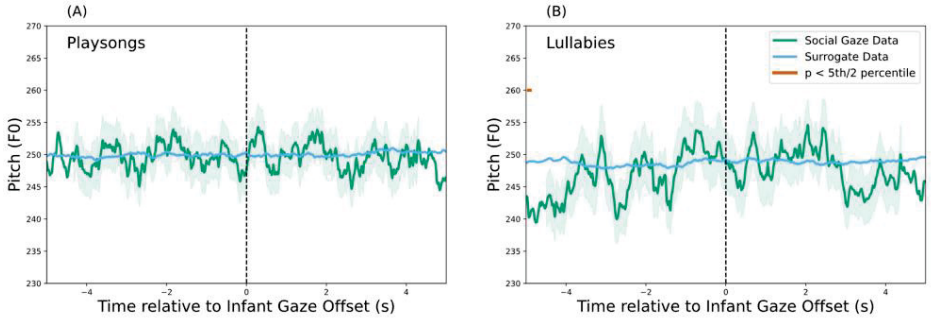
Table 8S: Exact time points of above-chance changes in pitch in playsongs and lullabies from five seconds before (T -5.0) to five seconds after (T +5.0) the onset of infant social gaze, chunked into half-second intervals. Time points where the observed spectral flux was significantly below the mean surrogate pitch are marked with an asterisk. The total length in seconds of every half-second interval is noted in bold and in curly brackets ($N = 74$).

Time (beats)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -4.0		[-4.23 – -4.19]* **{0.05}***
T -1.5	[-1.65 – -1.59]* **{0.07}***	
T -0.5		[-0.73 – -0.39]* **{0.35}***

2.5.2. S. Pitch Dynamics Around Infant Social Gaze Offset

In playsongs, pitch was not significantly different from chance level (p <5th/2 percentile) around infant social gaze offset (see Tab. 9S and Fig. 8SA). In lullabies, pitch was

194 significantly lower than chance level (p <5th/2 percentile) at T -5.0 before the offset of infant
195 social gaze (see Tab. 9S and Fig. 8SB).



196
197 *Figure 8S*. Mean pitch (fundamental frequency) in playsongs (A) and lullabies (B) five
198 seconds before and after infant social gaze offset (dashed line), depicting instances of infant
199 social gaze towards the mother (green) vs surrogate time points which excluded time points
200 within 5 s of an infant social gaze offset (blue). Horizontal orange lines indicate time points
201 where the pitch during infant social gaze was significantly below the pitch of non-social
202 surrogate looks (N = 74).

203
204
205 *Table 9S*. Exact time points of above-chance changes in pitch (fundamental
206 frequency) in playsongs and lullabies from five seconds before (T -5.0) to five seconds after
207 (T +5.0) the offset of infant social gaze, chunked into half-second intervals. Time points
208 where the observed spectral flux was significantly below the mean surrogate pitch are marked
209 with an asterisk. The total length in seconds of every half-second interval is noted in bold and
210 in curly brackets (N = 74).

Time (beats)	Playsongs **{Length (s)}**	Lullabies **{Length (s)}**
T -5.0		[-4.93 – -4.89]* **{0.05}***

211