frontiers | Research Topics

Early development of sound processing in the service of speech and music perception

Edited by
István Winkler, Judit Gervain, Marcela Pena, Laurel J. Trainor
and Teija Kujala

Published in
Frontiers in Human Neuroscience
Frontiers in Psychology

ISSN 1664-8714
ISBN 978-2-8325-5354-1
DOI 10.3389/978-2-8325-5354-1

Early development of sound processing in the service of speech and music perception

Topic editors

István Winkler — Research Centre for Natural Sciences, Hungarian Academy of Sciences (MTA), Hungary

Judit Gervain — Centre National de la Recherche Scientifique (CNRS), France

Marcela Pena — Pontificia Universidad Católica de Chile, Chile

Laurel J. Trainor — McMaster University, Canada

Teija Kujala — University of Helsinki, Finland

Citation

Winkler, I., Gervain, J., Pena, M., Trainor, L. J., Kujala, T., eds. (2024). *Early development of sound processing in the service of speech and music perception*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5354-1

Table of contents

frontiers | Frontiers in Human Neuroscience

TYPE Editorial
PUBLISHED 13 August 2024
DOI 10.3389/fnhum.2024.1471445

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Lutz Jäncke,
University of Zurich, Switzerland

*CORRESPONDENCE
Judit Gervain
✉ judit.gervain@parisdescartes.fr
István Winkler
✉ winkler.istvan@ttk.hu

RECEIVED 27 July 2024
ACCEPTED 30 July 2024
PUBLISHED 13 August 2024

CITATION
Gervain J, Kujala T, Peña M, Trainor LJ and Winkler I (2024) Editorial: Early development of sound processing in the service of speech and music perception.
Front. Hum. Neurosci. 18:1471445.
doi: 10.3389/fnhum.2024.1471445

Editorial: Early development of sound processing in the service of speech and music perception

Judit Gervain[1,2]*, Teija Kujala[3], Marcela Peña[4,5], Laurel J. Trainor[6] and István Winkler[7]*

[1]Integrative Neuroscience and Cognition Center, CNRS & Université Paris Cité, Paris, France, [2]Department of Social and Developmental Psychology, University of Padua, Padua, Italy, [3]Cognitive Brain Research Unit & Centre of Excellence in Music, Mind, Body, and Brain, Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [4]School of Psychology, Cognitive Neuroscience Lab, Pontificia Universidad Católica de Chile, Santiago, Chile, [5]National Center for Artificial Intelligence, CENIA FB210017, Basal ANID, Santiago, Chile, [6]Psychology, Neuroscience & Behaviour and the LIVELab, McMaster University, Hamilton, ON, Canada, [7]Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

KEYWORDS

early development, auditory perception, speech perception, music perception, infants, toddlers

Editorial on the Research Topic
Early development of sound processing in the service of speech and music perception

Speech and music are the two structurally most complex auditory signals that infants typically encounter. Yet, even in the presence of multiple sound streams, healthy human infants display signs of music perception, such as moving to a musical rhythm as early as a few months of life and by 18–24 months of age they understand many simple sentences. Extracting and analyzing speech and music at such an early age proves that many of the higher-order processing capabilities, such as regularity detection, auditory stream segregation, statistical learning, and rhythm processing are already present at birth or develop quite early during infancy. Understanding how the infant brain processes sound not only provides insights into the neural and processing prerequisites of speech and music perception and—compared to adults—a simpler model of the mechanics of these functions, but it is also essential for developing early interventions for atypically developing infants, such as designing training protocols for infants at risk of auditory developmental deficits.

The present Research Topic of papers consists of empirical studies and reviews examining the functioning of auditory information processing necessary for speech and language perception in infancy. The 11 papers collected can be sorted into four main research area: (1) higher-order auditory functions supporting speech and music processing, (2) rhythm, (3) speech processing in infants, and (4) predicting the quality of the outcome of language acquisition from data collected from prelingual infants and their families. Here we will summarize the main conclusions of the papers in the Research Topic in relation to the state-of-the-art of their larger research fields.

Higher-order auditory functions supporting speech and music processing

In developing their capabilities to extract information from speech and music, infants rely on general auditory functions supporting in-depth analysis of sound sequences. Because multiple sound sources are active in real-life situations, detailed analysis of any sound sequence must be preceded by separating it from the rest of the sounds (auditory stream segregation). Calcus' review of the development of auditory scene analysis from infancy to adulthood concludes that while some of the processes segregating auditory streams by sequential as well as by simultaneous cues are present at birth, these functions have long developmental trajectories.

Both speech and music abound in dependencies between nonadjacent elements of sound sequences. Mueller et al., showed that 3-year-old children can learn the dependency between two pure tones separated by a random third one. Further, the size of the electroencephalographic response to infrequent tones violating the dependency rule correlates with that to the response to deviance in tone intensity.

Rhythm processing in infants

The majority of previous studies on infants' rhythm processing have focused on auditory stimuli, but in real environments, rhythms occur in multimodal contexts. The two studies on rhythm in the present Research Topic targeted different multimodal interactions. Cirelli et al. explored whether infants attend (look longer) to videos where a hand taps are at the same tempo as an auditory rhythm compared to when there is a mismatch in tempo. They found no evidence for this, but exploratory analyses suggested there are complex interactions between pitch, tempo and rhythmic complexity that affect infants' attention to audio-visual synchrony. Boll-Avetisyan et al. explored auditory-motor interactions in the context of duration, pitch and intensity cues to rhythmic grouping structure in isochronous speech-syllable streams. While, as predicted, duration cues led to the greatest amount of infant rhythmic movement, they found the opposite of their prediction that infants who engaged in more rhythmic movement would show better speech segmentation. This leads to questions about how infant rhythmic movements relate to language learning. Both studies are intriguing in demonstrating the complexity of multi-modal rhythm processing and open the door for more research in this important area.

Speech processing in infants

Infancy includes rapid development of skills needed for understanding speech: identification of phonemes, integrating them to wordforms, and detecting morpheme and word relations. Whereas it has been shown earlier that already neonates can distinguish between phonemes, the study of Hegde et al. showed that the developmental trajectories for acquiring vowels and consonants may differ. They found that between 6 and 10 months, infants have different trajectories in the perceptual weight of temporal acoustic cues for consonant and vowel processing. Piot et al., in turn, found that 9-month-old infants are even sensitive to regularities of phoneme contrasts, which are difficult to discriminate at an early age. The intensive period of phoneme learning during the 1st year of life is followed by, and partly overlapping with, vocabulary acquisition. Ylinen et al. assessed in 1-year-olds the processing of word forms right after learning them compared to words learned earlier. They found that newly learned and earlier learned words appear to be processed similarly. A less studied but potentially important stage of language learning is the fetal stage. The study of Gorna-Careta et al. found that neonates of bilingual compared to monolingual mothers are more sensitive to a wider range of speech frequencies, possibly due to the greater speech signal complexity of bilingual mothers.

These early steps of speech analysis are followed by the acquisition of linguistic meaning, of which Forgács presents an interesting review and proposes that the access to meaning is possible thanks to the development of the theory of mind ability, providing some of the few available infant brain data in the field. This article opens the door to provoking questions such as how the attribution of mental beliefs arises in the absence of meaning.

Predicting language developmental outcomes

In the last decade, individual predictors of language outcomes have received increasing attention, in part in an attempt to provide early and efficient interventions for children at risk for language delay and disorders, as these have a strong negative impact on academic and later professional outcomes. In this vein, Ortiz-Barajas showed that newborn infants' differential theta oscillations discriminating between their native language and a rhythmically different unfamiliar language predict infants' vocabulary sizes at 12 and 18 months. Non-linguistic abilities also contribute to language development. Balázs et al. found that apart from gender and gestational age, already known predictors of language development, temperament also contributed to linguistic abilities in infants and toddlers, with more sociable and responsive children showing better language skills.

The current Research Topic of papers demonstrates that while studying the roots and early development of speech and music perception requires much ingenuity and often poses methodological challenges, there is substantial progress in the field through the efforts of an ever-widening circle of researchers world-wide.

Author contributions

JG: Writing – original draft, Writing – review & editing. TK: Writing – original draft, Writing – review & editing. MP: Writing – original draft, Writing – review & editing. LT: Writing – original draft, Writing – review & editing. IW: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

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

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 January 2024
DOI 10.3389/fpsyg.2023.1321311

OPEN ACCESS

EDITED BY
István Winkler,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences (MTA),
Hungary

REVIEWED BY
Evelyne Mercure,
Goldsmiths University of London,
United Kingdom
Anna Kohari,
Hungarian Academy of Sciences, Hungary

*CORRESPONDENCE
Monica Hegde
✉ monica.s.hegde@gmail.com

RECEIVED 17 October 2023
ACCEPTED 11 December 2023
PUBLISHED 24 January 2024

CITATION
Hegde M, Nazzi T and Cabrera L (2024) An auditory perspective on phonological development in infancy.
Front. Psychol. 14:1321311.
doi: 10.3389/fpsyg.2023.1321311

An auditory perspective on phonological development in infancy

Monica Hegde*, Thierry Nazzi and Laurianne Cabrera

Integrative Neuroscience and Cognition Center (INCC-UMR 8002), Université Paris Cité-CNRS, Paris, France

Introduction: The auditory system encodes the phonetic features of languages by processing spectro-temporal modulations in speech, which can be described at two time scales: relatively slow amplitude variations over time (AM, further distinguished into the slowest <8–16 Hz and faster components 16–500 Hz), and frequency modulations (FM, oscillating at higher rates about 600–10 kHz). While adults require only the slowest AM cues to identify and discriminate speech sounds, infants have been shown to also require faster AM cues (>8–16 Hz) for similar tasks.

Methods: Using an observer-based psychophysical method, this study measured the ability of typical-hearing 6-month-olds, 10-month-olds, and adults to detect a change in the vowel or consonant features of consonant-vowel syllables when temporal modulations are selectively degraded. Two acoustically degraded conditions were designed, replacing FM cues with pure tones in 32 frequency bands, and then extracting AM cues in each frequency band with two different low-pass cut- off frequencies: (1) half the bandwidth (Fast AM condition), (2) <8 Hz (Slow AM condition).

Results: In the Fast AM condition, results show that with reduced FM cues, 85% of 6-month-olds, 72.5% of 10-month-olds, and 100% of adults successfully categorize phonemes. Among participants who passed the Fast AM condition, 67% of 6-month-olds, 75% of 10-month-olds, and 95% of adults passed the Slow AM condition. Furthermore, across the three age groups, the proportion of participants able to detect phonetic category change did not differ between the vowel and consonant conditions. However, age-related differences were observed for vowel categorization: while the 6- and 10-month-old groups did not differ from one another, they both independently differed from adults. Moreover, for consonant categorization, 10-month-olds were more impacted by acoustic temporal degradation compared to 6-month-olds, and showed a greater decline in detection success rates between the Fast AM and Slow AM conditions.

Discussion: The degradation of FM and faster AM cues (>8 Hz) appears to strongly affect consonant processing at 10 months of age. These findings suggest that between 6 and 10 months, infants show different developmental trajectories in the perceptual weight of speech temporal acoustic cues for vowel and consonant processing, possibly linked to phonological attunement.

KEYWORDS

infants, amplitude modulations, frequency modulations, vocoder, speech perceptual attunement, phonetic processing

1 Introduction

The auditory system encodes the phonetic features of a given language by processing fine spectro-temporal acoustic changes in the speech signal. Even with a relatively immature auditory system (Moore, 2002), infants have been shown to distinguish phonetic contrasts in a language-specific manner before the end of their first year of life (see Kuhl, 2004; Saffran et al., 2006). However, it remains unclear whether infants and adults rely on the exact same acoustic information when discriminating native phonetic contrasts. To this aim, the current study compares the reliance upon spectro-temporal acoustic cues of speech in a phonetic feature discrimination task between infants at two ages (6 and 10 months) and adults. This study aims to investigate whether infants at different developmental stages, as well as adults, use the same acoustic information to discriminate vowels and consonants in their native language.

To explore infants auditory processing of speech, the present study uses a psychoacoustic approach that has been described extensively over the last decades and modeled the stages of auditory processing in adult listeners (c.f., Moore and Linthicum, 2007). A key concept of this psychoacoustic approach is to consider that the human auditory system decomposes any complex acoustic signal (including speech) into its fine spectral and its fine temporal modulations. The decomposition of the spectral modulations is related to the sensitivity of inner hair cells within the basilar membrane of the cochlea to a specific audio frequency range. The selective spectral processing of audio frequency from the high frequencies at the base of cochlea to low frequencies at the apex can be modeled as a bank of narrowband filters with a passband equal to one equivalent-rectangular bandwidth (ERB, Glasberg and Moore, 1990; Moore, 2003). Then, the auditory system is thought to decompose the temporal components of each extracted narrowband signal at two main time scales: relatively slow amplitude variations over time (amplitude modulations or AM, often referred to as temporal envelope), and relatively fast oscillations over time (frequency modulation or FM, often referred to as temporal fine structure). These models helped to develop speech analysis-synthesis tools, called vocoders, to assess selectively the specific role of spectral and temporal components in speech perception. Using vocoders, the spectro-temporal complexity of an original speech can be selectively manipulated.

In adults, a wealth of studies using vocoders showed that FM cues convey essential information related to voice pitch, and play an important role in speech perception in quiet for lexical-tone languages (using pitch at the syllable level, e.g., Zeng et al., 2005; Kong and Zeng, 2006). Moreover, sentence recognition has been found to be more difficult when only FM cues are preserved in the signal (Gilbert and Lorenzi, 2006; Lorenzi et al., 2006; Sheft et al., 2008; Hopkins et al., 2010), but FM cues provide crucial information in noisy environments (e.g., Zeng et al., 2005; Hopkins et al., 2008; Hopkins and Moore, 2009; Ardoint and Lorenzi, 2010). Nevertheless, AM cues have been found to convey information related to syllabic and phonetic information that allow word and sentence identification in quiet listening conditions (Rosen, 1992; Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005; Lorenzi et al., 2006; Sheft et al., 2008). This was initially demonstrated by Shannon et al. (1995) using noise-excited vocoders to investigate the impact of spectro-temporal degradation on speech identification. In that study, the researchers took original input sentences and applied a filter-bank to decompose the signal into 1, 2, 3, or 4 frequency bands from which the original AM and FM cues were decomposed. While the FM was replaced by a noise carrier in each band, the AM cues were low-pass filtered at different cutoff frequencies (16, 50, 160, or 500 Hz). Sentence identification scores in quiet were almost perfect in the 4 band-AM condition but decreased with a reduced number of frequency bands. Moreover, sentence recognition scores were worse in the condition where AM cues were preserved only below 16 Hz. Other studies showed that faster AM cues transmit some information regarding voice pitch information (Kong and Zeng, 2006) as well as formant transitions (Rosen, 1992).

While it has been repeatedly observed that adults are able to correctly identify speech in quiet with only the slowest AM cues (<8–16 Hz), the identification of individual phonetic features becomes more nuanced in terms of what acoustic cues are used. Using confusion matrices of phonemes, Shannon et al. (1995) showed that the reduction of faster AM cues (>16 Hz) significantly affected consonant identification, but not vowel identification. Moreover, for consonants, the identification of place of articulation remained challenging even in the 4-band AM condition. More recently, Xu et al. (2005) conducted a systematic study to determine the importance of various spectral and temporal information in phoneme identification. English-speaking adults were asked to identify consonants and vowels that varied in voicing, place of articulation, manner of articulation, duration, first formant (F1) frequency and second formant (F2) frequency. Syllables were vocoded using different numbers of bands (ranging from 1 to 16) and different low-pass filters for AM extraction (ranging from 1 to 512 Hz). Their findings showed that the optimal low-pass cutoff frequency for consonant recognition was 16 Hz, whereas for vowel recognition it was 4 Hz. Regarding spectral information, consonant recognition performance reached a plateau at 8 bands, while for vowel recognition it was 12 bands. These findings from adult studies show that AM cues are the most important cue for overall speech recognition in quiet (i.e., at the sentence recognition level), but that identification of consonants and vowels require different contributions of fast and slow AM, and FM cues. In other words, this demonstrates that various spectro-temporal cues play distinct functional roles in phoneme identification. However, it is important to note that these conclusions concern listeners with a mature auditory system and a well developed linguistic system.

To tackle developmental issues, vocoders have also been used to investigate how young listeners and especially infants use acoustic cues when processing speech sounds. Although this field of research is still largely emerging, the first infants studies using vocoders suggest that AM and FM cues have a different role at early ages compared to adults. For vowels, only one study to date has assessed English-learning 6-month-olds ability to detect a phonetic change in degraded speech. This study tested discrimination between /a/ and /i/ in vocoder conditions reducing FM cues and the number of spectral bands for AM extraction. Infants were found to detect a vowel change when the original AM (160 Hz cut-off frequency) was presented within 32 bands, but not when it was presented within 16 bands (Warner-Czyz et al., 2014). It is however not clear yet whether infants require faster fluctuations of AM to process vowels.

For consonants, on the other hand, a handful of studies have investigated phonetic discrimination in young infants. Two studies used looking-time recording procedures to familiarize or habituate French-learning infants to one specific vowel-consonant-vowel sequence processed in one vocoder condition. The findings reveal that 6-month-olds were able to distinguish /aba/ from /apa/ when the slowest (<16 Hz) AM cues were preserved in only 32 bands, but that they required an increased time of listening to display this behavior compared to a condition where the original (<ERB/2) AM cues were preserved (Cabrera et al., 2013, 2015a). These studies demonstrate that 6-month-old infants can effectively use slow (<16 Hz) AM cues for consonant voicing or place discrimination, but that faster AM cues may play an important role in early phonetic discrimination. Results along this line were also found in younger infants in a more recent study by Cabrera and Werner (2017) using an observer-based psychophysical procedure, the method used in the present study. English-speaking adult and English-learning 3-month-old participants were presented with one of five consonant categories (voiceless, voiced, labial, coronal, velar). In a yes-no task, participants were presented with a series of background syllables that exemplified the category under examination (e.g., voiced syllables like /ba/, /da/, /ga/, randomly repeated). They were evaluated based on their ability to detect change trials, where a single randomly selected “target” syllable (e.g., voiceless syllables like /pa/, /ta/, or /ka/) was played, and to withhold responses during no-change trials, where a background syllable was presented. Both infants and adults were tested on their ability to discriminate consonants under quiet or noisy conditions in two vocoder conditions: (1) Fast AM, in which the original AM (filtered < 256 Hz) was preserved in 32 bands and FM was replaced by a pure tone, and (2) Slow AM, in which only the slowest AM (filtered < 8 Hz) was preserved in 32 bands and FM was replaced by a pure tone. Adults were able to discriminate consonants in both vocoder conditions in quiet environments. However, in noisy environments, the percentage of adults who correctly discriminated the consonant changes decreased from 70% to 20% between the Fast and the Slow AM conditions. These results confirmed that the slowest AM cues are not sufficient for adults consonant discrimination in noise. Infants did not discriminate consonants equally in both vocoder conditions in quiet environments. The percentage of infants who discriminated decreased from 81% to 50% between the Fast and Slow AM conditions. In noisy environments, a similar pattern emerged, with the percentage of infants discriminating decreasing from 96% to 48% between the Fast and Slow AM conditions. In summary, these first infant studies using vocoders suggest that 3- and 6-month-old infants may not rely on exactly the same spectro-temporal modulations as adults when processing phonemes. However, the age at which infants start to use, or weight, the acoustic cues found to be used by adults to process speech remains unknown. This developmental shift must occur between early infancy and adulthood, and possibly when infants start to process speech sounds in a language-specific manner, that is, during the second half of the first year of life.

The present study aims to investigate the development of the early auditory processing of speech to provide further insights into the acquisition of the phonological properties specific to one’s native language. Interestingly, during the first year of life, infants show asynchronous perceptual attunement to the vowels and the consonants of their native language. Specifically, infants start becoming attuned to native language vowels around 4–6 months of age (Trehub, 1976; Kuhl et al., 1992; Polka and Werker, 1994), earlier than when they start becoming attuned to native language consonants around 8-10 months of age (Trehub, 1976; Werker and Tees, 1984; Best et al., 1988, 1995). Furthermore, at the lexical level, differences in processing of vowels and consonants are also found, showing a shift in infants reliance from vowels to consonants between 6 and 8-11 months of age when detecting word forms (Bouchon et al., 2015; Poltrock and Nazzi, 2015; Nazzi et al., 2016; Nishibayashi and Nazzi, 2016). The question arises as to whether changes in spectro-temporal cue processing occur during this same developmental time window, and could thus be linked to phonological acquisition during the first year of life.

While no study to date has explored this issue directly, one study investigated the development of spectro-temporal cue weighting in a cross-linguistic study comparing French- versus Mandarin-learning infants. Cabrera et al. (2015b) investigated whether native language exposure influences reliance upon AM and FM cues in a discrimination task measuring looking times for two syllables varying in lexical tone (that is a change in pitch at the syllable level, such contrasts being phonological in tonal languages such as Mandarin Chinese, but not in French). Results showed that at 6 months, French- and Mandarin-learning infants display the same pattern of response: they detected a change in lexical tones in an intact condition (without acoustic degradation), suggesting that French-learning infants were not yet attuned to this speech contrast, and both groups did not detect the change when fine spectral and FM cues were degraded, showing that these acoustic cues are required for lexical-tone detection at 6 months. However, at 10 months, an influence of language background was observed: Mandarin-learning 10-month-olds showed the same pattern of response as 6-month-olds, but French-learning 10-month-olds were not able to detect the lexical-tone change in the intact condition, showing perceptual reorganization for this speech contrast. Moreover, French-learning 10-month-olds were able to discriminate the lexical tones when fine spectral and FM cues were degraded. These results suggest that native language exposure plays a role in the development of acoustic cue weighting during the phonological reorganization period.

Accordingly, the current study focused on infants of 6 and 10 months of age exposed to French and compares their reliance upon FM and AM cues when detecting native vowel or consonant feature contrasts to assess whether with age infants rely more on slow or faster temporal cues when processing native phonemes. The present study will extend the findings of Cabrera and Werner (2017), using an observer-based psychophysical yes-no task to measure the proportions of listeners able to detect a phonetic change in two vocoder conditions. Particularly, we compared the number of adults, 10-month-old and 6-month-old infants correctly detecting vowel or consonant changes in quiet based on various types of phonetic features, and in two vocoder conditions reducing increasingly FM and AM cues.

Three groups of participants were tested in the exact same experimental conditions and setup: 6-month-olds, who have started to attune to the vowels but not the consonants of their native language; 10-month-olds, who have started to attune to both vowels and consonants of their native language; and adults. Eight phonetic conditions were designed to assess the ability of listeners to detect a change in: Vowel Place, Vowel Height, Consonant Place and Consonant Voicing, each tested in two vocoder conditions, a Fast AM condition (preserving the original AM cues by using a cutoff frequency of ERB/2 that preserves fast and slow AM, in 32 bands, with reduced FM cues), and a Slow AM condition (preserving only the slowest AM cues below 8 Hz, in 32 bands, with reduced FM cues). Listeners were exposed to only one phonetic feature contrast in its two vocoder conditions, starting with the Fast AM condition and then, if they succeeded, moving to the Slow AM condition. Therefore, this study specifically examines: (1) the contributions of FM, Fast AM, and Slow AM cues for phonetic categorization (2) the role of these cues at distinct developmental points, (3) how these cues influence the categorization of vowels and consonants, and (4) the impact of different phonetic features in the aforementioned categorization.

Based on prior behavioral studies, we expected a higher success rate among 6-month-olds in the Fast AM condition compared to the Slow AM condition, as these infants typically exhibit a stronger weighting of Fast AM cues. Nonetheless, as 6-month-olds have already started to attune to the vowels of their native language, we hypothesized that temporal degradation may have a more pronounced effect on consonant detection than on vowel detection. For 10-month-olds, we predicted similar performance for both the Fast AM and Slow AM conditions, as they have started to attune to both vowels and consonants of their native language. As such, we expected any difference between the effect of temporal degradation on vowels and on consonants to be less pronounced in 10-month-olds than in 6-month-olds. For adults, we predicted near-ceiling performance in all conditions.

2 Methods

2.1 Participants

Participants were recruited through the Babylab Participant Pool at the Integrative Neuroscience and Cognition Center. The data of 40 6-month-old infants (mean: 28.2 weeks, range: 25.9 weeks–31.8 weeks; 24 girls, 16 boys), 40 10-month-old infants (mean: 45.9 weeks, range: 42.4 weeks–49.6 weeks; 16 girls, 24 boys) and 20 adults (mean: 21 years; range: 18 to 29 years; 13 females, 7 males) were included in the analyses. All infants were born full term, had no history of otitis media within 3 weeks of testing with no more than 2 prior occurrences of otitis media, had no risk factors for hearing loss, were French monolinguals (French input > 90% of the time) and had no history of health or developmental concerns. All adult participants were native French monolingual speakers, reported typical hearing bilaterally and had no history of noise exposure. Informed consent forms were obtained from all infants legal guardians and adult participants as approved by the university ethics committee. Data from an additional three 6-month-olds and two 10-month-old were excluded because the infants were too tired or fussy to complete the task; and data from three 6-month-olds and four 10-month-olds were excluded because parents did not come back for the second testing session.

2.2 Stimuli

A total of 16 Consonant-Vowel (CV) syllables were chosen so that by different recombinations, they could be used to define two vowel categories contrasted on place, or two vowel categories contrasted on height or two consonant categories contrasted on place, or two consonant categories contrasted on voicing. Each of these categories was made up of eight syllables, in which both vowels and consonants were varied (see Table 1). The 16 CV syllables used in the study are as follows: pu, bu, tu, du, po, bo, to, do, py, by, ty, dy, pø, bø, tø, and dø. The CV syllables were recorded in a sound-attenuated room and digitized with 16-bit resolution at a 44.1-kHz sampling rate. A female French native speaker who was instructed to "speak clearly" produced several tokens of all the CVs and five tokens for each were selected for their clarity. All tokens were comparable in duration (range = 263:411 ms, mean = 338 ms; SD = 31 ms) and F0 (mean = 238 Hz). All stimuli were equated at the global root-mean-square (RMS) level.

The original stimuli were processed by two vocoders to alter the spectro-temporal modulations. Tone-excited vocoders were used instead of noise-excited vocoders, because they distort speech AM cues less (e.g., Kates, 2011). In each vocoder condition, the original speech signal was passed through a bank of 32 2nd-order gammatone filters (Patterson, 1987; Gnansia et al., 2009), each 1-ERB wide with center frequencies (CFs) uniformly spaced along an ERB scale ranging from 80 to 8,020 Hz. The Hilbert transform was then applied to each bandpass filtered speech signal to extract the AM component and FM carrier. The FM carrier in each frequency band was replaced by a sine wave carrier with a frequency at the CF of the gammatone filter and random starting phase. The AM component was low-pass filtered using a zero-phase Butterworth filter (36 dB/octave roll off) with a cutoff frequency set to either ERBN/2 (Fast AM Condition) or 8 Hz (AM < 8 Hz condition, Slow AM Condition). Each tone carrier was multiplied by the corresponding filtered AM function. The narrow-band speech signals were finally added up and the level of the wideband speech signal was adjusted to have the same RMS value as the input signal. Figure 1 represents the spectrograms of exemplary tokens illustrating the four different contrast types (i.e., vowel place, vowel height, consonant place, consonant height) in the two vocoded conditions.

2.3 Procedure, material, and apparatus

Infants were tested using an observer-based psychophysical procedure (Werner, 1995). This procedure is similar to the classical head-turn conditioning procedure used in psycholinguistic studies (Werker et al., 1997), with two key differences: (1) any behavioral change from the infants is considered a response to a sound change, not just head turns, and 2) false alarms are recorded to ensure that the detected behavioral change corresponds to a sound change (Olsho et al., 1987). During testing, infants sat on a caregiver's lap with an assistant inside a sound-attenuating booth. A TV screen was placed on the right of the participant. The infant listened to sounds through an insert earphone (ER-2), calibrated to deliver the sounds at 65 dB SPL, ensuring that none of the adults involved

TABLE 1 Eight distinct phonetic conditions were established.

	Feature contrast	Background	Target
Vowel contrasts	Place	Back: pu, bu, tu, du, po, bo, to, do	Front: py, by, ty, dy, pø, bø, tø, dø
		Front: py, by, ty, dy, pø, bø, tø, dø	Back: pu, bu, tu, du, po, bo, to, do
	Height	Open : po, bo, to, do, pø, bø, tø, dø	Closed: py, by, ty, dy, pu, bu, tu, du
		Closed: py, by, ty, dy, pu, bu, tu, du	Open: po, bo, to, do, pø, bø, tø, dø
Consonant contrasts	Place	Labial: py, pø, pu, po, by, bø, bu, bo	Coronal: ty, tø, tu, to, dy, dø, du, do
		Coronal: ty, tø, tu, to, dy, dø, du, do	Labial: py, pø, pu, po, by, bø, bu, bo
	Voicing	Voiced: by, bø, bu, bo, dy, dø, du, do	Voiceless: py, pø, pu, po, ty, tø, tu, to
		Voiceless: py, pø, pu, po, ty, tø, tu, to	Voiced: by, bø, bu, bo, dy, dø, du, do

In every condition, background syllables were played in a random sequence. During a "change" trial, one randomly selected "target" syllable replaced a background syllable. Conversely, in a "no-change" trial, only a background syllable was played. The specific target syllables were determined based on the phonetic condition being tested.

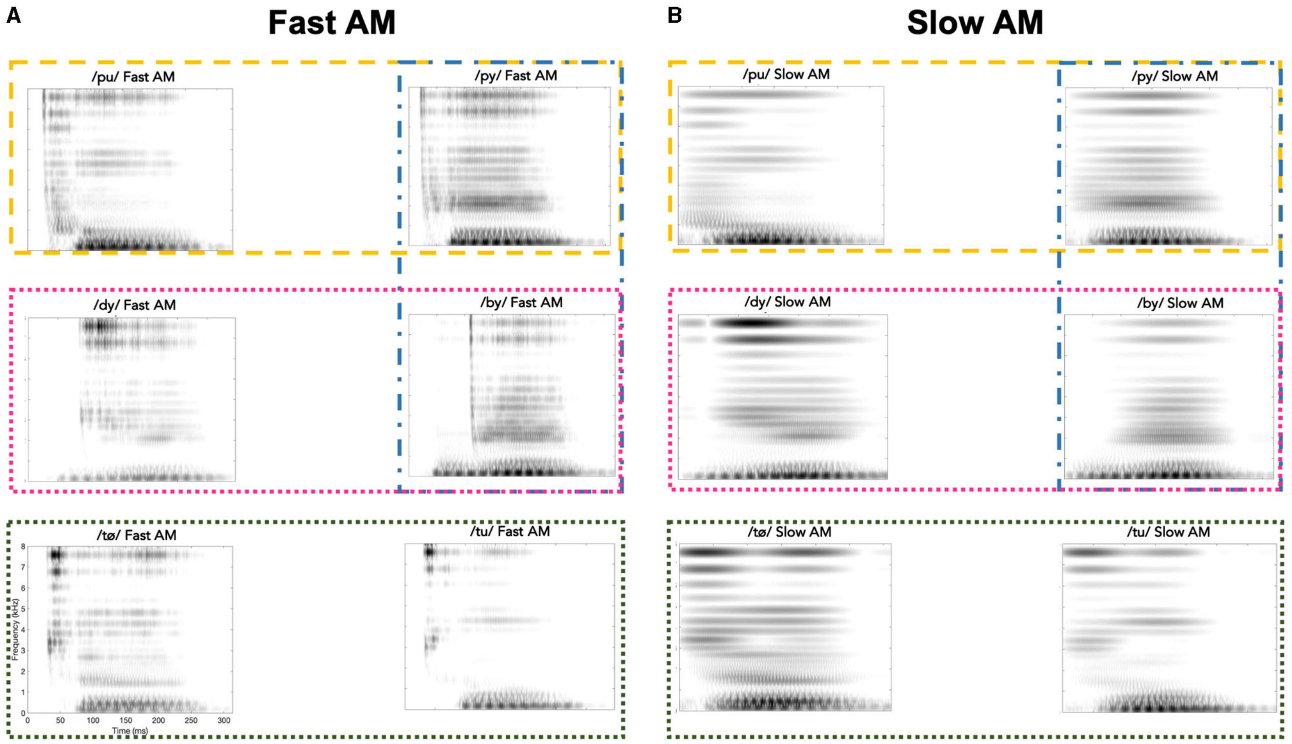


FIGURE 1
Spectrograms of exemplary syllable tokens in the two vocoded conditions: **(A)** for Fast AM condition and **(B)** for Slow AM condition. The syllables /pu/, /dy/, /tø/ are represented on the left columns, and the syllables /py/, /by/, /tu/ on the right columns. The four different phonetic feature contrasts are illustrated by colored rectangles (vowel place, vowel height, consonant place, consonant voicing by dashed yellow lines, dotted dark green lines, dotted pink lines and dash-dotted blue lines, respectively).

could hear the stimuli presented to the infant. The caregiver was instructed to avoid interacting with the infant.

The experimenter (or "observer", who was the same for all infants) sat outside the booth and observed the infant through a one-way mirror. A microphone inside the booth enabled the experimenter to listen to the infant and assistant, and a microphone outside the booth allowed the experimenter to communicate with the assistant who was wearing headphones. The assistant listened to the experimenter's instructions and manipulated toys silently to keep infants facing midline. A computer controlled the experiment. Adult participants were tested using the same setup, except that they sat alone in the booth. An advantage of the observer-based procedure over procedures previously used to assess infants discrimination of vocoded speech is that adults can be tested in the same procedure as a basis of comparison.

The participant heard repeated, randomly selected tokens from one "background" category, separated by silences of 800 ms. Each infant participant was tested in only one phonetic condition from Table 1, so that 10 infant participants completed the task in each age group and phonetic condition. Each adult was tested in 2 conditions (one Vowel, one Consonant) in a random order, varying which vowel and consonant condition was presented, so that 10 adult participants completed the task in each phonetic condition.

Test trials were initiated by the experimenter at moments when the participant was quietly listening to the syllables from the background category and facing midline. There were two trial types: on change trials, a syllable from the target category was presented once, while on no-change trials, a syllable from the background category was presented once. On each trial, the experimenter, blind to trial type, had 4 s from trial onset to decide whether the participant had reacted, that is, had produced a behavior during that time window, and to press a button if such a behavioral change was detected. For infants, the behaviors coded as response by the experimenter varied from infant to infant, and commonly observed behaviors included eye movements, increases and decreases in body movement, and facial expressions. Adults were instructed to raise their hand when they detected a change in the sounds. Computer feedback was provided to the experimenter at the end of a trial to indicate hit, miss, correct rejection, or false alarm. Participants responses were automatically reinforced with the presentation of a video for 4 s only if the participant correctly reacted during a change trial.

The experiment consisted of 3 phases, a demonstration phase and 2 test phases. The phases were presented in a fixed sequence: Participants were required to reach criterion on one phase before moving to the next. In the demonstration phase and in the first test phase, the stimuli were from the Fast AM Condition. In the second test phase, the stimuli were from the Slow AM condition.

The purpose of the demonstration phase was to familiarize the participant with the association between the reinforcer (i.e., video) and the target sounds. In this phase, the probability of a change trial was 0.80, and the reinforcer was activated after every change trial regardless of the participant's response. The demonstration phase, which lasted a maximum of 12 trials, ended as soon as the participant had responded correctly to 1 change trial (hence had reacted to the category change) and to 1 no-change trial (hence had not reacted to the lack of category change).

In the following test phases, change and no-change trials were presented in random order, with the probability of change and no-change trials being 0.5. The criterion to end the test phase was evaluated on sliding windows of 10 trials, and corresponded to responding correctly on at least 4 out of 5 change trials and at least 4 out of 5 no-change trials, which corresponds to a hit rate of more than 80% and a false alarm rate of <20%. If the criterion was not reached within a maximum number of trials, the session ended and a new session was started after a short break. If the participants could not reach the criterion within a maximum number of sessions, the participant was judged to be unable to complete the phase. In the Fast AM test phase, the maximum number of trials was 40 and the maximum number of sessions was 4; in the Slow AM test phase, the maximum number of trials was 32 and maximum number of sessions was 3 to minimize the effect of training (Cabrera and Werner, 2017).

To accommodate the anticipated difficulty in the Slow AM condition, a reminder procedure, similar to the one used by Clarkson and Clifton (1995), was used to assess whether an infant failure was due to factors such as sleepiness or boredom rather than an inability to discriminate. Figure 2 outlines the different scenarios in which the experiment could play out. If a participant responded incorrectly on three consecutive trials in the Slow AM test phase (responding to no-change trials or not responding to change trials), stimuli were presented from the previously completed (and thus succeeded) Fast AM Condition. Up to 10 trials of such "reminder" trials were presented, and if the participant responded correctly on three out of four consecutive trials, the participant returned to the Slow AM phase. If this criterion was not met, the session was discontinued, and infants were given a short break or returned on another day for a new session. Additionally, we ensured that infants were given frequent breaks during the testing process, whenever they appeared to need them, allowing them time to play inside the testing booth, feed, or crawl around as needed. If a participant reached criterion in the Fast AM Condition and reached criterion in three reminder periods without reaching criterion in the Slow AM Condition in three sessions, the participant was judged to be unable to discriminate the phonetic contrast based on the slow AM cues. Because the infant could still perform the discrimination in the Fast AM reminder trials, we could then conclude that the infant's failure in the Slow AM condition did not result from fatigue or loss of interest. As data collection is in line with the infants' individual rhythms, and that infants are more active in such a procedure compared to passive looking time recording procedures, we observed low attrition rates (see Table 2), which are comparable with previous studies using this technique (Olsho et al., 1987; Cabrera and Werner, 2017). It is important to note that, infant testing was completed in one or two visits (lasting around 60 minutes each) on 2 separate days within a 2-week period , which helped to adapt to the infants states. Adult testing was completed in one visit lasting around 60 minutes.

The main dependent variable analyzed was the proportion of participants who reach success criterion in each phonetic category in each test phase (Fast versus Slow AM). The probability for participants to succeed in the Fast AM condition and then in the Slow AM condition was compared across age groups (6 months versus 10 months versus adults) and (1) phonetic conditions (Vowel versus Consonant) and 2) phonetic features (Vowel Place versus Vowel Height; Consonant Place versus Consonant Voicing). In order to take into account the fact that participants who failed in the Fast AM condition were not tested in the Slow AM condition, we used a modified logistic regression approach called survival analysis to compare the proportion of participants reaching criterion ("survival") according to age and phonetic condition (or phonetic features). This analysis calculates a survival function for each group representing the cumulative probability that a participant who started the experiment reached criterion in each vocoder condition. The log-rank test for equality, a nonparametric statistic, was used to compare the survival functions for infants and adults, and for vowels and consonants. When a significant difference was found between functions, it meant that either the slope of the function was different between groups (i.e., groups were affected differently by the vocoders if the slopes were not parallel), or the proportion of participants succeeding in either or both vocoder conditions was different between groups (and this was further assessed using χ^2 tests). Survival function is a non-parametric test well-suited for analyzing smaller sample sizes and effective in discerning temporal patterns, particularly in the non-independent Time 1 and Time 2 conditions of the Fast and Slow tests. Additionally, as a secondary analysis, we used

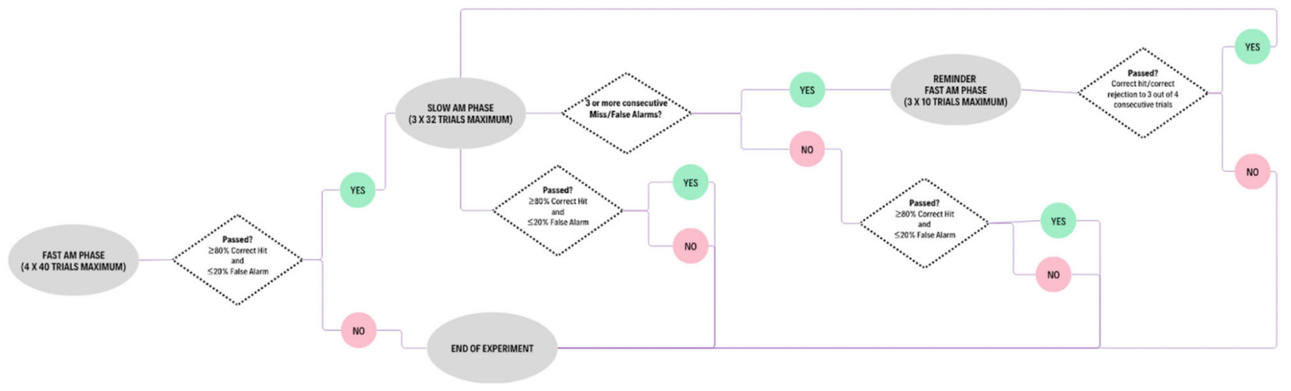

FAST AM PHASE (4 X 40 TRIALS MAXIMUM)
Passed? ≥80% Correct Hit and ≤20% False Alarm
YES
NO
SLOW AM PHASE (3 X 32 TRIALS MAXIMUM)
3 or more consecutive Miss/False Alarms?
Passed? ≥80% Correct Hit and ≤20% False Alarm
YES
NO
END OF EXPERIMENT
YES
NO
REMINDER FAST AM PHASE (3 X 10 TRIALS MAXIMUM)
Passed? Correct hit/correct rejection to 3 out of 4 consecutive trials
YES
NO
Passed? ≥80% Correct Hit and ≤20% False Alarm
YES
NO

FIGURE 2
Schematic representation of experimental procedure.

TABLE 2 Breakdown of participants' success rates across different age groups for the different phonetic conditions.

Age group	Phoneme	Category	Fast AM	Slow AM	Change in proportion (Slope)
6-month-olds	Vowels	Overall	16 of 20 (80%)	9 of 16 (56%)	-0.238
		Place	9 of 10 (90%)	4 of 9 (44%)	-0.456
		Height	7 of 10 (70%)	5 of 7 (71%)	0.014
	Consonants	Overall	18 of 20 (90%)	14 of 18 (78%)	-0.122
		Place	8 of 10 (80%)	5 of 8 (63%)	-0.175
		Voicing	10 of 10 (100%)	9 of 10 (90%)	-0.1
10-month-olds	Vowels	Overall	14 of 20 (70%)	10 of 14 (72%)	0.014
		Place	8 of 10 (80%)	6 of 8 (75%)	-0.05
		Height	6 of 10 (60%)	4 of 6 (67%)	0.067
	Consonants	Overall	15 of 20 (75%)	8 of 13 (53%)	-0.217
		Place	5 of 10 (50%)	0 of 3 (0%)	-0.5
		Voicing	10 of 10 (100%)	8 of 10 (80%)	-0.2
Adults	Vowels	Overall	All 20 (100%)	All 20 (100%)	0
		Place	10 of 10 (100%)	10 of 10 (100%)	0
		Height	10 of 10 (100%)	10 of 10 (100%)	0
	Consonants	Overall	All 20 (100%)	18 of 20 (90%)	-0.1
		Place	10 of 10 (100%)	8 of 10 (80%)	-0.2
		Voicing	10 of 10 (100%)	10 of 10 (100%)	0

logistic regression to compare the proportion of infants succeeding in the Fast condition across conditions and groups, as well as the proportion succeeding in the Slow condition. It is important to emphasize that these analyses serve as *post-hoc* analyses to provide a better understanding of the differences highlighted by the primary survival function analyses, but given the small number of participants in these exploratory analyses (max $N = 10$), results can only be seen as indications to be further tested in future research.

Additionally, for each age group and phonetic conditions, we compared the number of trials needed to achieve success in each vocoder condition, a metric often used as a measure of processing difficulty in infant studies (Clarkson et al., 1988; Clarkson and Clifton, 1995; Lau and Werner, 2012), using linear models (LM).These analyses thus explored whether infants and adults were able to detect (1) Vowel (Place and Height) and (2) Consonant (Place and Voicing) feature categories when FM is reduced and also when faster AM is reduced.

3 Results

3.1 Survival function analyses comparing vowels vs. consonants

The proportion of participants who reached the 80-20 criterion ($d' = 1$), considered as a measure of detection success, is represented in Figure 3 for each age group and in both vocoder conditions for consonant and vowel categories. See Table 2 for a summary of survival functions for all conditions.

The probability for participants to succeed in the Fast AM condition and then in the Slow AM condition was compared across age groups and phonetic conditions using survival analyses. When comparing all six survival functions defined by Age and Phonetic condition (illustrated in Figure 3), the functions were significantly different [$\chi^2(5) = 38.60, p < 0.001$]. Follow-up analyses were conducted first comparing the functions for Vowels versus Consonants within each Age group. A marginally significant difference was observed at 6 months [$\chi^2(1) = 3.10$, $p = 0.08$], and no significant difference was observed in the other two age groups [10-month-olds: $\chi^2(1) = 0.10$, $p = 0.80$; adults: $\chi^2(1) = 2.10$, $p = 0.20$], suggesting that the detection of vowel or consonant change was affected similarly by vocoding in the three groups. In other words, a similar proportion of participants reached criterion in the Fast and then in the Slow AM condition when exposed to vowel or to consonant change.

The next analyses investigated age effects within each Phonetic Condition (Vowels or Consonants). For Vowels, the functions were not significantly different between 6-month-olds and 10-month-olds [$\chi^2(1) = 0.10$, $p = 0.80$]. However, there was a significant difference between 6-month-olds and adults [$\chi^2(1) = 18.70$, $p < 0.001$] because fewer 6-month-olds reached criterion in both conditions [Fast AM: $\chi^2(1) = 4.44$, $p = 0.035$; Slow AM: $\chi^2(1) = 10.86$, $p = 0.001$] compared to adults. Moreover, while adults performed at ceiling, 6-month-olds showed a decrease from 80% to 56% between the Fast and Slow conditions when detecting vowel changes. There was also a significant difference in survival functions between 10-month-olds and adults [$\chi^2(1) = 19.20$, $p < 0.001$], again because fewer 10-month-olds reached criterion in both conditions [Fast AM: $\chi^2(1) = 7.06$, $p = 0.008$; Slow AM: $\chi^2(1) = 6.48$, $p = 0.011$] compared to adults, but 10-month-olds showed similar proportions of success in both vocoder conditions (70% at Fast; 72% at Slow).

For Consonants, functions showed a significant difference between 6-month-olds and 10-month-olds [$\chi^2(1) = 4.90$, $p = 0.03$] and further comparisons between the distribution of succeeding participants showed no significant difference in the Fast AM condition [$\chi^2(1) = 1.56$, $p = 0.21$] and a trend for fewer 10-month-olds succeeding in the Slow AM condition compared to 6-month-olds [$\chi^2(1) = 2.20$, $p = 0.14$]. While 6-month-olds showed a decrease from 90% to 78%, 10-month-olds showed a decrease from 75% to 53% suggesting that the detection of consonant change was more affected by vocoding at 10 months than at 6 months (see Figure 3). A significant difference was also found between 6-month-olds and adults [$\chi^2(1) = 18.70$, $p < 0.001$], related to the fact that overall fewer 6-month-olds reached criterion in both conditions [Fast AM: $\chi^2(1) = 4.44$, $p = 0.035$; Slow AM: $\chi^2(1)$=10.86, $p = 0.001$]. Moreover, a significant difference is observed between 10-month-olds and adults [$\chi^2(1) = 16.00$, $p = < 0.001$], and fewer 10-month-olds reached criterion in both conditions [Fast AM: $\chi^2(1) = 5.71$, p= 0.017; Slow AM: $\chi^2(1) = 6.03$, $p = 0.014$] compared to adults.

In summary, no difference was observed between 6- and 10-month-olds for vowel change detection, but 10-month-olds were more affected by vocoding than 6-month-olds for consonant change detection. For both consonant and vowel change detection, fewer 6- and 10-month-olds succeeded compared to adults in both vocoder conditions.

3.2 Exploratory survival function analyses comparing subcategory of vowels and consonants

Next, as an exploratory analysis, given the limited sample size of only 10 participants per subgroup, we compared survival functions for vowel features (place versus height) and consonant features (place versus voicing) to assess whether the different phonetic categories rely differently upon temporal cues as a function of age. These functions are represented in Figure 4.

3.2.1 Vowels

When comparing all six survival functions defined by Age and Vowel Feature (3 ages x 2 features), a significant difference was found [$\chi^2(5) = 23.00$, $p < 0.001$]. To understand this difference, we first explored the impact of Features for each age group separately. No significant differences were found for the 6-month-olds [$\chi^2(1) = 0.10$, $p = 0.70$], the 10-month-olds [$\chi^2(1) = 1.50$, $p = 0.20$], or the adults [$\chi^2(1) = 0.00$, $p = 1.00$], suggesting that detection of vowel height and vowel place change in all age groups was affected similarly by vocoding (i.e., in each age group, a similar proportion of participants reached criterion in the Fast and then in the Slow AM condition when exposed to either vowel height or place change).

The next comparisons addressed differences between Age groups for each vowel feature. For vowel place, a main effect of Age was found [$\chi^2(2) = 8.80$, $p = 0.01$], and pairwise comparisons revealed no significant effect between the two infant groups [$\chi^2(1) = 0.10$, $p = 0.70$], but a significant difference between 6-month-olds and adults [$\chi^2(1) = 9.00$, $p = 0.003$], and between 10-month-olds and adults [$\chi^2(1) = 6.80$, $p = 0.009$]. The differences were related to significant lower proportions of 6-month-olds reaching criterion in the Slow AM condition compared to adults [Fast AM: $\chi^2(1) = 1.05$, $p = 0.305$; Slow AM: $\chi^2(1) = 7.54$, $p = 0.006$], and to marginally lower proportions of 10-month-olds reaching criterion in the Slow AM condition compared to adults [Fast AM: $\chi^2(1) = 2.22$, $p = 0.136$; Slow AM: $\chi^2(1) = 2.81$, $p = 0.093$]. Specifically, for 6-month-olds success rates decreased from 90% to 44% and for 10-month-olds performance decreased from 80% to 75% between Fast and Slow conditions.

For vowel height, age influenced the functions [$\chi^2(2) = 12.40$, $p = 0.002$], and subsequent pairwise comparisons revealed no significant difference between 6- and 10-month-olds [$\chi^2(1) = 0.40$, $p = 0.50$], but a significant difference between 6-month-olds and adults [$\chi^2(1) = 9.40$, $p = 0.002$] and between 10-month-olds and adults [$\chi^2(1) = 12.30$, $p < 0.001$]. These differences were related to marginally lower proportions of 6-month-olds reaching criterion in both conditions [Fast AM: $\chi^2(1) = 3.53$, $p = 0.06$; Slow AM: $\chi^2(1) = 3.24$, $p = 0.070$] compared to adults. Similar but significant effects were observed between 10-month-olds and adults [Fast AM:

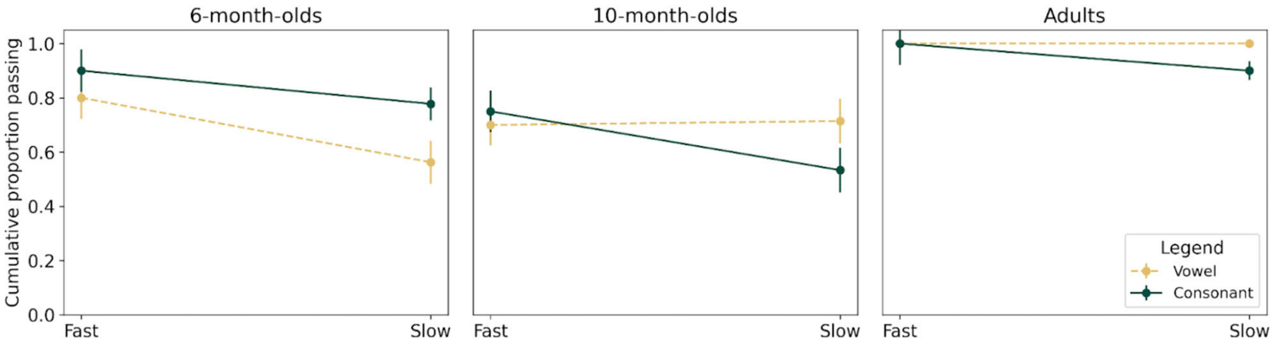


FIGURE 3
Overall survival plots between Fast and Slow AM conditions (on the x-axis) for Vowel (dashed yellow lines) and Consonant features (solid dark green lines) for 6-month-olds, 10-month-olds, and adults (in each panel). Error bars are standard errors from Kaplan-Meier analysis.

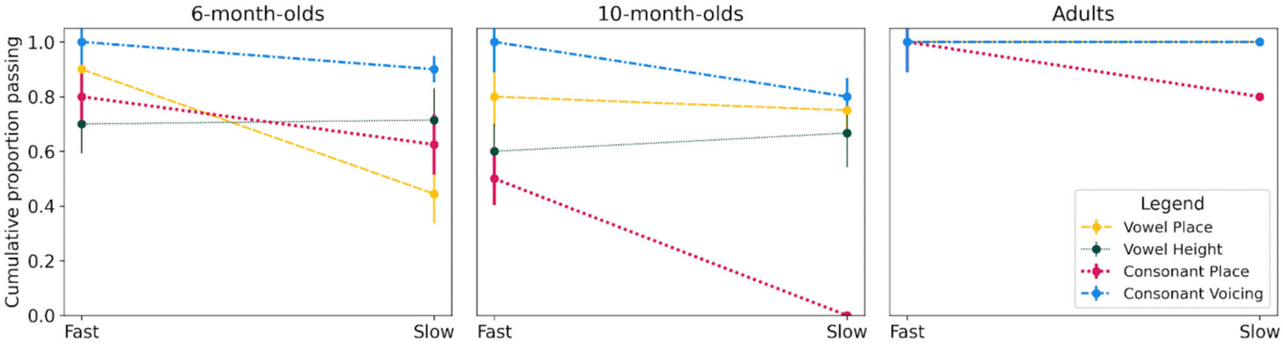


FIGURE 4
Survival plots between Fast and Slow AM conditions (on the x-axis) for Vowel (Place and Height, dashed yellow lines vs dotted dark green lines, respectively) and Consonant features (Place and Voicing, dotted pink lines vs dash-dotted blue lines, respectively) for 6-month-olds, 10-month-olds, and adults (in each panel). Error bars are standard errors from Kaplan-Meier analysis.

$\chi^2(1) = 5.00$, $p = 0.03$; Slow AM: $\chi^2(1) = 3.81$, $p = 0.05$]. Here, again, adults perform at ceiling while both infant groups show an overall lower success rate, albeit similarly affected in the Fast (6 months: 70%; 10 months: 60%) and Slow (6 months: 71%; 10 months: 67%) conditions.

In summary, no difference was observed between vowel height and vowel place in any group. However, fewer 6- and 10-month-olds succeeded in detecting vowel place and vowel height change under the current vocoder conditions compared to adults. For vowel place, a lower proportion of infants succeeded the detection compared to adults in the Slow AM condition, while for vowel height, lower proportions were observed in both vocoder conditions.

3.2.2 Consonants

When comparing the six survival functions defined by Age and Consonant Feature (3 ages × 2 features), a significant difference was found [$\chi^2(5) = 48.00$, $p < 0.001$]. To understand this difference, subsequent comparisons assessed the impact of Features for each age group separately. There were significant differences between the survival functions for place and voicing for 6-month-olds [$\chi^2(1) = 5.70$, $p = 0.02$] and for 10-month-olds [$\chi^2(1) = 17.70$, $p < 0.001$]. A marginal difference emerged for adults [$\chi^2(1) = 2.10$, $p = 0.10$] because two participants only failed to detect place change in the Slow AM condition. At 6 months, the comparison of participants reaching the criterion between voicing and place contrasts did not show statistical significance in either the Fast [$\chi^2(1) = 2.22$, $p = 0.136$] or Slow AM conditions [$\chi^2(1) = 1.94$, $p = 0.163$]. This suggests that the observed differences in survival functions for voicing and place contrasts are not statistically significant. However, there is an overall difference in the proportion of participants meeting the criterion, with a higher proportion for voicing (90%) compared to place (63%) in both conditions. At 10 months, this difference is characterized by a lower proportion of participants able to detect the place change compared to the voicing change in both vocoder conditions [Fast: $\chi^2(1) = 6.67$, $p = 0.01$; Slow: $\chi^2(1) = 8.57$, $p = 0.003$] and by a steeper decrease in proportion of participants reaching criterion from Fast to Slow AM conditions for place than for voicing. For place, 10-month-olds showed a decrease between the two vocoder conditions from 50% to 0% whereas for voicing the decrease was much smaller from 100% to 80%.

Next, we addressed differences between Age groups for each consonant feature. For place, a significant Age effect was observed on the survival functions, [$\chi^2(2) = 18.50$, $p < 0.001$], and

2-by-2 comparisons revealed significant differences between 6-month-olds and adults [$\chi^2(2) = 3.70$, $p = 0.05$] and 10-month-olds and adults [$\chi^2(2) = 17.70$, $p < 0.001$], with no significant difference between the two infant groups [$\chi^2(1) = 0.40$, $p = 0.50$]. These differences were related to lower proportions of 10-month-olds reaching criterion in both vocoder conditions [Fast AM: $\chi^2(1) = 6.67$, $p = 0.01$; Slow AM: $\chi^2(1) = 8.57$, $p = 0.003$] compared to adults. Moreover, 10-month-olds showed a stark decrease from 50% to 0% success rates between the two conditions, while adults went from a 100% success rate to an 80% one. Similar but non-significant trends were found in the Fast AM condition between 6-month-olds and adults [Fast AM: $\chi^2(1) = 2.22$, $p = 0.14$; Slow AM: $\chi^2(1) = 0.68$, $p = 0.41$], suggesting an overall effect of less 6-month-olds reaching criterion (71.5% compared to adults (90%). For voicing, no significant effect of Age was found [$\chi^2(2) = 2.10$, $p = 0.30$].

In summary, a difference in the proportion of participants able to succeed detection between the Fast AM and the Slow AM conditions was observed between voicing and place for both infant groups only. Moreover, while no difference was observed between either infant groups or adults for voicing detection, fewer 6- and 10-month-old infants were able to reach success in both vocoder conditions for place compared to adults, and 10-month-olds showed a strong decrease between the two vocoder conditions.

3.3 Linear Models comparing the numbers of trials to reach criterion

In order to further understand whether task difficulty was affected by Vocoder condition (Fast AM versus Slow AM), Phonetic Condition (Vowels vs Consonants) or Phonetic Feature (vowel place versus vowel height versus consonant place versus consonant voicing) and Age (6 months vs 10 months), Linear models were used to analyze the average number of trials needed to succeed the task (see Figure 5). Adults' data were analyzed in individual models to assess the effect of Phonetic Conditions or Phonetic Features. All analyses were conducted in R (version 4.3.1, R Core Team, 2019). We fitted linear models using the lm function.

In the Fast AM test phase, the maximum number of trials was set at 40, with a limit of 4 sessions. Conversely, in the Slow AM test phase, we limited the number to 32 trials and a maximum of 3 sessions. In the following analysis, the average number of trials required to achieve the success criterion thus corresponds to the average of the total number of trials over the sessions required by each infant in each phase. This analysis was conducted first for the Fast AM condition, followed by the Slow AM condition.

For infants, we used the following Linear model to analyze the average number of trials needed to achieve the success criterion, first in the Fast AM, then in the Slow AM condition:

$$\text{Average Number of Trials} \\ \sim \text{Age} * \text{Phonetic Condition (Vowel/Consonant)}$$

For the Fast AM condition, the ANOVA failed to find significant effects of Age [$F_{(1,59)} = 0.61, p = 0.436$], Phonetic Condition [$F_{(1,59)} = 0.08, p = 0.774$], or the Age x Phonetic Condition interaction [$F_{(1,59)} = 0.001, p = 0.975$]. Likewise, for the Slow AM condition, the ANOVA failed to find significant effects of Age [$F_{(1,37)} = 1.91, p = 0.176$], Phonetic Condition [$F_{(1,37)} = 0.42, p = 0.521$] or the Age x Phonetic Condition interaction [$F(_{1,37}) = 0.27, p = 0.605$].

For Adults, we used the following Linear Models to evaluate the average number of trials needed to pass the Fast and Slow conditions:

$$\text{Average Number of Trials} \\ \sim \text{Phonetic Condition (Vowel/Consonant)}$$

For the Fast condition, the ANOVA failed to find a significant effect of Phonetic Condition [$F_{(1,38)} = 0.01, p = 0.926$]. For the Slow condition, the ANOVA revealed a significant effect of Phonetic Condition [$F_{(1,36)} = 7.09, p = 0.012$], and *post-hoc* analyses revealed that it took more trials to achieve success for consonants (29 trials in average) than for vowels (17 trials in average), which indicates a greater level of difficulty for consonants than vowels. Given the significant effect of Phonetic Condition, follow up analyses were conducted comparing average number of trials within the vowel and consonant categories, using the following LM:

$$\text{Average Number of Trials} \sim \text{Phonetic Feature}$$

For vowels, the ANOVA found a marginally significant effect of Phonetic Feature [$F_{(1,18)} = 3.30, p = 0.086$].

4 Discussion

The present study explores the reliance of 6-month-olds, 10-month-olds and adults upon spectro-temporal modulations of speech when categorizing consonant and vowel contrasts based on different phonetic features (for vowels: place and height; for consonants: place and voicing). Results show that 6-month-olds, 10-month-olds and adults are able to use AM cues, and even the slowest AM cues only (< 8 Hz) for both vowel and consonant categorization. Indeed, in the Fast AM condition, in which FM cues were replaced but original AM cues were preserved in a large number (N = 32) of spectral bands, the overall proportion of participants succeeding the detection of vowels and consonants averaged together, was 85% in 6-month-olds, 73% in 10-month-olds, and 100% in adults. This first result establishes that at the three ages, the participants could successfully detect the vowel/consonant changes based solely on AM cues. It suggests that, in quiet, FM is not necessary for phonetic categorization for the majority of 6-month-olds, 10-month-olds, or adults. Moreover, among participants who succeeded in the Fast AM condition, the overall success rates in the Slow AM condition, in which only the slowest AM cues (<8 Hz) were preserved, were 67% in 6-month-olds, 63% in 10-month-olds, and 95% in adults. This again establishes that

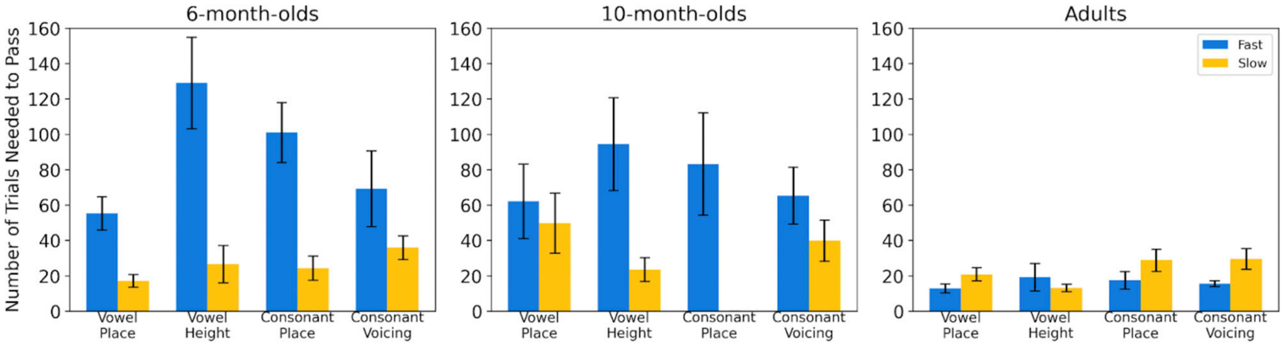


FIGURE 5
Average number of trials (and standard errors) needed to succeed Fast (blue bars) and Slow (yellow bars) phases in each phonetic feature condition (x-axis) for 6-month-olds, 10-month-olds and adults (in each panel).

at the three ages, most of the participants who could successfully detect the vowel/consonant changes using only AM cues could also detect those changes based on Slow AM cues only.

Our adult results show that although adults required more trials to reach success criterion when detecting consonant compared to vowel changes in the Slow AM condition, they were at ceiling in both vocoder conditions for all phonetic feature contrasts tested. This indicates that FM cues are not necessary for phoneme processing in adults, and that they are able to rely solely on slow AM cues (< 8 Hz). This pattern is similar to what was found in previous studies with adult listeners showing near perfect identification or discrimination scores on the basis of slow AM cues in quiet (Drullman et al., 1994a,b; Cabrera and Werner, 2017). Moreover, the higher number of trials required when detecting consonants based on the slowest AM cues is also consistent with previous studies showing a stronger impact of temporal reduction when processing consonants compared to vowels (Xu et al., 2005).

Our infant results suggest that slow AM cues (<8 Hz) provide enough information for most infants to successfully process the phonetic contrasts used in our task. They are consistent with previous infant experiments showing that 3- and 6-month-olds are able to discriminate consonant place and voicing on the basis of the original AM or slow AM cues (Bertoncini et al., 2011; Cabrera et al., 2013, 2015a; Cabrera and Werner, 2017). Crucially though, they add new data for vowel processing, as only one previous vocoder study had been conducted testing discrimination of a very large vowel contrast (/a/ versus /i/, Warner-Czyz et al., 2014). Here we demonstrate, for the first time, that both 6- and 10-month-olds can process vowel place and vowel height in conditions of reduced FM and AM cues. This aligns with evidence that young infants possess auditory mechanisms with relatively mature auditory temporal and spectral resolution (Folsom and Wynne, 1987; Spetner and Olsho, 1990; Levi and Werner, 1996).

4.1 Contributions of temporal cues to the categorization of vowels vs. consonants

Importantly, no overall difference was observed between the consonant and vowel conditions (that is, when the two phonetic feature conditions are averaged) in any age group. In other words, the proportion of participants succeeding the task was not different when they had to detect a change in vowel or a change in consonant. This finding suggests that the ability in detecting these changes are affected similarly by temporal degradation. However, differences between age groups appeared, indicating that with age the reliance upon temporal modulations may differ when processing native vowels and consonants. For vowels, while the 6- and 10-month-old groups did not differ from one another for overall detection in the vocoded conditions, they both independently differed from adults. Contrary to our expectations, we did not observe any significant difference in the reliance upon FM and faster AM cues between 6 and 10 months of age for vowel categorization. However, the number of participants succeeding in detecting the vowel contrasts was significantly lower for both groups of infants compared to adults in both vocoder conditions. More precisely, both 6- and 10-month-olds are more affected by FM degradation compared to adults and are also more affected by faster AM degradation (if they succeeded the Fast AM condition) compared to adults. Six-month-olds also displayed a specific result, that is, they showed a more important decrease of success rate in the Slow AM condition compared to adults. Altogether, these findings reveal that both FM and fast AM cues play a critical role in vowel categorization for both infant age groups, likely due to the role of faster temporal cues in conveying fine spectro-temporal details that are probably important for vocalic processing (Rosen, 1992). The fact that significantly fewer 6-month-olds succeeded the task compared to adults when the faster AM cues are reduced may also suggest that these temporal modulations are important for them to successfully detect vocalic changes.

For consonants, an overall difference was observed between the two infant groups, with 10-month-olds showing a stronger impact of vocoding on consonant detection compared to 6-month-olds, and a significantly greater decline in detection success rates between the Fast AM and Slow AM conditions compared to 6-months (75% and 53% vs. 90% and 78%, respectively). Thus, the degradation of faster AM cues (>8 Hz) appears to strongly affect consonant processing at 10 months of age. This difference in overall consonant processing is contrary to our hypothesis as we expected any difference between the effect of temporal degradation on consonants to be less pronounced in 10-month-olds than in 6-month-olds, the former being more advanced in their speech perceptual attunement to their native consonants. However, as will be further discussed later, this result can be further nuanced based on consonant feature category. It is then possible that while 10-month-olds become more attuned to the consonants of their native language, they are also more impeded by their linguistic experience to rely on the residual AM information. This is in line with previous cross-linguistic studies using vocoders with infants and adults showing that native listeners are more impaired by reduction of fine spectro-temporal cues compared to non-native listeners for consonant and tone processing (Cabrera et al., 2014, 2015a). Moreover, both infant groups independently were more affected by vocoding compared to the adult group, and this, in both vocoder conditions. Importantly, adults showed more robust detection of consonants in degraded conditions compared to 10-month-olds, who are supposed to have started attuning to the consonants of their native language. This different perceptual weight on temporal modulations for consonant change detection suggests that further changes in acoustic processing take place after the onset of perceptual attunement around 10 months. This is congruent with some studies showing that phonological categorization continues to develop until 12 years of age, and that children do not rely on the same acoustic (i.e., spectral or VOT) cues as adults to distinguish between native phonemes (e.g., Lehman and Sharf, 1989; Hazan and Barrett, 2000; Mayo et al., 2003; Nittrouer, 2004; Nittrouer and Lowenstein, 2007).

4.2 Contributions of temporal cues to the categorization of phonetic features

In the present experimental design, we also manipulated the phonetic feature to be detected within the vowel and the consonant condition: vowel place versus vowel height, and consonant place versus consonant voicing. In our task, participants heard a string of syllables that varied in consonants and vowels but shared one phonetic feature (for example, back vowels: pu, bu, tu, du, po, bo, to, do), and they were required to react to a new syllable corresponding to a feature change (for example, front vowels: py, by, ty, dy, pø, bø, tø, dø). To perform this task, participants could have discriminated the eight new syllables from the eight syllables presented as background. Alternatively, they could have categorized the syllables according to phonetic features, and discriminate categories of syllables based on the contrasting feature (in the example above, vowel place). If so, albeit exploratory, our findings would add to a small number of studies showing that phonetic features appear to be used in infant processing. In both vocoder conditions, all age groups were able to successfully detect a change in phonetic features on the basis of AM cues. Around 9 to 10 months, infants can form generalizations across different speech segments on the basis of place of articulation (Seidl and Buckley, 2005), they can learn constraints between non-adjacent consonants, but only when the consonants share a phonetic feature (Saffran and Thiessen, 2003) and their phonotactic knowledge appears constrained by phonetic features (Gonzalez-Gomez and Nazzi, 2015). Moreover, between 4 and 7 months, infants' acquisition and generalization of phonological constraints on consonant categories becomes constrained by the fact that those categories are defined by a single phonetic feature (Cristià and Seidl, 2008; Cristià et al., 2011). Our exploratory findings would add another piece of evidence in support of a role of phonetic features in early language processing and acquisition, providing the first piece of evidence of an early use also of vowel features, and that the cues needed to process these features are contained in the AM information.

These findings also reveal that the ability to detect phonetic feature changes was affected differently by vocoder and age of the listeners. The detection of vowel place and vowel height was affected similarly by vocoding in all age groups, meaning that one vocalic feature was not easier to detect than the other when FM or faster AM cues are degraded. However, age differences occurred within each phonetic feature category between infant groups and adults, while no difference was observed between 6- and 10-month-olds. For vowel place detection, 6-month-olds were more affected by the reduction of fast AM cues compared to adults, while 10-month-olds were overall worse than adults in both vocoder conditions (with no further decrease in the proportion of participants succeeding the task in the Slow AM condition). For vowel height, both 6-month-olds and 10-month-olds showed lower success rates compared to adults in both vocoder conditions. These findings may suggest that at 6 months infants require faster AM cues (> 8 Hz) to efficiently detect changes in vowel features. Importantly, the vocoders used in the current experiment did not drastically affect the original formants of the vowels, as the spectral resolution of the vocoded signals was pretty high including 32 spectral bands. Furthermore, in adults, it has been shown that vowel perception is more affected by spectral degradation than by temporal degradation (Xu and Pfingst, 2003). In our results, infants are more sensitive to a degradation of the temporal modulations of vowels compared to adults, suggesting a stronger perceptual weight on relatively fast temporal modulations in infancy even for vowel processing. FM cues and faster AM cues convey information about the spectrum of speech and thus, about the formant pattern (Rosen, 1992). It is possible that infants from 6 to 10 months of age rely more strongly on these temporal cues compared to adults, and are more sensitive to subtle modification of the speech spectrum, like older children have been shown to be more sensitive to the dynamic spectral structure in vowel identification (Nittrouer and Lowenstein, 2007). For consonant features, a different pattern was observed, that is, both infant groups were less affected by temporal cue reduction for voicing contrasts compared to place of articulation contrasts. No age difference was observed in the proportion of infants and adults succeeding the detection of voicing change when FM cues were degraded or when faster AM cues were degraded. One result of note is that in the Fast AM condition, all age groups achieved a 100% success rate for voicing contrasts. On the other hand, the detection of place change for consonants was significantly different with age and as a function of the vocoder condition. Significantly less infants, at both 6 and 10 months of age, compared to adults were able to complete the task in the vocoded conditions. Six-month-olds were overall more affected than adults when FM cues and faster AM cues were reduced. Ten-month-olds were strongly affected by reduction of fast AM cues compared to adults as none of the infants tested were able to detect the place change in the Slow AM condition. These findings align with prior adult research which indicated that the identification of place is notably vulnerable to spectro-temporal degradation. For instance, Shannon et al. (1995) highlighted that, unlike other phonetic features, the identification of place suffered in scenarios with limited spectral bands where FM cues were degraded. Drullman et al. (1994a,b) also observed that place of articulation for stop consonants was difficult to identify by adult listeners when reducing FM and faster AM cues. Again, as FM cues convey information about the spectrum, it has been suggested that place is particularly affected by such degradation compared to voicing (Rosen, 1992). In the present task, adults were not impacted by this FM degradation even for place, which reveals that even though the task was implicit (i.e., without direct instruction about what the participants should be attending in the signal), adult listeners were extremely good at detecting any phonetic change even the ones usually more difficult to identify.

In sum, the developmental differences between infants and adults in success rates between the Fast and the Slow AM conditions reveal different reliance upon temporal cues for phonetic perception between infancy and adulthood. To some extent, the present results are consistent with previous behavioral studies showing that younger infants more strongly rely on fast AM (> 0.8Hz) compared to adults. However, Cabrera and Werner (2017) using similar methods and vocoder conditions with 3-month-old infants and young adults showed that less than half of the infants differentiated between consonants (contrasting on either voicing or place) when only slow AM cues were preserved (i.e., Slow AM condition). In the current study, the filtering of faster AM cues did not impact as drastically the success rate of 6-month-olds for consonant detection. This discrepancy might be attributed to the age difference between the two studies, but also to the fact that in the previous study, the same vocalic context /a/ was used when presenting the vocoded syllables, and more consonant types were presented within the background (e.g., /b/, /p/, /d/, /t/, but also /k/, /g/), while in the current study, multiple vocalic contexts are presented /o/, /ø/, /u, /y/, and only two different consonants were presented in the background. Thus, it is possible that infants in the present design might have been able to leverage different mechanisms to compensate for the impact of acoustic temporal degradation on consonant discrimination, potentially due to vowel and consonant variability. Finer differences are then observed in the present study compared to the previous ones that did not find any difference in the detection of voicing and place in such vocoded conditions (Cabrera et al., 2015a; Cabrera and Werner, 2017). The present design may have "helped" infants to detect changes in voicing, not requiring FM or faster AM cues, while it may have impeded their detection of the place contrasts known to be more sensitive to any acoustic degradation (Miller and Nicely, 1955). Moreover, it is important to note that the present stimuli were from the French language where /p/ and /t/ are voiceless and unaspirated, and /b/ and /d/ are pre-voiced, whereas in English /p/ and /t/ are aspirated and /b/ and /d/ are partially voiced. These differences may also contribute to the discrepancy between the two studies. Finally, no significant differences were observed in the proportion of participants succeeding the detection of vowel changes between 6- and 10-month-old infants, but a significant difference emerged for consonant changes. This relates to the specific difficulty of 10-month-olds to detect the place change in the consonant condition when only the slowest AM cues are available. These findings thus suggest a similar weighting of FM and fast AM cues between 6 and 10 months of age for vowel categorization, perhaps because at these two ages infants have already started to attune to their native vowels, but a stronger reliance upon faster AM cues at 10 months for processing some native consonant features. This difference may relate to later onset of perceptual attunement for consonants than vowels. Future studies are required to determine whether differences in the perceptual weight of acoustic temporal cues for consonants are related to some other language milestones for instance lexical acquisition.

4.3 Conclusions

The present results indicate that infants, in comparison to adults, are more sensitive to the deterioration of FM and faster AM cues (> 8 Hz). They further indicate that infants between 6 and 10 months of age assign a similar perceptual weight to FM and fast AM cues when categorizing vowels, possibly because they already process vowels in a language-specific way (since they have started to attune to their native language vowels). However, at 10 months, there appears to be a stronger reliance for faster AM cues for consonants, especially when processing place of articulation. This difference between vowels and consonants might be linked to the later onset of infants' perceptual reorganization to consonant sounds, which begins between 6 and 10 months. Altogether, this study underscores the significant role of speech temporal cues in vowel and consonant categorization during infancy and suggests that the ability to rely solely on slow AM cues for phonetic categorization develops later in life.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Le comité d'éthique de la recherche (CER)-Paris Descartes. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

MH: Conceptualization, Formal analysis, Writing—original draft, Writing—review & editing. TN: Conceptualization, Writing—original draft, Writing—review & editing. LC:

Conceptualization, Formal analysis, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Agence Nationale de la Recherche [Grant Number: ANR-17-CE28-008 DESIN] awarded to LC, the ANR's French Investissements d'Avenir—Labex EFL Program under Grant No. ANR-10-LABX-0083 awarded to TN and MH.

Acknowledgments

We wish to thank Maxine dos Santos and Viviane Huet for help in recruiting the participants, and the undergraduate and master's students who served as assistants to the experimenter during testing. We also thank the families who agreed to participate in this study.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1321311/full#supplementary-material

References

Ardoint, M., and Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hear. Res.* 260, 89–95. doi: 10.1016/j.heares.2009.12.002

Bertoncini, J., Nazzi, T., Cabrera, L., and Lorenzi, C. (2011). Six-month-old infants discriminate voicing on the basis of temporal envelope cues (l). *J. Acoust. Soc. Am.* 129, 2761–2764. doi: 10.1121/1.3571424

Best, C. T., McRoberts, G. W., LaFleur, R., and Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behav. Dev.* 18, 339–350. doi: 10.1016/0163-6383(95)90022-5

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. *J. Exp. Psychol.* 14, 345. doi: 10.1037/0096-1523.14.3.345

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., and Nazzi, T. (2015). Call me alix, not elix: Vowels are more important than consonants in own-name recognition at 5 months. *Dev. Sci.* 18, 587–598. doi: 10.1111/desc.12242

Cabrera, L., Bertoncini, J., and Lorenzi, C. (2013). Perception of speech modulation cues by 6-month-old infants. *J. Speech, Lang. Hear. Res.* 56, 1733–1744. doi: 10.1044/1092-4388(2013/12-0169)

Cabrera, L., Lorenzi, C., and Bertoncini, J. (2015a). Infants discriminate voicing and place of articulation with reduced spectral and temporal modulation cues. *J. Speech, Lang. Hear. Res.* 58, 1033–1042. doi: 10.1044/2015_JSLHR-H-14-0121

Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoncini, J., and Lorenzi, C. (2014). The role of spectro-temporal fine structure cues in lexical-tone discrimination for french and mandarin listeners. *J. Acoust. Soc. Am.* 136, 877–882. doi: 10.1121/1.4887444

Cabrera, L., Tsao, F.-M., Liu, H.-M., Li, L.-Y., Hu, Y.-H., Lorenzi, C., et al. (2015b). The perception of speech modulation cues in lexical tones is guided by early language-specific experience. *Front. Psychol.* 6, 1290. doi: 10.3389/fpsyg.2015.01290

Cabrera, L., and Werner, L. (2017). Infants and adults use of temporal cues in consonant discrimination. *Ear Hear.* 35, 497–506. doi: 10.1097/AUD.0000000000000422

Clarkson, M. G., and Clifton, R. K. (1995). Infants pitch perception: inharmonic tonal complexes. *J. Acoust. Soc. Am.* 98, 1372–1379. doi: 10.1121/1.413473

Clarkson, M. G., Clifton, R. K., and Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. *Perc. Psychophys.* 43, 15–20. doi: 10.3758/BF03208968

Cristià, A., and Seidl, A. (2008). Is infants' learning of sound patterns constrained by phonological features? *Lang. Learn. Dev.* 4, 203–227. doi: 10.1080/15475440802143109

Cristià, A., Seidl, A., and Gerken, L. (2011). Learning classes of sounds in infancy, in *University of Pennsylvania Working Papers in Linguistics*, 9.

Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.* 95, 2670–2680. doi: 10.1121/1.409836

Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 1053–1064. doi: 10.1121/1.408467

Folsom, R. C., and Wynne, M. K. (1987). Auditory brain stem responses from human adults and infants: wave v tuning curves. *J. Acoust. Soc. Am.* 81, 412–417. doi: 10.1121/1.394906

Gilbert, G., and Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *J. Acoust. Soc. Am.* 119, 2438–2444. doi: 10.1121/1.2173522

Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Gnansia, D., Péan, V., Meyer, B., and Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *J. Acoust. Soc. Am.* 125, 4023–4033. doi: 10.1121/1.3126344

Gonzalez-Gomez, N., and Nazzi, T. (2015). Constraints on statistical computations at 10 months of age: the use of phonological features. *Dev. Sci.* 18, 864–876. doi: 10.1111/desc.12279

Hazan, V., and Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *J. Phon.* 28, 377–396. doi: 10.1006/jpho.2000.0121

Hopkins, K., and Moore, B. C. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J. Acoust. Soc. Am.* 125, 442–446. doi: 10.1121/1.3037233

Hopkins, K., Moore, B. C., and Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J. Acoust. Soc. Am.* 123, 1140–1153. doi: 10.1121/1.2824018

Hopkins, K., Moore, B. C., and Stone, M. A. (2010). The effects of the addition of low-level, low-noise on the intelligibility of sentences processed to remove temporal envelope information. *J. Acoust. Soc. Am.* 128, 2150–2161. doi: 10.1121/1.3478773

Kates, J. M. (2011). Spectro-temporal envelope changes caused by temporal fine structure modification. *J. Acoust. Soc. Am.* 129, 3981–3990. doi: 10.1121/1.3583552

Kong, Y.-Y., and Zeng, F.-G. (2006). Temporal and spectral cues in mandarin tone recognition. *J. Acoust. Soc. Am.* 120, 2830–2840. doi: 10.1121/1.2346009

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831. doi: 10.1038/nrn1533

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364

Lau, B. K., and Werner, L. A. (2012). Perception of missing fundamental pitch by 3-and 4-month-old human infants. *J. Acoust. Soc. Am.* 132, 3874–3882. doi: 10.1121/1.4763991

Lehman, M. E., and Sharf, D. J. (1989). Perception/production relationships in the development of the vowel duration cue to final consonant voicing. *J. Speech, Lang. Hear. Res.* 32, 803–815. doi: 10.1044/jshr.3204.803

Levi, E., and Werner, L. (1996). Amplitude modulation detection in infancy: Update on 3-month-olds. *Association for Research in Otolaryngology*, 142.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Nat. Acad. Sci.* 103, 18866–18869. doi: 10.1073/pnas.0607364103

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *J. Speech, Lang. Hear. Res* 46, 1184–1196. doi: 10.1044/1092-4388(2003/092)

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *J. Acoust. Soc. Am.* 27, 338–352. doi: 10.1121/1.1907526

Moore, B. C. (2003). Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. *Speech Commun.* 41, 81–91. doi: 10.1016/S0167-6393(02)00095-X

Moore, D. R. (2002). Auditory development and the role of experience. *Br. Med. Bull.* 63, 171–181. doi: 10.1093/bmb/63.1.171

Moore, J. K., and Linthicum Jr, F. H. (2007). The human auditory system: a timeline of development. *Int. J. Audiol.* 46, 460–478. doi: 10.1080/14992020701383019

Nazzi, T., Poltrock, S., and Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Curr. Dir. Psychol. Sci.* 25, 291–296. doi: 10.1177/0963721416655786

Nishibayashi, L.-L., and Nazzi, T. (2016). Vowels, then consonants: early bias switch in recognizing segmented word forms. *Cognition* 155, 188–203. doi: 10.1016/j.cognition.2016.07.003

Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J. Acoust. Soc. Am.* 115, 1777–1790. doi: 10.1121/1.1651192

Nittrouer, S., and Lowenstein, J. H. (2007). Children's weighting strategies for word-final stop voicing are not explained by auditory sensitivities. *J. Speech, Lang. Hear. Res.* 50, 58–73. doi: 10.1044/1092-4388(2007/005)

Olsho, L. W., Koch, E. G., Halpin, C. F., and Carter, E. A. (1987). An observer-based psychoacoustic procedure for use with young infants. *Dev. Psychol.* 23, 627. doi: 10.1037/0012-1649.23.5.627

Patterson, R. D. (1987). A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* 82, 1560–1586. doi: 10.1121/1.395146

Polka, L., and Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human perception and performance* 20, 421. doi: 10.1037/0096-1523.20.2.421

Poltrock, S., and Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *J. Exp. Child Psychol.* 131, 135–148. doi: 10.1016/j.jecp.2014.11.011

R Core Team. (2019). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 336, 367–373. doi: 10.1098/rstb.1992.0070

Saffran, J. R., and Thiessen, E. D. (2003). Pattern induction by infant language learners. *Dev. Psychol.* 39, 484. doi: 10.1037/0012-1649.39.3.484

Saffran, J. R., Werker, J. F., and Werner, L. A. (2006). "The infant's auditory world: hearing, speech, and the beginnings of language," in *Handbook of Child Psychology: Cognition, Perception, and Language*, eds D. Kuhn, R. S. Siegler, W. Damon, and R. M. Lerner (John Wiley & Sons, Inc.), 58–108.

Seidl, A., and Buckley, E. (2005). On the learning of arbitrary phonological rules. *Lang. Learn.. Dev.* 1, 289–316. doi: 10.1207/s15473341lld0103&amp;4_4

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303

Sheft, S., Ardoint, M., and Lorenzi, C. (2008). Speech identification based on temporal fine structure cues. *J. Acoust. Soc. Am.* 124, 562–575. doi: 10.1121/1.2918540

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90. doi: 10.1038/416087a

Spetner, N. B., and Olsho, L. W. (1990). Auditory frequency resolution in human infancy. *Child Dev.* 61, 632–652. doi: 10.1111/j.1467-8624.1990.tb02808.x

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Dev.* 47, 466–472. doi: 10.2307/1128803

Warner-Czyz, A. D., Houston, D. M., and Hynan, L. S. (2014). Vowel discrimination by hearing infants as a function of number of spectral channels. *J. Acoust. Soc. Am.* 135, 3017–3024. doi: 10.1121/1.4870700

Werker, J. F., Polka, L., and Pegg, J. E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Infant Child Dev.* 6, 171–178.

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

Werner, L. A. (1995). Observer-based approaches to human infant psychoacoustics. *Biomethods* 6, 135–135.

Xu, L., and Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (l). *J. Acoust. Soc. Am.* 114, 3024–3027. doi: 10.1121/1.1623786

Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., et al. (2005). Speech recognition with amplitude and frequency modulations. *Proc. Nat. Acad. Sci.* 102, 2293–2298. doi: 10.1073/pnas.04064 60102

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 March 2024
DOI 10.3389/fpsyg.2024.1367240

Infants' sensitivity to phonotactic regularities related to perceptually low-salient fricatives: a cross-linguistic study

Leonardo Piot[1,2]*, Thierry Nazzi[2] and Natalie Boll-Avetisyan[1]

[1]Department of Linguistics, Cognitive Sciences, University of Potsdam, Potsdam, Germany,
[2]Integrative Neuroscience and Cognition Center, CNRS & Université Paris Cité, Paris, France

Check for updates

OPEN ACCESS

EDITED BY
Marcela Pena,
Pontificia Universidad Católica de Chile, Chile

REVIEWED BY
Vsevolod Kapatsinski,
University of Oregon, United States
Marilyn May Vihman,
University of York, United Kingdom

*CORRESPONDENCE
Leonardo Piot
✉ leo1997p@gmail.com

RECEIVED 08 January 2024
ACCEPTED 22 February 2024
PUBLISHED 06 March 2024

CITATION
Piot L, Nazzi T and Boll-Avetisyan N (2024) Infants' sensitivity to phonotactic regularities related to perceptually low-salient fricatives: a cross-linguistic study.
Front. Psychol. 15:1367240.
doi: 10.3389/fpsyg.2024.1367240

Introduction: Infants' sensitivity to language-specific phonotactic regularities emerges between 6- and 9- months of age, and this sensitivity has been shown to impact other early processes such as wordform segmentation and word learning. However, the acquisition of phonotactic regularities involving perceptually low-salient phonemes (i.e., phoneme contrasts that are hard to discriminate at an early age), has rarely been studied and prior results show mixed findings. Here, we aimed to further assess infants' acquisition of such regularities, by focusing on the low-salient contrast of /s/- and /ʃ/-initial consonant clusters.

Methods: Using the headturn preference procedure, we assessed whether French- and German-learning 9-month-old infants are sensitive to language-specific regularities varying in frequency within and between the two languages (i.e., /st/ and /sp/ frequent in French, but infrequent in German, /ʃt/ and /ʃp/ frequent in German, but infrequent in French).

Results: French-learning infants preferred the frequent over the infrequent phonotactic regularities, but the results for the German-learning infants were less clear.

Discussion: These results suggest crosslinguistic acquisition patterns, although an exploratory direct comparison of the French- and German-learning groups was inconclusive, possibly linked to low statistical power to detect such differences. Nevertheless, our findings suggest that infants' early phonotactic sensitivities extend to regularities involving perceptually low-salient phoneme contrasts at 9 months, and highlight the importance of conducting cross-linguistic research on such language-specific processes.

KEYWORDS

phonotactics, salience, cross-linguistic, fricatives, infant language

Introduction

Infants' acquisition of their language-specific phonological system occurs rapidly. Indeed, already in their first year of life, infants start to specialize in speech sounds – or phonemes – that are used in their native language at the lexical level (e.g., Werker and Tees, 1984; Kuhl et al., 1992). Another important phonological property acquired during that period is the language-specific phonotactic system, that is, the legal and probabilistic positioning and sequencing of speech sounds within the words of a given language. Previous studies have shown that infants begin to acquire their language-specific phonotactic system in the first year (e.g., Friederici and Wessels, 1993; Jusczyk et al., 1994; Gonzalez-Gomez and Nazzi, 2012). However, the acquisition of regularities clearly contrasting in phonotactic frequencies but composed of perceptually low-salient phonemes has rarely been studied in infancy, and never with a cross-linguistic approach that makes it possible to ascertain that the findings result from the acquisition of language-specific properties rather than some intrinsic characteristics of the words presented to the infants. Indeed, only two studies investigated early phonotactic sensitivities to low-salient fricative patterns, and found contrasting results: French-learning infants were sensitive to frequent regularities of fricatives (Gonzalez-Gomez and Nazzi, 2015), whereas the evidence is more mixed for English-learning infants (Henrikson et al., 2020). Given these prior results, more research is needed to understand whether language-specific regularities involving low-salient phonemes are acquired early in development, and whether this depends on the infants' native language. This study aims to further investigate infants' acquisition of their language-specific phonotactic system from a cross-linguistic perspective by asking if German-and French-learning 9-month-old infants have already acquired perceptually low-salient regularities of fricative-plosive word-initial clusters.

Friederici and Wessels (1993) and Jusczyk et al. (1994) set the stage for the investigation of early phonotactic acquisition. Using the headturn preference procedure (HPP, Kemler Nelson et al., 1995) to investigate language-specific phonotactic sensitivities within one language, they demonstrated that sensitivity to language-specific phonotactic patterns emerges between 6 and 9 months: Friederici and Wessels (1993) tested Dutch-learning infants' preferences for nonwords including legal (e.g., /bref/, /murt/) versus illegal (e.g., /febr/, /rtum/) consonant clusters (henceforth CCs) in Dutch. Phonotactically legal CCs were attested and typical in their specific positions within Dutch words, whereas illegal CCs were clusters that could not appear in their specific positions within Dutch words. Nine-month-olds preferred listening to the nonwords containing legal CCs at onsets and offsets, whereas 4-and 6-month-olds showed no listening preference. In parallel, Jusczyk et al. (1994) investigated early sensitivity to phonotactic probabilities in English, presenting English-learning 6- and 9-month-olds with lists of monosyllabic nonwords with a consonant-vowel-consonant (CVC) structure that had high (e.g., riss, ghen, kazz) or low (e.g., yowdge, shawch, gushe) positional uni-and biphone probabilities in English. Infants at 9 but not 6 months listened significantly longer to the high than to the low probability lists of monosyllables. These two studies have been taken as evidence that knowledge of both language-specific phonotactic legality and phonotactic probability is acquired between 6 and 9 months.

Capitalizing on these early findings, subsequent studies showed that in their first year of life infants also acquire non-adjacent regularities (i.e., regularities between non-sequential phonemes) between consonants (Nazzi et al., 2009; Gonzalez-Gomez and Nazzi, 2012; Gonzalez-Gomez et al., 2014) and vowels[1] (Altan et al., 2016; Gonzalez-Gomez et al., 2019). Moreover, it has been found that infants' phonotactic knowledge supports other early linguistic processes, such as wordform segmentation (Mattys and Jusczyk, 2001; Gonzalez-Gomez and Nazzi, 2013) and word learning (Graf Estes et al., 2011; MacKenzie et al., 2012; Gonzalez-Gomez et al., 2013).

Most studies mentioned above [with the exception of Gonzalez-Gomez et al. (2014, 2019)] tested infants within one native language. Although infants' performance was in line with the phonotactic properties of their native language, which might attest to acquisition of native properties, it remains possible that language-general acoustic or structural properties of the words led to the preferences found. One way to avoid such a confound is to test, with the same task and words, groups of infants acquiring different languages with different phonotactic properties, and establish that their phonotactic preferences differ and relate to the properties of each language. The first crosslinguistic study on phonotactics compared Spanish-and Catalan-learning monolingual 10-month-olds, and two groups of Spanish-Catalan bilingual 10-month-olds differing in their predominant language (Sebastián-Gallés and Bosch, 2002). Using HPP, infants were presented with lists of nonwords containing coda CCs that were legal (e.g., /birt/, /dort/) versus illegal in Catalan (e.g., /ketr/ /bepf/). Importantly, Spanish being a language which forbids CCs in coda position, the two types of nonwords were phonotactically illegal in Spanish. The Catalan-learning monolinguals and both groups of bilinguals, but not the Spanish-learning monolinguals, preferred the list with phonotactic patterns legal in Catalan. A second study compared French-and Japanese-learning infants' acquisition of non-adjacent dependencies, showing that French-learning infants develop a preference for labial-coronal sequences between 6 and 10 months, while Japanese-learning infants develop a preference for coronal-labial sequences between 10 and 13 months, in line with the respective phonotactic properties of French and Japanese (Gonzalez-Gomez et al., 2014). Lastly, a study comparing Hungarian-and French-learning infants established the emergence of a preference for harmonic words between 10 and 13 months in Hungarian, a harmonic language, but no such preference at 13 months in French, a non-harmonic language (Gonzalez-Gomez et al., 2019). Taken together, these cross-linguistic studies demonstrate that infants' phonotactic sensitivity is tied to experience and knowledge of native language(s), rather than the acoustic/phonetic properties of the words presented. Interestingly, cross-linguistic studies also showed that phonotactic knowledge applies top-down, such that infants' ability to discriminate between words composed of specific phonotactic patterns differs depending on the phonotactic system of the language they are learning. Since Japanese does not allow CCs, adult speakers of Japanese tend to perceptually repair these by perceiving an illusory/u/ between the two consonants, whereas speakers of languages allowing CCs do not perceive this illusory vowel (see Dupoux et al., 2011). Cross-linguistic infant studies have shown that this effect emerges early in life: at 14 months, while French-and English-learning infants can clearly discriminate between word pairs in which one of the words contains a consonant cluster and the other one contains a/u/between the two consonants of the cluster (e.g., abna vs. abuna), Japanese-learning infants show a reduced or no ability to do so (Mazuka et al., 2011; but see also Kajikawa et al., 2006).

While early phonotactic sensitivity has been demonstrated for infants learning a number of different languages, meta-analytic evidence suggests that the age at which it emerges partly depends on the specific type of regularity investigated, and on the specific language infants are acquiring (Sundara et al., 2022). However, little is currently known about the factors modulating acquisition, as most previous studies investigating infants' sensitivity to phonotactic regularities used many different regularities in each phonotactic condition (e.g., 8 different legal versus 8 different illegal CCs embedded in 224 different nonwords; Sebastián-Gallés and Bosch, 2002) rather than focusing on

[1] Omane, P., Benders, T., and Boll-Avetisyan, N. (under review). Vowel harmony preferences in infants growing up in multilingual Ghana (Africa).

specific phonotactic contrasts (but see, e.g., Gonzalez-Gomez and Nazzi, 2015). Furthermore, in most studies, phonotactic regularities between conditions usually differ by more than one phoneme (e.g., / rt./versus/pf/) with at least one perceptually salient phoneme contrast, involving early acquired phonemes that infants can produce early in life (i.e., at the babbling stage): vowels, plosives, and nasals (de Boysson-Bardies, 1996; Morgan and Wren, 2018; Lorenzini and Nazzi, 2022). Accordingly, there are important gaps in our knowledge of the specific types of phonotactic regularities acquired in infancy. As a result, several factors likely to modulate acquisition have been posited, such as input properties (e.g., how much evidence in different languages supports a specific pattern), usefulness of phonotactic knowledge (e.g., how much phonotactic regularities inform other linguistic processes such as word-learning), type of regularity (e.g., adjacent vs. non-adjacent, or consonant vs. vowel dependencies), and perceptual salience/position of the regularities within the words (e.g., Bonatti et al., 2005; Zamuner, 2006; Sundara et al., 2022). Yet, no study has directly established the role of these factors. In the present study, we test the potential involvement of one such factor: perceptual salience.

To assess infants' acquisition of phonotactic regularities related to perceptually low-salient phonemes, the present study focused on fricative consonants. The definition of salience is still not clear across studies: understanding why some phonemes and phoneme contrasts are more salient than others, and which acoustic properties are linked to perceptual salience, is still a matter of investigation (for discussions, see Cristia et al., 2011; Chládková and Paillereau, 2020). Here, we define perceptual salience as infants' ability to discriminate between phonemes that contrast in one phonetic feature. Fricatives are a class of sounds that can be considered as low-salient given that previous studies have presented mixed findings with regard to infants' ability to perceive and discriminate fricative contrasts, which taken together suggest that discrimination of such contrasts is difficult in infancy. A set of studies (Eilers and Minifie, 1975; Eilers, 1977) suggests that English-learning infants cannot discriminate between fricative pairs such as /s/−/z/ and /f/−/θ/, but can discriminate between /s/−/v/ and /s/−/ʃ/. Moreover, Nittrouer (2001) tested English-learning infants on their discrimination between the fricative contrast /s/−/ʃ/ and either the vowel contrast /a/−/u/ or the stop voicing contrast /t/−/d/. Out of 15 infants who could discriminate vowel quality (either /sa/−/su/ or /ʃa/−/ʃu/), only 6 could also discriminate /sa/−/ʃa/, while out of 8 infants who could distinguish a stop voicing contrast (/ta/−/da/), none discriminated /sa/−/ʃa/. Importantly, the acquisition and perception of fricatives categories (or more specifically the /s/ and /ʃ/ categories) seems to be tightly linked to infants' ambient linguistic input: Cristia (2011) showed that fine-grained, subphonemic aspects of the acoustic realization of /s/ in caregivers' speech predicts 6-to 14-month-old infants' discrimination of this sound from /ʃ/, suggesting that learning based on acoustic cue distributions of /s/ and /ʃ/ in the infants' surrounding environment drives the acquisition of such categories. The fact that fricative contrasts are difficult to perceive by infants is relevant to phonotactic acquisition because the ability to discriminate between two phonemes is often a necessary requirement for phonotactic acquisition involving those phonemes. This was suggested by Zamuner (2006), who showed that Dutch-learning 9- and 11-month-olds do not display knowledge of the phonotactic final devoicing rule (resulting in allowing voiceless but no voiced plosives at Dutch word endings), possibly because infants at 9, 11 and 16 months do not show an ability to discriminate the voicing contrast in that position either.

At present, the only two studies that have explored infants' acquisition of phonotactic patterns related to fricatives have provided contrasting results. The first study (Gonzalez-Gomez and Nazzi, 2015) focused on the acquisition of non-adjacent dependencies linked to the relative order of labial and coronal consonants, asking whether infants' knowledge of such phonotactic constraints can be found when testing them with different types of consonants. The rationale was that in French, labial-coronal structures are more frequent than coronal-labial structures if both consonants are plosives or nasals, but coronal-labial structures are more frequent than labial-coronal structures if both consonants are fricatives. French-learning 10-month-olds preferred the most frequent phonotactic structures in all cases, showing a labial-coronal preference for plosive and nasal sequences, and a coronal-labial preference for fricative sequences. With respect to fricatives, this established that at 10 months, these infants could discriminate place of articulation of the fricatives, and had learned phonotactic regularities on these fricatives, hence indicating phonotactic acquisition linked to these perceptually low-salient phonemes. In the other study (Henrikson et al., 2020), English-learning full-and pre-term infants aged 7 to 14 months were tested on their preferences for phonotactic regularities involving low-salient phonemes, namely fricatives and liquids: frequent /ʃr/ and /sl/ versus infrequent /ʃl/ and /sr/ in word-initial position. No significant effects were found for the pre-term infants. For the full-term infants, an analysis restricted to the 9-month-olds showed a significant preference for the fricative-liquid patterns with the higher phonotactic probability in their language. However, when all four age groups (7, 9, 11, and 14 months) were analyzed together, no significant preference, nor interaction with age, was found, strongly reducing the significance of the effect found at 9 months. Hence, the results of this second study at best provide weak evidence of infants' preferences for frequent fricative-liquid patterns.

In this context of contrasting results between the two previous studies, the main goal of the present crosslinguistic study was to further investigate whether 9-month-old infants are sensitive to phonotactic regularities involving perceptually low-salient phonemes, namely fricatives. This was done by assessing whether French-and German-learning 9-month-old infants are sensitive to language-specific word-initial fricative-plosive regularities, specifically word-initial /s/− versus /ʃ/−plosive CCs. We chose to focus on such regularities with a cross-linguistic perspective for two reasons. First, we wanted a very clear contrast in phonotactics, meaning that the frequent regularities had to be very frequent in both their uniphone frequencies (i.e., frequencies of the word-initial fricative) and biphone frequencies (i.e., frequencies of the word-initial fricative-plosive cluster), whereas the infrequent ones had to be very infrequent in both uniphone and biphone frequencies. Word-initial fricative-and specifically fricative-plosive clusters are overall very frequent in these two languages, but the frequency of distribution of these two types of regularities contrasts between French and German: word-initial /s/ and /s/−plosive clusters are very frequent in French and very infrequent (or even illegal, but present in loanwords) in German, while word-initial /ʃ/ and /ʃ/−plosive clusters are very infrequent (or even illegal, but present in loanwords) in French and very frequent in German. Second, while being clearly contrasted at the phonotactic level, these two patterns are also composed of low-salient, perceptually very similar phonemes, differing only in place of articulation.

In the present study, we hypothesized that infants' phonotactic sensitivities at 9 months would extend to sensitivities to highly frequent versus infrequent regularities involving low-salient phonemes. Thus, given the between-language distributional contrast of /s/− and /ʃ/−plosive regularities described above, we predicted opposite preferences between infants of the two language groups (namely, a preference for /s/−plosive compared to /ʃ/−plosive in the French-learning group, and the opposite preference in the German-learning group). Note that novelty preferences can also be found in such designs, but given that it has rarely been documented in studies investigating phonotactic sensitivities (see Sundara et al., 2022; Figures 2 and 6), we consider a preference for the frequent phonotactic regularity within a language to be the more likely outcome. If confirmed, these findings would provide strong additional evidence of infants' acquisition by 9 months of language-specific phonotactic properties related to low-salient phonemes, adding to the evidence found in the two prior related studies each testing only one language group (Gonzalez-Gomez and Nazzi, 2015; Henrikson et al., 2020).

Methods

Ethical statement

Parents of all infant participants provided written informed consent prior to the experiment. Both the experimental protocol and consent procedure were according to the principles expressed in the Declaration of Helsinki and approved by the ethics committees of both Université Paris Cité (Nr. 2011-03) and University of Potsdam (Nr. 42_2023).

Participants

A total of 48 9-month-old infants from monolingual French (N = 24, mean age = 9.6 months, range = [9.1–10.1]) and German-speaking (N = 24, mean age = 9.5 months, range = [9.1–9.9]) families were included in the analyses. Seventeen additional infants (French-learning: N = 5; German-learning: N = 12) were tested but their data was not included in the final sample, because they were fussy (10), caregivers interacted with them during testing (2), they had otitis media within 5 days before the experimental session (1) or there was a technical/experimenter error (4). Finally, infants were included in the final sample only if they managed to complete at least one entire block (out of two blocks) before the experiment was stopped (one block included 12 experimental trials and two familiarization trials, see procedure below). Among all the participants included, 4 out of 24 French-learning infants did not finish the experiment, completing 20, 22, 24, and 26 trials out of 28, respectively. All German-learning infants completed the entire experiment.

All infants included in the analyses were in good health, had been born full-term (36–41 weeks of gestation), and had no known hearing or vision impairments. They were considered monolinguals, with a daily exposure to a single language (either French or German) above 80% of total language exposure as assessed through parental estimates (French-learning infants: M = 89.6%, SD = 8.7; German-learning infants: M = 97.5%, SD = 5.9). Families were contacted via the two babylabs' participant databases and received a small gift for participation (i.e., a colorful diploma with the picture of their infant). In Germany, caregivers were additionally compensated for their time and travel by a small fee.

Materials

Stimuli

Our stimuli consisted of 144 unique CCVCV nonwords, half of them starting with /s/ and the other half starting with /ʃ/, followed by either a /t/ or a /p/, thus giving four possible word-initial consonant clusters (i.e., /st/, /sp/, /ʃt/, /ʃp/), our phonotactic patterns of interest. These clusters were followed by one of 36 unique *VCV* tails and distributed such that the same tails appeared in both conditions (s-initial: /st/, /sp/ & ʃ-initial: /ʃt/, /ʃp/). The *VCV* structure of the tails was as follows: the first vowel was one of a set of six selected vowels attested and highly frequent in both French and German (/a/, /i/, /o/, /u/, /e/ & /y/). The onset of the second syllable included a variety of obstruents and sonorants (/k/, /g/, /d/, /b/, /m/, /n/, and /r/) to have acoustic variation between nonwords. No stimuli violated the OCP-Place constraint regarding the non-adjacent consonants C2 and C3: when the phoneme /p/ was present in the word-initial consonant cluster, C3 was never /m/ or /b/. When the phoneme /t/was present, C3 was never /n/or/d/. Finally, the word-final vowel was /a/ in one half and /i/ in the other half of the stimuli because these two vowels have a relatively comparable phonotactic probability in word-final position in the two languages, which was not the case for the rest of the vowels (e.g., notably, 70% of word-final vowels in German are schwas). The complete stimulus list can be found in Appendix 1 in Supplementary material.

Phonotactic probability

/s/−plosive and /ʃ/−plosive clusters were selected because they differ in phonotactic probability between French and German: word initially, both /s/− and /s/−plosive clusters are very frequent in French but infrequent in German, while both /ʃ/− and /ʃ/−plosive clusters are very frequent in German but infrequent in French. Calculations of phonotactic probability were performed using four different phonemically transcribed lexical databases: two based on adult speech, and two on infant-directed speech (IDS). For German, we used the German lemma database of CELEX (Baayen et al., 1995), including 51,322 number of different lemmas, and a lexical database by Stärk et al. (2022) derived from various CHILDES corpora (MacWhinney, 2000) including 1,660 number of word types. Similarly, for French, we used the LEXIQUE database (New et al., 2004), including 47,342 number of lemmas, and a French lexical database of infant-directed speech, including 5,533 number of word types. As no lexical database was publicly available for IDS, we derived our French IDS database from a corpus of phonemically transcribed IDS utterances (Carbajal et al., 2018), from which we segmented all unique words and extracted their frequency of occurrence to create a lexical database similar to the German IDS database. Following previous research showing that adults' phonotactic intuitions are better captured by type frequency measures (Denby et al., 2018), we used type frequency for all our phonotactic measures, meaning that the frequency of occurrence of a given word in the database was not taken into account. Because the tails were identical for /s/− and /ʃ/−initial nonwords, differences in phonotactic probability between experimental lists within a language were driven solely by the word-initial clusters in the nonwords.

As in Jusczyk et al. (1994), phonotactic probability was operationally defined based on two main measures, and calculated word-initially in the two languages (see Tables 1, 2).

TABLE 1 Frequency counts (and probability) of /s/ & /ʃ/ and /s/– & /ʃ/–consonant clusters in German and French (adult lexical databases), calculated in word-initial positions.

	Uniphone			Biphone	
	German	French		German	French
/s/	165 (0.0032)	3,875 (0.0832)	/st/	8 (0.0002)	296 (0.0068)
			/sp/	6 (0.0001)	174 (0.0040)
/ʃ/	4,478 (0.0873)	804 (0.0173)	/ʃt/	1,452 (0.0283)	11 (0.0003)
			/ʃp/	816 (0.0159)	2 (<0.0001)

TABLE 2 Frequency counts (and probability) of /s/ & /ʃ/ and /s/– & /ʃ/–consonant clusters in German and French (lexical databases based on IDS corpora), calculated in word-initial positions.

	Uniphone			Biphone	
	German	French		German	French
/s/	2 (0.0012)	411 (0.0748)	/st/	1 (0.0006)	18 (0.0035)
			/sp/	0 (0)	11 (0.0021)
/ʃ/	142 (0.0855)	139 (0.0253)	/ʃt/	44 (0.0266)	2 (0.0004)
			/ʃp/	26 (0.0157)	3 (0.0006)

TABLE 3 Frequency counts (and probability) of the phonemes /s/ and /ʃ/ in German and French (adult language and IDS) among all consonants (not positional).

	ADS		IDS	
	German	French	German	French
s	12,687 (0.0288)	17,915 (0.0579)	324 (0.036)	1,399 (0.053)
ʃ	11,975 (0.0272)	2,732 (0.0088)	254 (0.0283)	347 (0.013)

(1) Positional uniphone probability (i.e., how often a given phoneme occurs in a specific position within a word).
(2) Positional biphone probability (i.e., the phoneme-to-phoneme co-occurrence probability in a specific position within a word).

We also computed the overall probability of encountering the phonemes /s/ and /ʃ/ in each of the two languages' lexicons (see Table 3).

Results of the lexical statistics suggest that /s/ and /s/—initial CCs are more frequent than /ʃ/ and /ʃ/—initial CCs word-initially in French in ADS (Table 1) and IDS (Table 2). The opposite pattern is found in German: both /ʃ/ and /ʃ/—initial CCs are more frequent than both /s/ and /s/—initial CCs word-initially in ADS (Table 1) and IDS (Table 2). When calculating overall frequency of /s/ and /ʃ/, it can be seen that the former is more frequent than the latter in both languages, although this difference is much more marked in French than in German (Table 3).

Recordings and trial lists

We asked two different speakers, one monolingual German-native female and one monolingual French-native female, to pronounce all our experimental stimuli. To control for indexicality (e.g., voice characteristics), at the same time allowing us to assess if infants' phonotactic sensitivities are robust to acoustic-phonetic variability, we presented both language-specific pronunciations to our participants, in two experimental blocks counterbalanced for order of presentation across infants. Note that one set of studies had shown that French-learning infants' preferences for the phonotactic regularities of their native language were not impacted by the native language of the person recording the stimuli, which was either a French (Nazzi et al., 2009; Gonzalez-Gomez and Nazzi, 2012; Gonzalez-Gomez et al., 2013) or a Japanese (Gonzalez-Gomez et al., 2014) monolingual speaker.

The two speakers were instructed to read the stimuli in an IDS register. Their productions were recorded in the same sound-proof booth with the same technical equipment. Nonwords were then organized into 12 lists of 12 items each for each speaker. Six lists contained stimuli starting with the /s/—plosive clusters, and six lists contained stimuli starting with the /ʃ/—plosive clusters. The length of the lists was kept constant. In each list, items were presented with a silent interstimulus interval which varied in duration, so that each list would last exactly 18 s. The average intensity of the nonwords was normalized at 70 dB using PRAAT (Boersma, 2001). The average duration (ms) and pitch (Hz) of the nonwords can be found in Table 4. Overall, /s/—initial nonwords were longer than /ʃ/—initial nonwords in both German (Mean difference = 100 ms) and French (mean difference = 20 ms), the difference being more marked in German. In contrast, /ʃ/—initial nonwords were on average characterized by a higher pitch than /s/—initial nonwords, in both German (Mean difference = −12.60 Hz) and French (Mean difference = −7.28 Hz), the difference being more marked in German. The German stimuli followed a strong-weak stress pattern, which is typical for German, whereas the French stimuli were pronounced with even stress on all syllables, which is typical for French.

Procedure, apparatus, and design

For the experimental procedure, we used the HPP set-ups in the babylabs of Paris and Potsdam. During testing, caregivers were seated in a sound-attenuated testing booth with their infants on their lap, facing forward. Loudspeakers were mounted into the walls of the two side panels at about the level of the infants' heads. There were three

TABLE 4 Mean (SD) duration and pitch of nonwords separated by condition and pronunciation.

	German		French	
	s-initial	ʃ-initial	s-initial	ʃ-initial
Duration (ms)	779 (68)	676 (43)	461 (31)	441 (25)
Pitch (Hz)	262 (14)	274 (17)	272 (17)	280 (13)

lights mounted on the walls: a small green light directly in front of the infants, and two small red lights on either side of the infants, close to the two speakers. A video camera was also connected from below the central light (i.e., in front of the infants) to a monitor in an adjacent control room where the experimenter was located.

The experiment took place as follows. Infants and their caregiver (s) were welcomed by an experimenter. Caregivers first completed the consent form and were explained how the experiment would take place. Then a caregiver entered the testing booth with the infant. Once they were seated, the experiment started. The caregiver, who was instructed not to interact with the infant during testing, wore headphones playing experimental stimuli overlaid over music to efficiently mask the test stimuli. The experimenter, who recorded the infant's looking behavior via button presses, was also blind to the conditions of the study as no sound from the testing booth reached the control room. Each trial began by drawing the infant's attention to the center by flashing the central light. Once achieved, the central light was turned off and one of the two side lights started flashing. Once the infant turned and looked at it, the stimulus began to play (and the side light kept flashing during the trial). The trial ended when the entire stimulus for that trial had been played (an entire trial list lasted maximally 18 s) or when the infant turned away for at least a continuous period of 2 s. Infants' attention to the stimuli was measured based on their looking time toward the target side light on a given trial.

The experiment was made up of a total of 28 trials, divided into two blocks, with one block consisting of the stimuli pronounced by the French native speaker, and the other of the stimuli pronounced by the German native speaker. Each block started with two warm-up trials consisting of classical music, one on each side, and was followed by 12 test trials consisting of six trials with the nonwords starting with /s/−plosive and six trials with the nonwords starting with /ʃ/−plosive. During each trial, a unique list of 12 stimuli was played (e.g., list 1: /stika/, /stoga/, /spiki/, /spogi/, /steba/, /spedi/, /spuni/, /stuma/, /spyni/, /stara/, /styga/, /spari/). Trial type (/s/−plosive versus /ʃ/−plosive), pronunciation (German versus French) and side orders (left versus right) were pseudorandomized across participants. We created eight versions of the experiment such that half of the participants started the experiment with the French pronunciation block and the other half with the German pronunciation block. Furthermore, half of the participants started the experiment with an /s/−plosive trial, and the other half with a /ʃ/−plosive trial. Finally, for half of the participants the first trial was on the left side of the booth, and for the other half it was on the right side of the booth. There were never more than two consecutive trials on the same side and no more than two consecutive trials of the same phonotactic condition in a row.

Data pre-processing and analysis

All analyses were conducted in R-studio. We used linear mixed-effect models using the function lmer of the R package *lme4* (Bates et al., 2009), and the package *lmerTest* (Kuznetsova et al., 2017) to obtain *p*-values. We conducted a nested linear mixed-effect model, with infants' log-transformed looking times (hereafter LT, in seconds) as the dependent variable. Given the distribution of the data, LTs were log-transformed (as also recommended by Csibra et al., 2016). As fixed factors, we included native language (French vs. German) as between-participant nesting factor, phonotactics (/s/−initial vs. /ʃ/−initial) and pronunciation (French vs. German) in interaction, and block (block1 vs. block2) as within-participant factors. All factors were sum-contrasted for the model (i.e., effect coding: −1 vs. 1). Individual participant intercepts and by-participant random slopes for pronunciation were included in the random effects structure. Phonotactics was not added as a by-participant random slope because the model failed to converge when doing so. The full equation was as follows:

$$\begin{aligned}\log(\mathrm{LT}) \sim\ & \text{Native Language}\,/\\ & \left(\text{Phonotactics}\times\left(\text{Pronunciation}+\text{Block}\right)\right)\\ & +\left(1+\text{Pronunciation}|\text{Participant}\right).\end{aligned}$$

Note that we initially discussed whether we should compute a model with all factors in full interaction or the present nested one. Ultimately, we chose to compute the nested model as our confirmatory model for two reasons: (1) we were underpowered to reliably test for such interactions and (2) because both familiarity and novelty effects can emerge from experiments testing infants' preferences, the nested model provided the best possible way to assess infants' sensitivity to phonotactics without being restricted to one specific direction of preference, while also informing about the direction of this sensitivity.

Results

Eighteen out of 24 French-learning infants had longer LTs to the /s/−initial stimuli; 16 out of 24 German-learning infants had longer LTs to the /s/−initial stimuli. The results of the statistical model are presented in Table 5 and raw means and *CIs* are illustrated in Figure 1. Within the French-learning group, we found a significant main effect of phonotactics ($\beta = -0.07$, $SE = 0.03$, $p = 0.032$), indicating that infants' LTs to the /s/−initial nonwords ($mean = 7.99$ s, $SD = 1.90$s) were longer than their LTs to the /ʃ/−initial nonwords ($mean = 7.01$ s, $SD = 1.54$ s). Within the German-learning group, the main effect of phonotactics was not significant ($\beta = -0.03$, $SE = 0.031$, $p = 0.359$) [s-initial: $mean = 7.85$ s, $SD = 2.49$ s; ʃ-initial: $mean = 7.44$ s, $SD = 2.13$ s]. There was also a significant main effect of block in both language groups (French: $\beta = -0.23$, $SE = 0.03$, $p < 0.001$; German: $\beta = -0.09$, $SE = 0.03$, $p = 0.006$), indicating that infants' overall LTs decreased between the first and the second block. No further significant main effects or interactions were found.

TABLE 5 Results of the linear mixed model.

Predictors	log(LT)		
	Estimates	CI	*p*
(Intercept)	1.75	1.67–1.84	**<0.001**
Language	0.03	−0.05 – 0.12	0.442
Language [French]: Phonotactics	−0.07	−0.13 – −0.01	**0.031**
Language [German]: Phonotactics	−0.03	−0.09 – 0.03	0.359
Language [French]: Pronunciation	0.02	−0.04 – 0.08	0.564
Language [German]: Pronunciation	0.02	−0.04 – 0.09	0.447
Language [French]: block	−0.23	−0.29 – −0.17	**<0.001**
Language [German]: block	−0.09	−0.15 – −0.03	**0.006**
Language [French]:Phonotactics × Pronunciation	−0.04	−0.10 – 0.02	0.199
Language [German]: Phonotactics × Pronunciation	−0.04	−0.10 – 0.02	0.223
Language [French]:Phonotactics × block	−0.01	−0.07 – 0.05	0.764
Language [German]: Phonotactics × block	0.02	−0.04 – 0.08	0.585

P-values are highlighted in bold when smaller than 0.05.

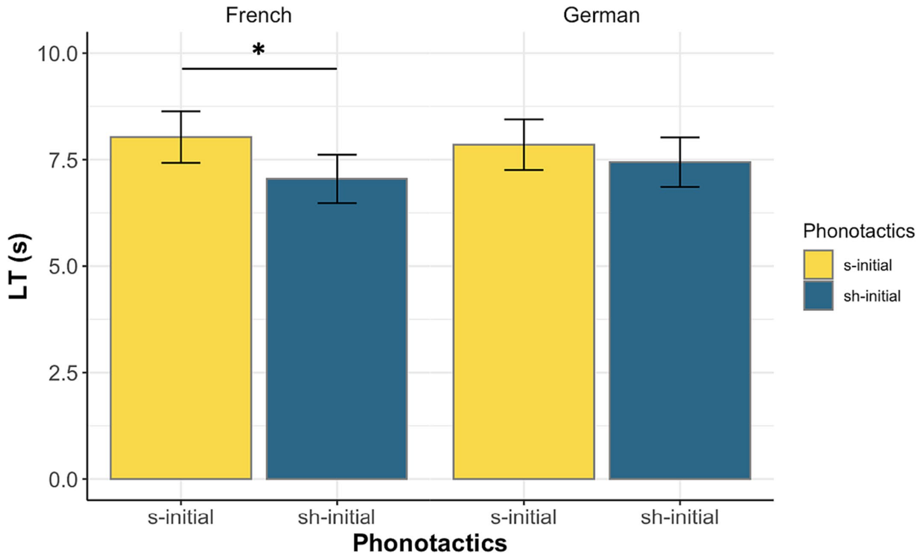


FIGURE 1
Raw infants' LTs (Means and CIs), broken down by phonotactics and language group.

Discussion

In a cross-linguistic design, the present study investigated French-and German-learning infants' sensitivity to low-salient, fricative-plosive word-initial phonotactic regularities at 9 months of age. To do so, we presented participants with lists of nonwords starting either with /s/—plosive clusters, with /s/ and /s/—plosive clusters being frequent word-initially in French but infrequent in German, or /ʃ/—plosive clusters, with /ʃ/ and /ʃ/—plosive clusters being frequent word-initially in German but infrequent in French, and measured their attention to either types of nonwords. Since previous studies found that infants start acquiring the phonotactics of their native language between 6 and 9 months (e.g., Friederici and Wessels, 1993), we expected our participants to show a significant preference for the frequent phonotactic regularities of their ambient language (French-learning participants: /s/—plosive clusters; German-learning participants: /ʃ/—plosive clusters). However, since the two previous studies on the acquisition of phonotactic properties related to perceptually low-salient fricatives found partly conflicting findings [with infants showing sensitivity in Gonzalez-Gomez and Nazzi (2015); but not at all ages in Henrikson et al. (2020)], it remained possible that infants would fail in the current experiment, or that performance would differ across languages. Our results show that the French-learning infants exhibited significantly longer LTs to the /s/—initial patterns, the more frequent regularities in their native language, than to the /ʃ/—initial patterns. In contrast, the German-learning group did not show a statistical preference for the frequent regularities in their native language.

These findings might be taken as evidence of cross-linguistic differences in infants' phonotactic sensitivities, and in their trajectory of acquisition of phonotactic regularities. Results for the French-learning group are compatible with previous studies on phonotactic acquisition, and suggests that infants' early phonotactic sensitivities extend to regularities involving perceptually low-salient, later-acquired phoneme contrasts in the French language. This is in line with the findings from Gonzalez-Gomez and Nazzi (2015) for another phonotactic regularity involving fricatives. Results for the German-learning group fail to provide evidence of knowledge of language-specific regularities on this low-salience fricative-based regularity. They contrast with the French results, but are in line with the difficulty found with English-learning infants tested on another fricative-based regularity in Henrikson et al. (2020).

To assess whether the difference in outcomes between our language-learning groups is statistically significant, we conducted an exploratory mixed effect model with a reduced number of parameters, with log (LT) as dependent variable and the two independent factors language and phonotactics in interaction (i.e., Log (LT) ~ Native Language x Phonotactics + (1|Participant)). This analysis only showed a main effect of phonotactics ($\beta = -0.049$, $SE = 0.023$, $p = 0.034$), indicating longer LTs to the /s/− than /ʃ/−plosive patterns, but failed to show a significant interaction between language and phonotactics ($\beta = 0.02$, $SE = 0.023$, $p = 0.380$). From this pattern of results, we cannot statistically conclude that the two groups of infants differed in their phonotactic sensitivities. To further assess if the data provides evidence for a null interaction or is just inconclusive, we used the Bayesian information criterion (BIC) for statistical inference, following Wagenmakers (2007). We computed two mixed effect models, one with and one without the interaction between language and phonotactics. We then extracted their respective BIC, and converted the BIC difference into a Bayes Factor, used to calculate the posterior probability of finding a null interaction (H0). This resulted in a Bayes Factor of 22.851, which amounts to a posterior probability of H0 of .96. This result can be interpreted as strong evidence that the data favors the null interaction, instead of being inconclusive (see Appendix 1 in Supplementary material for more details).

Based on this exploratory analysis, it is statistically more probable that both groups preferred the same phonotactic patterns (i.e., the /s/−plosive word-initial regularities), but the effect being smaller in the German-learning group, it did not emerge as significant in our confirmatory nested mixed-effect model. Another statistically less likely possibility, given that the data favors a null interaction between language and phonotactics, is that the two groups differed in their phonotactic sensitivities, with the French-learning group being sensitive to the frequent patterns in French and the German-learning group not being sensitive to the frequent pattern in German. However, the exploratory model was not sensitive enough to detect such a cross-linguistic difference with the current data. We discuss how we would interpret each of these two possibilities in what follows.

Let us consider first the possibility that both language groups do show a preference for the /s/−plosive word-initial regularity compared to the /ʃ/−plosive word-initial regularity. While this entails that both German-and French-learning infants are able to distinguish between the two low-salient phonemes /s/ and /ʃ/, results from the German-learning group are not in line with previous findings showing that infants prefer listening to the more frequent phonotactic regularities in their language. Since novelty phonotactic preferences have rarely been found (see Sundara et al., 2022), it seems unlikely that our findings could result from a cross-linguistic familiarity versus novelty preference.

An explanation for the overall preference for /s/−initial words could be found in the acoustic properties of our stimuli. Our /s/−initial nonwords were longer than our /ʃ/−initial nonwords, in both pronunciations (this difference being much more marked in the nonwords pronounced by the German native speaker). If 9-month-old infants are sensitive to durational differences in the range of 20-to 100-ms then the preference for /s/−plosive stimuli in both of our groups might partly be explained by such acoustic differences. Additionally, phonotactic sensitivity would reinforce French-but not German-learning infants' preference for the /s/−plosive stimuli, which could explain the larger effect in French. Note that our stimuli also differed in pitch (with higher pitch for the /ʃ/−initial nonwords), which either did not affect performance, or contributed to the pattern observed if infants preferred the stimuli with lower pitch, an unlikely preference given data on IDS preference showing infants' preference for stimuli with higher pitch (e.g., ManyBabies Consortium, 2020).

Finally, the general preference for the /s/− over /ʃ/−initial regularities might be linked to production factors. Productions studies suggest that /s/ is relatively easier to produce than /ʃ/, as illustrated by children experiencing a period of postalveolar fronting (the so-called "fis-effect," e.g., pronouncing "fish" as "fis," Jakobson, 1968; Kokkelmans, 2021), found in many languages including English (Vihman and Greenlee, 1987), German (Fox and Dodd, 1999), and French (Lemieux, 2011). Given studies showing that infants' ability to produce consonants impacts their processing of speech between 9 to 11 months (DePaolis et al., 2011, 2013; Majorano et al., 2014), our perceptual preference for /s/ over /ʃ/ could relate to its production advantage. Note, however, that recent evidence from French-learning infants reports no production of /s/ or /ʃ/ in 32 11-month-olds, and only 1 infant producing /s/ and 1 producing /ʃ/ out of 32 14-month-olds (Lorenzini and Nazzi, 2022), so it is not clear that production of /s/ is favored at this developmental stage, and could have impacted their performance. One direct way to explore this possibility would have been to ask the caregivers about their infant's babbling repertoire, and assess whether their production abilities are associated with their perceptual preferences in the current study. Since we did not collect this information, we leave this issue open for future research.

Let us now consider that our confirmatory nested mixed effect model shows cross-linguistic differences, with the French-learning group being sensitive to frequent regularities of low-salient phonemes in French while the German-learning group are not sensitive to frequent regularities of low-salient phonemes in German. What could explain this discrepancy between our two language-learning groups? Since French-learning infants were sensitive to our phonotactic manipulation, it seems unlikely that the lack of significant results in the German-learning group could be due to language-general processing abilities, such as for example an inability to discriminate fricatives at the age tested. Could the crosslinguistic difference be explained by phonotactic properties? This seems unlikely, since it was not the case that the difference in phonotactic frequency between conditions was more marked in French than in German. Indeed, both word-initial /ʃ/ and word-initial /ʃ/−plosive are comparatively much more common in German than word-initial /s/ and word-initial /s/−plosive are in French (see Tables 1, 2). This should have made it easier for German-than French-learning infants to acquire the respective phonotactic regularities. In contrast, our behavioral results might best be explained by sensitivity to overall/non-positional sound frequency within a language: while overall the phoneme /s/ is much more frequent than

/ʃ/in French, the frequencies of these two phonemes are relatively similar in German (see Table 3). Thus, it might be that the preferences in our study are related to phoneme frequency rather than phonotactic frequency. This is reminiscent of another study investigating infants' phonotactic preferences, which found that 7- but not 10-month-old French-learning infants prefer listening to coronal consonants compared to labial consonants, presumably because coronals are overall, as well as word-initially and word-finally, more frequent in French than labials (Gonzalez-Gomez and Nazzi, 2012). In our study, such sensitivity to non-positional phoneme frequencies is found at the intermediate age of 9 months, a possible delay related to the low-salience of the phonemes tested here. Further research will be needed on this finding, which was not predicted in the current study.

At any rate, our results further point toward the importance and the challenges of conducting cross-linguistic studies in language acquisition research, notably in the field of phonotactic acquisition. While in the current study, the frequency calculations would have predicted clear preferences in two opposite directions for our two language-learning groups, the lack of a significant result in the German-learning group suggests that frequencies are not enough to account for infants' preferences at that age: it is possible that learners of different languages rely and process phonotactic information differently, possibly giving them different weights at different ages. Indeed, previous studies suggest that adult listeners' use of specific phonotactic regularities for phoneme categorization differs depending on whether the adults were English or Dutch native-speakers, with English adults' perception being affected to a greater extent by diphone probabilities than Dutch adults' perception (Warner et al., 2005; Park et al., 2018; see Sundara et al., 2022 for a discussion). Relatedly, one cross-linguistic wellformedness judgment task points toward greater sensitivity to phonotactics in French compared to German adult listeners, although phonotactic probabilities predicted wellformedness judgments by both groups (Piot et al., 2024). It is possible that a differential use of phonotactic knowledge across individuals speaking different languages might already emerge in infancy and thus affect their sensitivities to phonotactic regularities. This is suggested by Jusczyk et al. (1993), documenting a similar discrepancy between English-and Dutch-learning infants, with stronger phonotactic sensitivity in English. For the specific case of fricatives tested here, findings for French suggest mastery of fricative-based properties by 9/10 months (Gonzalez-Gomez and Nazzi, 2015; current French data), while findings from English and German suggest failure or difficulties in acquiring fricative-based properties by the same age (Henrikson et al., 2020; current German data). The factors that drive these cross-linguistic differences (e.g., variable lexical stress, vocalic reduction, numerous complex codas and stress-timed rhythm in English/German, versus lack of lexical stress and vocalic reduction, less complex codas and syllable-timed rhythm in French) will have to be identified in future research.

Before concluding, we would like to point out some limitations of the present study. First, our experiment was relatively long for a 9-months-old infants' preference study: it took between 7 to 10 min to complete. Experimental length, coupled with the large amount of information that our participants had to process (i.e., 144 different nonwords, pronounced by two speakers with different native languages), might have been cognitively too demanding for them, resulting in noisy data. It is possible that a shorter experiment, or a reduced number of different stimuli might have better suited our purposes. Nevertheless, additional exploratory analyses, for each experimental block separately, showed results that are similar, although not significant, to our main analysis: the main effect of phonotactics was marginally significant in both blocks for the French group (1st block: $p = 0.108$, 2nd block: $p = 0.088$), while it was not significant for the German group. The lack of significance for these exploratory analyses could relate to low statistical power, especially when considering only one block: although rather typical compared to similar studies on infants' phonotactic acquisition, our sample size was relatively small. As a result, our effect would need to be further replicated, possibly with a bigger sample size. In any case, these explorations suggest that the relatively high number of trials was beneficial for detecting infants' phonotactic preferences in our experiment.

In sum, our findings suggest that infants' sensitivity to subtle, perceptually low-salient phonotactic patterns in their language at 9 months of age differs cross-linguistically. An implication for this finding is that infants' early phonotactic knowledge is already detailed and fine-grained, at least in French-learners, while our German-learning infants failed to show a preference for the frequent phonotactic pattern in German. Further studies are needed to understand whether this cross-language discrepancy can be explained by overall phoneme frequency, the use of challenging phonological categories or differences between French and German.

Data availability statement

The original contributions presented in the study are publicly available. The data and the R-script for data analysis are available on OSF: https://osf.io/pbk59/.

Ethics statement

The studies involving humans were approved by Ethics Committees of both Université Paris Cité (Nr. 2011-03) and University of Potsdam (Nr. 42_2023). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

LP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. TN: Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing. NB-A: Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing.

Funding

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project ID 317633480 & 491466077—SFB 1287, Project C07; and by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics - EFL.

Acknowledgments

The authors thank all the infants (and their parents) who participated in the study. A special thank to Tom Fritzsche & the Potsdam babylab team, as well as Flora Chartier & Maxine Dos Santos for their help in conducting this study.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1367240/full#supplementary-material

References

Altan, A., Kaya, U., and Hohenberger, A. (2016). Sensitivity of Turkish infants to vowel harmony in stem-suffix sequences: Preference shift from familiarity to novelty. In Proceedings of the 40th Boston University Conference on Language Development.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The CELEX lexical database (release 2). Distributed by the linguistic data consortium, University of Pennsylvania.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2009). Package 'lme4'. Available at: https://lme4.r-forge.r-project.org.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.* 5, 341–345.

Bonatti, L. L., Pena, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* 16, 451–459. doi: 10.1111/j.0956-7976.2005.01556.x

Carbajal, M. J., Bouchon, C., Dupoux, E., and Peperkamp, S. (2018). *A toolbox for phonologizing French infant-directed speech corpora*. IASCL Child Language Bulletin.

Chládková, K., and Paillereau, N. (2020). The what and when of universal perception: a review of early speech sound acquisition. *Lang. Learn.* 70, 1136–1182. doi: 10.1111/lang.12422

Cristia, A. (2011). Fine-grained variation in caregivers'/s/predicts their infants'/s/ category. *J. Acoust. Soc. Am.* 129, 3271–3280. doi: 10.1121/1.3562562

Cristia, A., McGuire, G. L., Seidl, A., and Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *J. Phon.* 39, 388–402. doi: 10.1016/j.wocn.2011.02.004

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., and Lengyel, M. (2016). Statistical treatment of looking-time data. *Dev. Psychol.* 52, 521–536. doi: 10.1037/dev0000083

de Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants* Odile Jacob.

Denby, T., Schecter, J., Arn, S., Dimov, S., and Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 280–294. doi: 10.1037/xlm0000465

DePaolis, R., Vihman, M. M., and Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behav. Dev.* 34, 590–601. doi: 10.1016/j.infbeh.2011.06.005

DePaolis, R., Vihman, M. M., and Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behav. Dev.* 36, 642–649. doi: 10.1016/j.infbeh.2013.06.007

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., and Peperkamp, S. (2011). Where do illusory vowels come from? *J. Mem. Lang.* 64, 199–210. doi: 10.1016/j.jml.2010.12.004

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *J. Acoust. Soc. Am.* 61, 1321–1336. doi: 10.1121/1.381435

Eilers, R. E., and Minifie, F. D. (1975). Fricative discrimination in early infancy. *J. Speech Hear. Res.* 18, 158–167. doi: 10.1044/jshr.1801.158

Fox, A. V., and Dodd, B. J. (1999). 'The phonological acquisition of German'. Sprache Stimme Gehoer.

Friederici, A. D., and Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Percept. Psychophys.* 54, 287–295. doi: 10.3758/BF03205263

Gonzalez-Gomez, N., Hayashi, A., Tsuji, S., Mazuka, R., and Nazzi, T. (2014). The role of the input on the development of the LC bias: a crosslinguistic comparison. *Cognition* 132, 301–311. doi: 10.1016/j.cognition.2014.04.004

Gonzalez-Gomez, N., and Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy* 17, 498–524. doi: 10.1111/j.1532-7078.2011.00104.x

Gonzalez-Gomez, N., and Nazzi, T. (2013). Effects of prior phonotactic knowledge on infant word segmentation: the case of nonadjacent dependencies. *J. Speech Lang. Hear. Res.* 56, 840–849. doi: 10.1044/1092-4388(2012/12-0138)

Gonzalez-Gomez, N., and Nazzi, T. (2015). Constraints on statistical computations at 10 months of age: the use of phonological features. *Dev. Sci.* 18, 864–876. doi: 10.1111/desc.12279

Gonzalez-Gomez, N., Poltrock, S., and Nazzi, T. (2013). A "bat" is easier to learn than a "tab": effects of relative phonotactic frequency on infant word learning. *PLoS One* 8:e59601. doi: 10.1371/journal.pone.0059601

Gonzalez-Gomez, N., Schmandt, S., Fazekas, J., Nazzi, T., and Gervain, J. (2019). Infants' sensitivity to nonadjacent vowel dependencies: the case of vowel harmony in Hungarian. *J. Exp. Child Psychol.* 178, 170–183. doi: 10.1016/j.jecp.2018.08.014

Graf Estes, K., Edwards, J., and Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy* 16, 180–197. doi: 10.1111/j.1532-7078.2010.00046.x

Henrikson, B., Seidl, A., and Soderstrom, M. (2020). Perception of sibilant–liquid phonotactic frequency in full-term and preterm infants. *J. Child Lang.* 47, 893–907. doi: 10.1017/S0305000919000825

Jakobson, R. O. (1968). "Child language: aphasia and phonological universals" in *Of Janua Linguarum, series minor*. ed. A. R. Keiler (Berlin, Boston: De Gruyter Mouton)

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *J. Mem. Lang.* 32, 402–420. doi: 10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *J. Mem. Lang.* 33, 630–645. doi: 10.1006/jmla.1994.1030

Kajikawa, S., Fais, L., Mugitani, R., Werker, J. F., and Amano, S. (2006). Cross-language sensitivity to phonotactic patterns in infants. *J. Acoust. Soc. Am.* 120, 2278–2284. doi: 10.1121/1.2338285

Kemler Nelson, D. G. K., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev.* 18, 111–116. doi: 10.1016/0163-6383(95)90012-8

Kokkelmans, J. (2021). *The phonetics and phonology of sibilants: A synchronic and diachronic OT typology of sibilant inventories*. Doctoral thesis. Italy: University of Verona.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606–608. doi: 10.1126/science.1736364

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82. 1–26. doi: 10.18637/jss.v082.i13

Lemieux, G. (2011). 'Le développement de la prononciation'. Online document.

Lorenzini, I., and Nazzi, T. (2022). Early recognition of familiar word-forms as a function of production skills. *Front. Psychol.* 13:947245. doi: 10.3389/fpsyg.2022.947245

MacKenzie, H. K., Curtin, S., and Graham, S. A. (2012). 12-month-olds' phonotactic knowledge guides their word-object mappings. *Child Dev.* 83, 1129–1136. doi: 10.1111/j.1467-8624.2012.01764.x

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Majorano, M., Vihman, M. M., and DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Lang. Learn. Dev.* 10, 179–204. doi: 10.1080/15475441.2013.829740

ManyBabies Consortium (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Adv. Methods Pract. Psychol. Sci.* 3, 24–52. doi: 10.1177/2515245919900809

Mattys, S. L., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121. doi: 10.1016/S0010-0277(00)00109-8

Mazuka, R., Cao, Y., Dupoux, E., and Christophe, A. (2011). The development of a phonological illusion: a cross-linguistic study with Japanese and French infants. *Dev. Sci.* 14, 693–699. doi: 10.1111/j.1467-7687.2010.01015.x

Morgan, L., and Wren, Y. E. (2018). A systematic review of the literature on early vocalizations and babbling patterns in young children. *Commun. Disord. Q.* 40, 3–14. doi: 10.1177/1525740118760215

Nazzi, T., Bertoncini, J., and Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *J. Acoust. Soc. Am.* 126, 1440–1446. doi: 10.1121/1.3158931

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004). Lexique 2: a New French lexical database. *Behav. Res. Methods Instrum. Comput.* 36, 516–524. doi: 10.3758/bf03195598

Nittrouer, S. (2001). Challenging the notion of innate phonetic boundaries. *J. Acoust. Soc. Am.* 110, 1598–1605. doi: 10.1121/1.1379078

Park, S., Hoffmann, M., Shin, P. Z., and Warner, N. L. (2018). The role of segment probability in perception of speech sounds. *J. Acoust. Soc. Am.* 143:1920. doi: 10.1121/1.5036263

Piot, L., Nazzi, T., and Boll-Avetisyan, N. Auditory phonotactic wellformedness intuitions depend on the nativeness of a speaker's pronunciation. Abstract, Linguistic Evidence (2024).

Sebastián-Gallés, N., and Bosch, L. (2002). Building phonotactic knowledge in bilinguals: role of early exposure. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 974–989. doi: 10.1037/0096-1523.28.4.974

Stärk, K., Kidd, E., and Frost, R. L. (2022). Word segmentation cues in German child-directed speech: a corpus analysis. *Lang. Speech* 65, 3–27. doi: 10.1177/0023830920979016

Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., and Steffman, J. (2022). Infants' developing sensitivity to native language phonotactics: a meta-analysis. *Cognition* 221:104993. doi: 10.1016/j.cognition.2021.104993

Vihman, M. M., and Greenlee, M. (1987). Individual differences in phonological development: ages one and three years. *J. Speech Hear. Res.* 30, 503–521. doi: 10.1044/jshr.3004.503

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105

Warner, N., Smits, R., McQueen, J. M., and Cutler, A. (2005). Phonological and frequency effects on timing of speech perception: a database of Dutch diphone perception. *Speech Comm.* 46, 53–72. doi: 10.1016/j.specom.2005.01.003

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

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9-to 16-month-old infants. *Infancy* 10, 77–95. doi: 10.1207/s15327078in1001_5

frontiers | Frontiers in Human Neuroscience

TYPE Mini Review
PUBLISHED 12 March 2024
DOI 10.3389/fnhum.2024.1352247

OPEN ACCESS

EDITED BY
István Winkler,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences
(MTA), Hungary

REVIEWED BY
Alexandra Bendixen,
Chemnitz University of Technology, Germany
Brigitta Toth,
Research Centre for Natural Sciences,
Hungarian Academy of Sciences
(MTA), Hungary

*CORRESPONDENCE
Axelle Calcus
✉ axelle.calcus@ulb.be

RECEIVED 07 December 2023
ACCEPTED 22 February 2024
PUBLISHED 12 March 2024

CITATION
Calcus A (2024) Development of auditory scene analysis: a mini-review.
Front. Hum. Neurosci. 18:1352247.
doi: 10.3389/fnhum.2024.1352247

Development of auditory scene analysis: a mini-review

Axelle Calcus*

Center for Research in Cognitive Neuroscience (CRCN), ULB Neuroscience Institute (UNI), Université Libre de Bruxelles, Brussels, Belgium

Most auditory environments contain multiple sound waves that are mixed before reaching the ears. In such situations, listeners must disentangle individual sounds from the mixture, performing the auditory scene analysis. Analyzing complex auditory scenes relies on listeners ability to segregate acoustic events into different streams, and to selectively attend to the stream of interest. Both segregation and selective attention are known to be challenging for adults with normal hearing, and seem to be even more difficult for children. Here, we review the recent literature on the development of auditory scene analysis, presenting behavioral and neurophysiological results. In short, cognitive and neural mechanisms supporting stream segregation are functional from birth but keep developing until adolescence. Similarly, from 6 months of age, infants can orient their attention toward a target in the presence of distractors. However, selective auditory attention in the presence of interfering streams only reaches maturity in late childhood at the earliest. Methodological limitations are discussed, and a new paradigm is proposed to clarify the relationship between auditory scene analysis and speech perception in noise throughout development.

KEYWORDS

development, auditory scene analysis (ASA), streaming, selective attention, neurophysiology

1 Introduction

Contrary to appearances, lively playgrounds and business meetings have one thing in common: they are noisy. In such complex auditory environments, sound waves are mixed before reaching the ears. Listeners must disentangle individual sounds from the mixture, performing what is called the *auditory scene analysis* (ASA; Bregman, 1990, 2015). Analyzing complex auditory scenes relies on the listeners' ability to segregate acoustic events into different streams, and to selectively attend to the stream of interest.

With respect to segregation, pioneer studies used sequences of tones organized temporally in repeated ABABAB patterns, where A and B represent successive tones of different frequencies (e.g., Miller and Heise, 1950; see Figure 1A). When listeners report hearing two streams, they are effectively experiencing stream segregation: they parse the *sequential* auditory events into distinct streams. At a given presentation rate, the larger the frequency distance between A and B, the more likely participants are to experience stream segregation. Later studies set out to evaluate segregation abilities in response to *simultaneous*, concurrent sounds. Listeners were presented with complex harmonic tones, of which one component had been mistuned (Moore et al., 1986; see Figure 1B) or delayed (Hedrick and Madix, 2009); manipulations that contributed to segregation into distinct auditory objects. With respect to selective attention, canonical studies investigated adults' ability to focus on a specific auditory feature in the

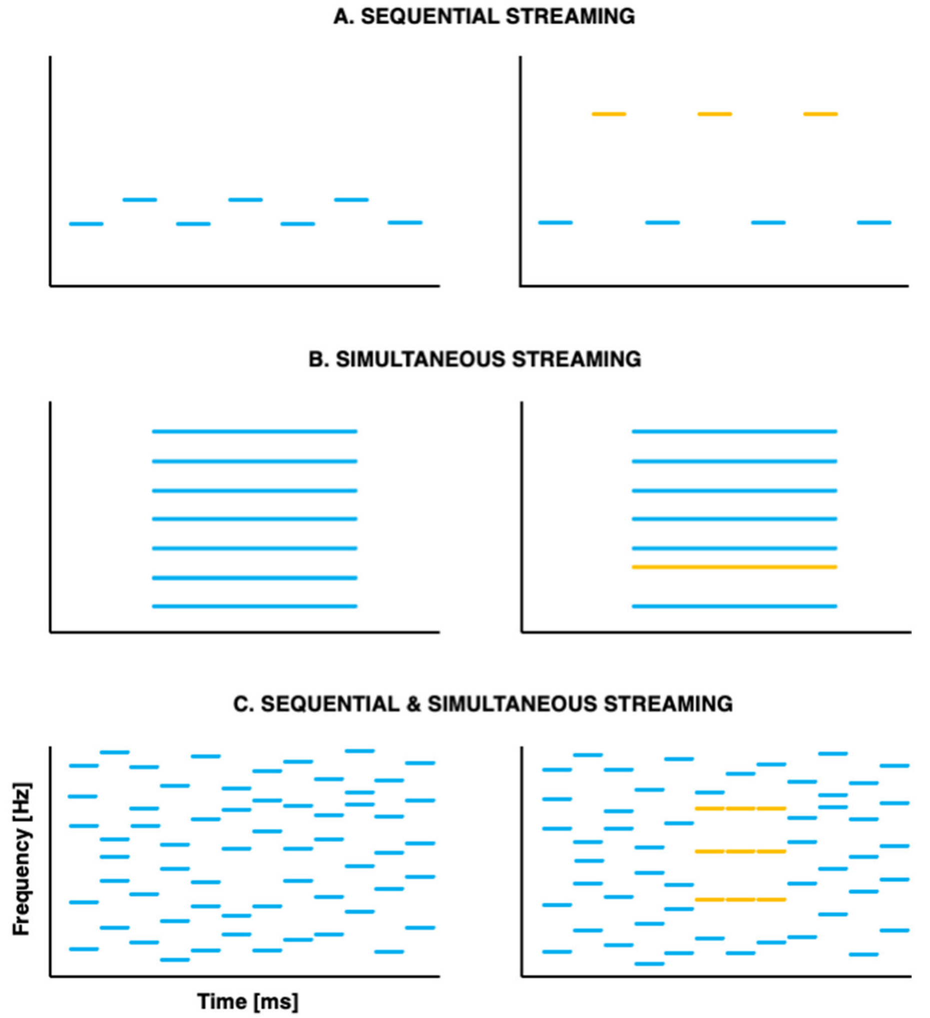

A. SEQUENTIAL STREAMING
B. SIMULTANEOUS STREAMING
C. SEQUENTIAL & SIMULTANEOUS STREAMING
Frequency [Hz]
Time [ms]

FIGURE 1
Schematic representation of the canonical paradigms used to investigate stream segregation **(A)** in sequences of successive tones; **(B)** in simultaneous concurrent sounds and **(C)** in stochastic tone clouds that combine successive and simultaneous tones. Stimuli that are typically perceived as one auditory stream are shown on the left; stimuli that are typically perceived as two auditory streams are shown on the right—with the two streams shown as different colors.

presence of simultaneous or sequential distractors (e.g., Greenberg and Larkin, 1968).

A major limitation of these early studies is their focus on either sequential or simultaneous stimuli. However, in everyday life, broadband streams that are temporally correlated often overlap with one another. In such situations, temporal coherence between different elements of the auditory scene appears essential for auditory segregation (Elhilali et al., 2009), potentially guiding selective attention such that it binds together coherent acoustic (spectral, spatial, and/or temporal) features into streams (Shamma et al., 2011). In this view, attention contributes not only to stream selection, but also to stream formation. An interesting development of the past decade was the creation of a paradigm in which the spectral coherence varies across time, requiring listeners to perform *both* simultaneous and sequential streaming at once (Teki et al., 2011, 2013; see Figure 1C).

How ASA develops in the first decades of life has attracted a lot of interest over the years. So far, studies focused on paradigms that tackled either sequential or simultaneous ASA. In a comprehensive review published about a decade ago on the topic, Leibold (2011) showed that sequential stream segregation and selective attention are functional early in life, albeit not yet as efficient as they are in adulthood. At the time, the author identified several open questions regarding the development of ASA: (i) How does simultaneous ASA develop from infancy to adulthood? (ii) Which acoustic cues are used by infants/children to perform ASA? (iii) How does sensorineural hearing loss affect the development of ASA? Here, we aim to review recent developmental data that answer some of these questions or raise new interrogations. We focus on studies using non-linguistic stimuli, to illustrate the development of basic auditory perception and processing involved in ASA, without the confound of language abilities.

2 Stream segregation

2.1 First year of life

Pioneer studies of ASA development investigated sequential streaming in the 1st year of life by habituating infants to a repeating (forward) sound sequence, then measuring their dishabituation to a reversed version of the sequence. Should infants parse the auditory scene based on each individual sound of the sequence, they would show a dishabituation response to the reversed pattern. On the contrary, newborns and 3-month-olds appeared to parse the streams of complex auditory scenes using the same cues adults use (Demany, 1982; McAdams and Bertoncini, 1997; Smith and Trainor, 2011)—albeit less accurately (for a detailed review, see Leibold, 2011).

Later studies investigated the neural correlates of sequential segregation in infants, using the mismatch negativity (MMN). The brain generates a MMN when it processes a difference between an unexpected auditory stimulus (a deviant) and the neural representation of a standard, expected pattern (for a review, see Näätänen et al., 2012). In adults, this "oddball paradigm" would even entail an MMN in the presence of interleaved sounds of a different frequency, as long as the interleaved sounds are perceived as separate streams (Sussman et al., 1999). Presented with this "interleaved" oddball paradigm, newborns also show an MMN, indicating that the neural correlates of sequential stream segregation are functional from birth (Winkler et al., 2003). Seven-month-olds also show an MMN if the deviant is placed in a chord component, and successive chords are played as a sequence (Marie and Trainor, 2013). Note that in this case, infants, like adults (Fujioka et al., 2005), show larger MMN to a deviant in the high than low voice, supporting early emergence of a preference for the highest stream.

In the last decade, a number of studies have set out to investigate the early development of *simultaneous* ASA, answering one of the open question identified by Leibold (2011). Folland et al. (2012) presented 6-month-old infants with complex tones consisting of 6 harmonic components. In half of the trials, one of the harmonic components was mistuned by 2–8% of its initial value. Infants were able to discriminate 4% mistuning or larger, whereas adults' thresholds were between 1 and 2% mistuning. Smith et al. (2017) paired in-tune and 8% mistuned complex tones with visual displays showing either one or two bouncing balls, hence being congruent or incongruent with the complex tones. Four-month-olds looked longer at incongruent audiovisual displays, indicating that they use harmonicity as a cue for stream segregation when integrating multisensory information. Whether newborns can segregate simultaneous auditory objects, or use acoustic cues to guide simultaneous streaming remains an open question.

To our knowledge, only two studies have investigated the neural correlates of simultaneous segregation in the 1st year of life, leading to contradictory results. Both studies used a similar paradigm, where half of the trials were 500 ms long complex tones of which the second harmonic was mistuned by 8% of its original value while the other half were in-tune complex tones. The object-related negativity (ORN) is an event-related potential that indexes listeners' processing of two simultaneous auditory objects (Alain et al., 2001). It is typically elicited by a mistuned component in otherwise harmonic complex tones (see Figure 1B). Whereas, newborns (Bendixen et al., 2015) and 4- to 12-month-old infants (Folland et al., 2015) showed an ORN in response to the mistuned complex tones, 2-month-olds did not (Folland et al., 2015). Future studies are needed to determine whether this discrepancy is due to methodological differences between the studies, or whether they reflect non-linearities in the development of the neural correlates of simultaneous stream segregation.

2.2 Childhood

For the sake of this review, childhood will be defined as ranging from 3 to 12 years of age. Most behavioral studies of stream segregation in children have been reviewed in Leibold (2011). They show that the acoustic difference required to segregate sequential or simultaneous sounds into distinct streams decreases as children grow older, but remains larger in late childhood than in adulthood (Alain et al., 2003; Sussman et al., 2007; Sussman and Steinschneider, 2009). Note that 5- to 13- year-old children benefit from visual cues helping simultaneous ASA to the same extent as adults (Bonino et al., 2013). Yet 5-year-olds show less benefit from spatial cues to perform simultaneous stream segregation than adults (Wightman et al., 2003).

Electrophysiological studies are in line with the behavioral observation of immature stream segregation in children up to 12 years of age. Like infants, children show an MMN when presented with stimuli that entail sequential streaming (Sussman et al., 2001; Lepistö et al., 2009). However, the frequency separation between the successive sounds of these sequences needs to be larger in passively attending 9–12 year-olds than adults to elicit an MMN (Sussman and Steinschneider, 2009).

With respect to simultaneous ASA, Alain et al. (2003) recorded the ORN in 8- to 13-year-old children and adults. Their results indicate that children have a *larger* ORN than adults, despite having poorer behavioral performance when segregating streams in the mistuned complex tones. This was interpreted as suggesting greater neuronal activity associated with the perception of separate auditory objects in children than adults. In a recent follow-up to that study, the same team investigated the ORN of 6–12 year-olds with a moderate to severe congenital hearing loss (55–70 dB HL), who were regular hearing aid users (Mehrkian et al., 2022). Note that children with a hearing loss were tested unaided, but sounds were presented at higher sound pressure level than for age-matched children with normal hearing, thus aiming to equate sensation level across groups. Children with a hearing loss had smaller and later ORN than age-matched children with normal hearing. Congenital sensorineural hearing loss thus seem to have a pervasive effect on the central processing of simultaneous streams, that is not merely due to an audibility loss.

2.3 Adolescence

In the past decade, researchers started to investigate the maturational trajectory of ASA at adolescence. The frequency separation needed to experience streaming of successive tones did not change between 7 and 15 years (Sussman et al., 2015). However, in the same study, there was a gradual improvement in the ability to detect an intensity deviant in one of two sequential streams. More studies are needed to investigate adolescent development of simultaneous ASA, and to explore the neural correlates of both sequential and simultaneous streaming at adolescence.

3 Selective attention in the context of ASA

3.1 First year of life

Do infants use selective attention to guide streaming in complex auditory scenes? To address this challenging question, researchers have investigated the effects of non-sensory factors on the detection of an auditory target in the presence of *simultaneous* distractors (for a detailed review, see Leibold, 2011). From 6 months of age, infants rely on temporal (Werner et al., 2009) but not spectral (Werner and Bargones, 1991; Bargones and Werner, 1994) expectations to selectively direct their attention toward a target in the presence of a simultaneous interference. Several questions remain open: are newborns able to selectively direct their attention in complex auditory scenes? Are infants able to selectively attend to a target that unfolds over time in a sequential stream? What are the neural correlates of infants' selective attention in the presence of auditory distractors?

3.2 Childhood

Behavioral studies of simultaneous ASA in children have been reviewed by Leibold (2011), and suggest a progressive improvement in selective auditory attention throughout the primary school years (Greenberg et al., 1970; Allen and Wightman, 1995; Stellmack et al., 1997; Leibold and Neff, 2007). A recent psychoacoustic study aimed at understanding the mechanism underlying this progressive improvement (Jones et al., 2015). Reverse correlations were used to estimate which spectral region children and adults paid attention to when asked to detect a 1 kHz target embedded in an unpredictable noise. Results confirmed that 4- to 7-year-olds had poorer thresholds than 8- to 11-year-olds and adults. In fact, younger children were less efficient at analyzing the spectral content of the stimuli than older children. Their poorer thresholds in noise thus likely reflect an inability to selectively attend to the target while ignoring the distractor. How selective attention to sequential sound streams develops during childhood remains so far unexplored.

Neural correlates of selective attention to sequential streams can be investigated using a variation of the "oddball paradigm" described above (Sussman et al., 1999; Winkler et al., 2003). Participants are presented with two streams of interleaved sounds, differing in frequency (see Figure 1A, right panel). They are asked to focus on one of the streams, and to indicate when they detect a deviant within this target stream, while ignoring deviants that appear in the distracting stream. This allows to compare the neural response of the to-be-attended deviant to that of the to-be-ignored deviant, which typically leads to an early frontal positivity followed by a difference negativity (Nd, for a review see Näätänen et al., 2001). Nds were recorded in a group of 9-year-olds, a group of 12-year-olds, and a group of adults (Gomes et al., 2007). Both groups of children exhibited a later Nd than adults, indicating persistent processing immaturities in sequential streaming in late childhood. Whether persistent processing immaturities would also be observed in the neural correlates of simultaneous streaming, despite the seemingly mature behavioral performance (Jones et al., 2015) remains an open question.

3.3 Adolescence

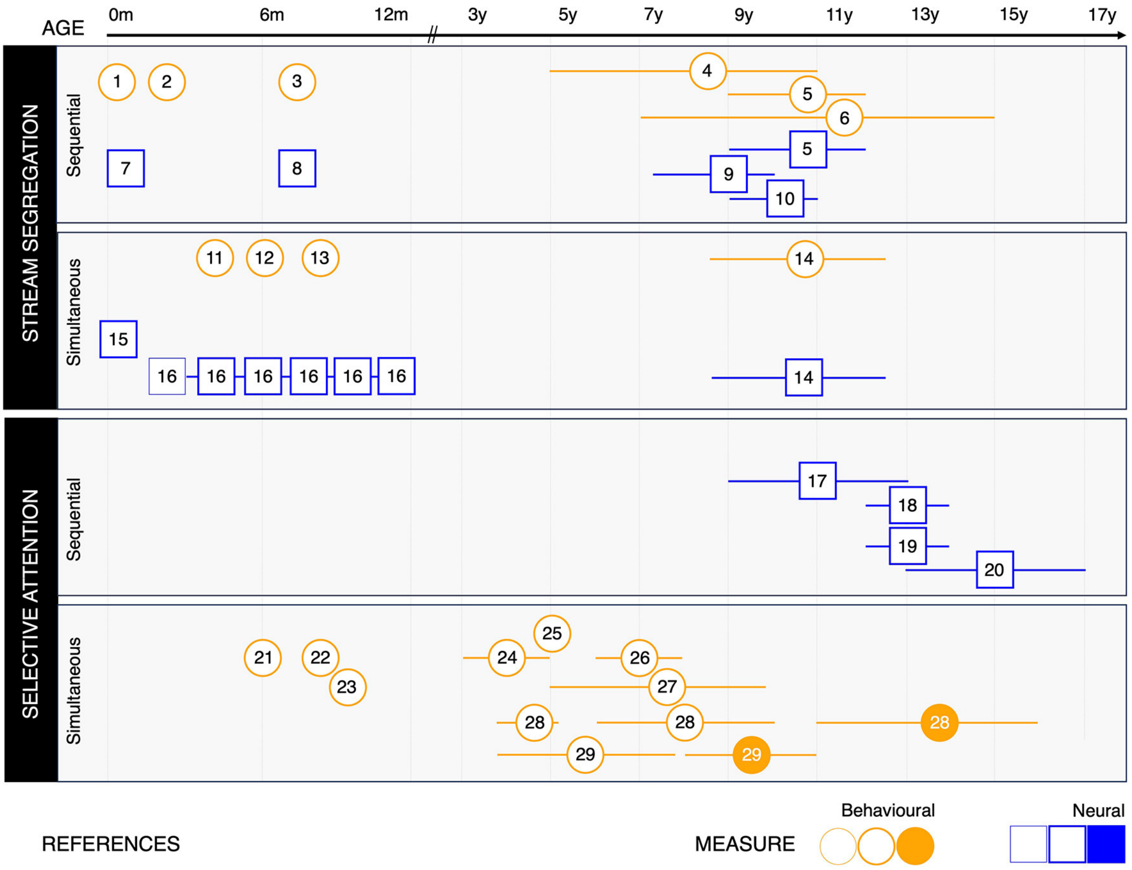
Selective auditory attention to a target in the presence of a simultaneous multitone masker seems to be mature by late childhood (Jones et al., 2015). This observation is consistent with earlier results collected in a small cohort of children as well as adolescents and adults (Lutfi et al., 2003). Whereas, 4- to 10-year-old children showed more masking than adults, there was no difference between adolescents (11–16 years) and adults. A principal component analysis was performed on the variance in masking performance, to investigate whether different age groups and/or individuals use different detection strategies. If so, several components would be identified as significantly contributing to the variance observed in masking performance. On the contrary, a single principal component was found to account for more than 80% of variance in masking performance, both across and within age groups. This suggests that children use similar target detection strategies to adults, but that they vary in their selective attention abilities.

A few studies have investigated the neural correlates of selective attention during sequential streaming at adolescence. Nds did not change between 11- and 14 year-olds as they were asked to detect a deviant in a target stream while ignoring those in the distracting stream (D'Angiulli et al., 2008). Interestingly, the early frontal positivity evoked by the to-be-ignored targets was larger in adolescents with poorer executive functioning skills than in those with higher executive function skills (Lackner et al., 2013). Last, an oddball paradigm was presented with different instructions, directing adolescents' attention toward different auditory cues, or away from the auditory modality and toward visual information (Sussman, 2013). The morphology of adolescents' event-related potentials and MMN varied with the instructions, like adults' (Sussman et al., 2002).

Overall, studies did not find developmental effects on the neural correlates of selective attention to sequential streams, which may indicate mature attentional responses at adolescence. Note however that none of the studies reviewed in the above paragraph included a group of adults, which limits interpretation in terms of the maturational trajectory at adolescence. Additionally, how the neural correlates of selective attention in simultaneous segregation tasks develop throughout adolescence remains unexplored.

4 Discussion

Figure 2 shows the studies reviewed in this paper, with respect to the age range of their pediatric population, the type of measure collected, and the specific ASA ability investigated. To sum up,



1. McAdams & Bertoncini, 1997; **2**. Demany, 1982; **3**. Smith & Trainor, 2011; **4**. Sussman et al., 2007; **5**. Sussman & Steinschneider, 2009; **6**. Sussman et al., 2015; **7**. Winkler et al., 2003; **8**. Marie & Trainor, 2013; **9**. Sussman et al., 2001; **10**. Lepisto et al., 2009; **11**. Smith et al., 2017; **12**. Folland et al., 2012; **13**. Newman et al., 2013; **14**. Alain et al., 2003; **15**. Bendixen et al. 2015; **16**. Folland et al., 2015; **17**. Gomes et al., 2007; **18**. D'Angiulli et al., 2008; **19**. Lackner et al., 2013; **20**. Sussman et al., 2013; **21**. Werner & Bargones, 1991; **22**. Bargones & Werner, 1994; **23**. Werner et al. 2009; **24**. Allen & Wightman, 1995; **25**. Stellmack et al. 1997; **26**. Greenberg et al., 1970; **27**. Leibold & Neff, 2007; **28**. Lutfi et al., 2003; **29**. Jones et al., 2015

FIGURE 2
Studies on auditory scene analysis (ASA) throughout development, organized according to the specific ASA ability investigated, the type of measure collected, and the developmental results reported. Behavioral studies are represented as circles (orange); neurophysiological studies are represented as squares (blue). Symbols are positioned at the mean age of the pediatric participants included in the study. Whiskers around the symbols indicate the age range included in the study, whenever relevant. Thin symbols indicate that participants did not show evidence of the ASA ability investigated (Folland et al., 2015). Regular symbols indicate that participants were able to perform ASA. Filled symbols indicate there was no significant difference between the performance of pediatric participants and a group of adults included in the study.

cognitive and neural mechanisms supporting both simultaneous and sequential stream *segregation* are functional from birth. Yet, their efficiency keeps improving throughout childhood and adolescence (Alain et al., 2003; Sussman et al., 2007, 2015; Sussman and Steinschneider, 2009).

Developmental studies of selective auditory *attention* in the context of ASA paint a seemingly contradictory picture. From 6 months of age, infants benefit from some (but not all) auditory cues to orient their attention toward a target in the presence of simultaneous interferers (Werner and Bargones, 1991; Bargones and Werner, 1994; Werner et al., 2009), in line with neurophysiological data showing developmental changes in arousal over the first 2 years of life (Richards et al., 2010). Yet, the existent developmental data on selective attention in ASA suggest that behavioral performance reaches maturity by late childhood (Lutfi et al., 2003; Jones et al., 2015), whereas its neural correlates keep maturing until adolescence (Gomes et al., 2007). Two explanations might account for this apparent discrepancy. First, children may perform similarly to adults by recruiting different cognitive resources (Trau-Margalit et al., 2023). Future studies are thus warranted to investigate the development of the neural markers of listening effort in noise. Second, the literature on selective attention in the context of ASA seems to present a blind spot. Indeed, to the best of our knowledge, behavioral studies all used simultaneous streaming tasks, whereas neurophysiological studies used sequential streaming tasks. Discrepancies between behavioral and neural results may thus stem from different maturational trajectories between simultaneous and sequential streaming tasks. Note however that speech-in-speech perception inherently requires *both* simultaneous and sequential ASA abilities, whereas the bulk of the literature reviewed here has focused on one or the other. Noteworthily, studies investigating selective attention to speech in the presence of distractors indicate a protracted development of neurophysiological attentional responses from childhood until adulthood (Berman and Friedman, 1995; Karns et al., 2015).

This supports the need to better understand the development of ASA in more ecological situations that require both simultaneous and sequential streaming abilities. The stochastic figure-ground paradigm (Teki et al., 2011, 2013; see Figure 1C) offers a unique opportunity in this respect. The paradigm consists in a series of identical chords (the figure) presented against a background of random chords. Adults are remarkably sensitive to the appearance of such figures in stochastic noise backgrounds—discrimination performance even improves as figure coherence increases. Additionally, the ORN and a later positive wave (P400) have been elicited in adults listening to such stochastic sequences, providing "neural signatures" of figure-ground discrimination (Tóth et al., 2016). Adapting this task to children and adolescents would further our understanding of the development of combined simultaneous and sequential streaming, as is often required in real-life.

Other limitations are that most studies focused on narrow age ranges, and a number of them did not include a group of adult participants. In addition, most of the results reported here stem from single studies that addressed a specific question. In the few cases where more than one study was conducted to address a research question in a specific age range, results were partly contradictory. This points toward the need for comprehensive developmental investigations, including replication studies. This would allow to examine the transition toward adult-like performance, and the factors that contribute to this transition, including those that relate to individual differences in maturation. Cognitive (executive functions and working memory), neurochemical (modulation of serotonin, dopamine and gamma-aminobutyric acid) and environmental factors (exposure to music and language) should be included as potential predictors of maturation, as they are thought to contribute to stream segregation and/or speech perception in noise in adults (Moore et al., 2008; Kondo et al., 2012; Lackner et al., 2013; van Loon et al., 2013; Chabal et al., 2015; Tierney et al., 2020; Porto et al., 2023). Last, future studies are warranted to disentangle the relationship between selective attention and auditory streaming throughout development.

Together, this would pave the way toward a model of ASA development from infancy to adulthood. This would be contribute to understand typical development, and to better grasp the difficulties faced by clinical populations in noisy environments. Many children seem to be disproportionally affected by the presence of background noise (Calcus et al., 2018; Sharma et al., 2019). Adding insult to injury, classrooms are notoriously noisy (Brill et al., 2018). A better understanding of ASA development may therefore have a significant societal impact on the academic performance of children/adolescents in noisy environments.

Author contributions

AC: Writing—original draft, Writing—review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. We gratefully acknowledge funding from the H2020 European Research Council under grant 101076968.

Acknowledgments

We are very thankful to two reviewers for their comments on a previous version of this manuscript.

Conflict of interest

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

Publisher's note

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

References

Alain, C., Arnott, S., and Picton, T. (2001). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol.* 27, 1072–1089. doi: 10.1037//0096-1523.27.5.1072

Alain, C., Theunissen, E., Chevalier, H., and Batty, M. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cogn. Brain Res.* 16, 210–218. doi: 10.1016/S0926-6410(02)00275-6

Allen, P., and Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *J. Speech Hear. Res.* 38, 503–511. doi: 10.1044/jshr.3802.503

Bargones, J. Y., and Werner, L. (1994). Adults listen selectively; infants do not. *Psychol. Sci.* 5, 170–174. doi: 10.1111/j.1467-9280.1994.tb00655.x

Bendixen, A., Haden, G., Nemeth, R., Farkas, D., Torok, M., and Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. *Dev. Neurosci.* 37, 172–181. doi: 10.1159/000370237

Berman, S., and Friedman, D. (1995). The development of selective attention as reflected by event-related brain potentials. *J. Exp. Child Psychol.* 59, 1–31. doi: 10.1006/jecp.1995.1001

Bonino, A. Y., Leibold, L. J., and Buss, E. (2013). Effect of signal-temporal uncertainty in children and adults: tone detection in noise or a random-frequency masker. *J. Acoust. Soc. Am.* 134, 4446–4457. doi: 10.1121/1.4828828

Bregman, A. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sounds*. Cambridge, MA: MIT Press.

Bregman, A. (2015). Progress in understanding auditory scene analysis. *Music Percept*. 33, 12–19. doi: 10.1525/mp.2015.33.1.12

Brill, L., Smith, K., and Wang, L. (2018). Building a sound future for students: considering the acoustics in occupied active classrooms. *Acoust. Tod*. 14, 14–21.

Calcus, A., Deltenre, P., Colin, C., and Kolinsky, R. (2018). Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Dev. Sci.* 21:12558. doi: 10.1111/desc.12558

Chabal, S., Schroeder, S., and Marian, V. (2015). Audio-visual object search is changed by bilingual experience. *Atten. Percept. Psychophys*. 77, 2684–2693. doi: 10.3758/s13414-015-0973-7

D'Angiulli, A., Herdman, A., Stapells, D., and Hertzman, C. (2008). Children's event-related potentials of auditory selective attention vary with their socioeconomic status. *Neuropsychology* 22, 293–300. doi: 10.1037/0894-4105.22.3.293

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behav. Dev*. 5, 261–276. doi: 10.1016/S0163-6383(82)80036-2

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A., and Shamma, S. (2009). Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61, 317–329. doi: 10.1016/j.neuron.2008.12.005

Folland, N. A., Butler, B. E., Payne, J. E., and Trainor, L. J. (2015). Cortical representations sensitive to the number of perceived auditory objects emerge between 2 and 4 months of age: electrophysiological evidence. *J. Cogn. Neurosci*. 27, 1060–1067. doi: 10.1162/jocn_a_00764

Folland, N. A., Butler, B. E., Smith, N. A., and Trainor, L. J. (2012). Processing simultaneous auditory objects: infants' ability to detect mistuning in harmonic complexes. *J. Acoust. Soc. Am*. 131, 993–997. doi: 10.1121/1.3651254

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci*. 17, 1578–1592. doi: 10.1162/089892905774597263

Gomes, H., Duff, M., Barnhardt, J., Barrett, S., and Ritter, W. (2007). Development of auditory selective attention: event-related potential measures of channel selection and target detection. *Psychophysiology* 44, 711–727. doi: 10.1111/j.1469-8986.2007.00555.x

Greenberg, G. Z., Bray, N. W., and Beasley, D. S. (1970). Children's frequency-selective detection of signals in noise. *Percept. Psychophys.* 8:BF03210199. doi: 10.3758/BF03210199

Greenberg, G. Z., and Larkin, W. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J. Acoust. Soc. Am*. 44, 1513–1523. doi: 10.1121/1.1911290

Hedrick, M. S., and Madix, S. G. (2009). Effect of vowel identity and onset asynchrony on concurrent vowel identification. *J. Speech Lang. Hear. Res*. 52, 696–705. doi: 10.1044/1092-4388(2008/07-0094)

Jones, P. R., Moore, D. R., and Amitay, S. (2015). Development of auditory selective attention: why children struggle to hear in noisy environments. *Dev. Psychol*. 51, 353–369. doi: 10.1037/a0038570

Karns, C. M., Isbell, E., Giuliano, R. J., and Neville, H. J. (2015). Auditory attention in childhood and adolescence: an event-related potential study of spatial selective attention to one of two simultaneous stories. *Dev. Cogn. Neurosci*. 13, 53–67. doi: 10.1016/j.dcn.2015.03.001

Kondo, H., Kitagawa, N., Kitamura, M., Koizumi, A., Nomura, M., and Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cerebr. Cortex* 22, 1915–1922. doi: 10.1093/cercor/bhr266

Lackner, C. L., Santesso, D. L., Dywan, J., Wade, T. J., and Segalowitz, S. J. (2013). Electrocortical indices of selective attention predict adolescent executive functioning. *Biol. Psychol*. 93, 325–333. doi: 10.1016/j.biopsycho.2013.03.001

Leibold, L. J. (2011). Development of auditory scene analysis and auditory attention. *Hum. Audit. Dev*. 42, 137–161. doi: 10.1007/978-1-4614-1421-6_5

Leibold, L. J., and Neff, D. L. (2007). Effects of masker-spectral variability and masker fringes in children and adults. *J. Acoust. Soc. Am*. 121, 3666–3676. doi: 10.1121/1.2723664

Lepistö, T., Kuitunen, A., Sussman, E., Saalasti, S., Jansson-Verkasalo, E., Wendt, T. N., et al. (2009). Auditory stream segregation in children with Asperger syndrome. *Biol. Psychol*. 82, 301–307. doi: 10.1016/j.biopsycho.2009.09.004

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Percept. Psychophys*. 65, 396–406. doi: 10.3758/BF03194571

Marie, C., and Trainor, L. J. (2013). Development of simultaneous pitch encoding: infants show a high voice superiority effect. *Cerebr. Cortex* 23, 660–669. doi: 10.1093/cercor/bhs050

McAdams, S., and Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *J. Acoust. Soc. Am*. 102, 2945–2953. doi: 10.1121/1.420349

Mehrkian, S., Moossavi, A., Gohari, N., Nazari, M. A., Bakhshi, E., and Alain, C. (2022). Long latency auditory evoked potentials and object-related negativity based on harmonicity in hearing-impaired children. *Neurosci. Res*. 178, 52–59. doi: 10.1016/j.neures.2022.01.001

Miller, G., and Heise, G. (1950). The trill threshold. *J. Acoust. Soc. Am*. 22, 637–638. doi: 10.1121/1.1906663

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am*. 80, 479–483. doi: 10.1121/1.394043

Moore, D., Ferguson, M., Halliday, L., and Riley, A. (2008). Frequency discrimination in children: perception, learning and attention. *Hear. Res*. 238, 147–154. doi: 10.1016/j.heares.2007.11.013

Näätänen, R., Alho, K., and Schröger, E. (2001). "Electrophysiology of attention," in *Steven's Handbook of Experimental Psychology, 3rd Edn, vol. 4*, ed. H. Pashler (New York, NY: John Wiley & Sons), 601–653.

Näätänen, R., Kujala, T., Escera, C., Baldeweg, T., Kreegipuu, K., Carlson, S., et al. (2012). The mismatch negativity (MMN)—a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin. Neurophysiol*. 123, 424–458. doi: 10.1016/j.clinph.2011.09.020

Porto, L., Wouters, J., and van Wieringen, A. (2023). Speech perception in noise, working memory, and attention in children: a scoping review. *Hear. Res*. 439:108883. doi: 10.1016/j.heares.2023.108883

Richards, J., Reynolds, G., and Courage, M. (2010). The neural bases of infant attention. *Curr. Direct. Psychol. Sci*. 19 41–46. doi: 10.1177/0963721409360003

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* 34, 114–123. doi: 10.1016/j.tins.2010.11.002

Sharma, M., Purdy, S., and Humburg, P. (2019). Cluster analyses reveals subgroups of children with suspected auditory processing disorders. *Front. Psychol*. 10:2481. doi: 10.3389/fpsyg.2019.02481

Smith, N. A., Folland, N. A., Martinez, D. M., and Trainor, L. J. (2017). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition* 164, 1–7. doi: 10.1016/j.cognition.2017.01.016

Smith, N. A., and Trainor, L. J. (2011). Auditory stream segregation improves infants' selective attention to target tones amid distracters. *Infancy* 16, 655–668. doi: 10.1111/j.1532-7078.2011.00067.x

Stellmack, M., Willihnganz, M., Wightman, F., and Lutfi, R. (1997). Spectral weights in level discrimination by preschool children: analytic listening conditions. *J. Acoust. Soc. Am*. 101, 2811–2821. doi: 10.1121/1.419479

Sussman, E. (2013). Attention matters: pitch vs. pattern processing in adolescence. *Front. Psychol*. 4:333. doi: 10.3389/fpsyg.2013.00333

Sussman, E., Ceponiene, R., Shestakova, A., Näätänen, R., and Winkler, I. (2001). Auditory stream segregation processes operate similarly in school-aged children and adults. *Hear. Res*. 153, 108–114. doi: 10.1016/S0378-5955(00)00261-6

Sussman, E., Ritter, W., and Vaughan, H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 36, 22–34. doi: 10.1017/S0048577299971056

Sussman, E., and Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia* 47, 771–785. doi: 10.1016/j.neuropsychologia.2008.12.007

Sussman, E., Steinschneider, M., Lee, W., and Lawson, K. (2015). Auditory scene analysis in school-aged children with developmental language disorders. *Int. J. Psychophysiol*. 95, 113–124. doi: 10.1016/j.ijpsycho.2014.02.002

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., and Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cogn. Brain Res*. 13, 393–405. doi: 10.1016/S0926-6410(01)00131-8

Sussman, E., Wong, R., Horváth, J., Winkler, I., and Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5-11-year-old children. *Hear. Res.* 225, 117–127. doi: 10.1016/j.heares.2006.12.013

Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife* 2:9. doi: 10.7554/eLife.00699.009

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci*. 31, 164–171. doi: 10.1523/JNEUROSCI.3788-10.2011

Tierney, A., Rosen, S., and Dick, F. (2020). Speech-in-speech perception, nonverbal selective attention, and musical training. *J. Exp. Psychol.* 46, 968–979. doi: 10.1037/xlm0000767

Tóth, B., Kocsis, Z., Háden, G. P., Szerafin, Á., Shinn-Cunningham, B. G., and Winkler, I. (2016). EEG signatures accompanying auditory figure-ground segregation. *NeuroImage* 141, 108–119. doi: 10.1016/j.neuroimage.2016.07.028

Trau-Margalit, A., Fostick, L., Harel-Arbeli, T., Nissanholtz-Gannot, R., and Taitelbaum-Swead, R. (2023). Speech recognition in noise task among children and young-adults: a pupillometry study. *Front. Psychol.* 2023:1188485. doi: 10.3389/fpsyg.2023.1188485

van Loon, A., Knapen, T., Scholte, S., St. John-Saaltink, E., Donner, T., and Lamme, V. (2013). GABA shapes the dynamics of bistable perception. *Curr. Biol.* 23, 823–827. doi: 10.1016/j.cub.2013.03.067

Werner, L., and Bargones, J. (1991). Sources of auditory masking in infants: distraction effects. *Percept. Psychophys.* 50, 405–412. doi: 10.3758/BF03205057

Werner, L. A., Parrish, H. K., and Holmer, N. M. (2009). Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity. *J. Acoust. Soc. Am.* 125, 1040–1049. doi: 10.1121/1.3050254

Wightman, F., Callahan, M., Lutfi, R., Kistler, D., and Oh, E. (2003). Children's detection of pure-tone signals: informational masking with contralateral maskers. *J. Acoust. Soc. Am.* 113, 1–9. doi: 10.1121/1.1570443

Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11812–11815. doi: 10.1073/pnas.2031891100

TYPE Original Research
PUBLISHED 04 April 2024
DOI 10.3389/fnhum.2024.1358380

Check for updates

OPEN ACCESS

EDITED BY
Teija Kujala,
University of Helsinki, Finland

REVIEWED BY
Gábor Péter Háden,
Hungarian Academy of Sciences (MTA),
Hungary
Yan H. Yu,
St. John's University, United States

*CORRESPONDENCE
Jutta L. Mueller
✉ jutta.mueller@univie.ac.at

RECEIVED 19 December 2023
ACCEPTED 15 March 2024
PUBLISHED 04 April 2024

CITATION
Mueller JL, Weyers I, Friederici AD and Männel C (2024) Individual differences in auditory perception predict learning of non-adjacent tone sequences in 3-year-olds.
Front. Hum. Neurosci. 18:1358380.
doi: 10.3389/fnhum.2024.1358380

Individual differences in auditory perception predict learning of non-adjacent tone sequences in 3-year-olds

Jutta L. Mueller[1,2]*, Ivonne Weyers[1], Angela D. Friederici[3] and Claudia Männel[3,4]

[1]Department of Linguistics, University of Vienna, Vienna, Austria, [2]Vienna Cognitive Science Research HUB, Vienna, Austria, [3]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [4]Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Berlin, Germany

Auditory processing of speech and non-speech stimuli oftentimes involves the analysis and acquisition of non-adjacent sound patterns. Previous studies using speech material have demonstrated (i) children's early emerging ability to extract non-adjacent dependencies (NADs) and (ii) a relation between basic auditory perception and this ability. Yet, it is currently unclear whether children show similar sensitivities and similar perceptual influences for NADs in the non-linguistic domain. We conducted an event-related potential study with 3-year-old children using a sine-tone-based oddball task, which simultaneously tested for NAD learning and auditory perception by means of varying sound intensity. Standard stimuli were *A* × *B* sine-tone sequences, in which specific *A* elements predicted specific *B* elements after variable × elements. NAD deviants violated the dependency between *A* and *B* and intensity deviants were reduced in amplitude. Both elicited similar frontally distributed positivities, suggesting successful deviant detection. Crucially, there was a predictive relationship between the amplitude of the sound intensity discrimination effect and the amplitude of the NAD learning effect. These results are taken as evidence that NAD learning in the non-linguistic domain is functional in 3-year-olds and that basic auditory processes are related to the learning of higher-order auditory regularities also outside the linguistic domain.

KEYWORDS
event-related potentials, sequence learning, non-adjacent dependencies, artificial grammar learning (AGL), auditory processing, infant development

Introduction

An important part of auditory cognition is the ability to build relations between sounds that do not occur in direct sequence, but are separated by other intervening sounds. The extraction and processing of such non-adjacent dependencies (NADs) has been suggested to operate both in the auditory-perceptual and linguistic domain (Peña et al., 2002; Gebhart et al., 2009; Bendixen et al., 2012; Mueller et al., 2012; Wilson et al., 2018, for review; Winkler et al., 2018). A specific case from the auditory-perceptual domain that often includes processing of NADs is the phenomenon that listeners need to split rapid sequences of variable sounds into different perceptual streams, depending on various physical characteristics of the stimuli (Bregman and Campbell, 1971). In everyday life, auditory streams which indicate objects or specific events are often interrupted by other sounds. One might hear, for example, songs of different birds and in order to identify the sounds of one specific species, the listener has to extract sequential patterns across intervening sounds from other species. The linguistic domain, which manifests mainly in speech in everyday life, involves the ability to use NADs in similar ways. NADs are basic building blocks of human language and occur in many grammatical constructions, for example, "The **cat** that is strolling on the roofs **is** very old." Such dependencies can occur between syntactic categories (in the example, noun and verb) or between specific morphemes (in the example, the third person and singular – s).

The learning mechanisms that have been suggested to underlie the human ability to extract sequential patterns from auditory input are statistical learning of transitional probabilities and rule-learning supported by non-distributional, perceptional cues (Endress and Bonatti, 2016). Evidence for both accounts will be reviewed to motivate the present investigation, but since this study is not designed to evaluate these two accounts, the general terms "artificial grammar learning" and "NAD learning" will be used in the following without any commitment to a specific mechanism supporting this learning process. Human sequence learning abilities have been shown to be domain-general on the one hand, and tuned to specific input signals, for instance language, on the other hand. Domain-generality is attested by many experiments demonstrating that statistical learning works in the auditory, visual, and even tactile domain (Saffran et al., 1996; Kirkham et al., 2002; Conway and Christiansen, 2005). This shows that in principle, learning takes place across different modalities and materials. Yet, many experiments also attest a modulation of artificial grammar learning processes dependent on the input modality and the specific type of (linguistic) materials (Frost et al., 2015; Milne et al., 2017; Rabagliati et al., 2019; van der Kant et al., 2020; Weyers and Mueller, 2022; Weyers et al., 2022). On top of modality and stimulus differences, individual differences have been found, suggesting, for instance, that statistical learning may be considered a well-defined cognitive ability (Siegelman and Frost, 2015). Other artificial grammar learning studies suggest that interindividual differences in basic auditory abilities may have an impact on performance in auditory artificial grammar learning tasks (Mueller et al., 2012, 2019; Studer-Eichenberger et al., 2016; Vasuki et al., 2017). In light of the mentioned potential domain-specificity of sequence learning abilities in the language domain and the perceptually based individual variation, we here aim to investigate whether the learning of NADs in early childhood, as previously investigated using speech materials (Mueller et al., 2012, 2019), shows similar characteristics in the non-linguistic auditory domain and, crucially, whether learning success shows similar relations to auditory perception as it does for speech materials.

When and how infants start to extract NADs is indicative of how infants approach the problem of acquiring structural knowledge in language. One method to assess whether and how NADs can be extracted by learners of different age groups is the measurement of electrophysiological (EEG) or hemodynamic (functional near-infrared spectroscopy; fNIRS) responses during natural language processing or artificial grammar learning experiments. These studies show that when exposed to sequences of speech sounds, infants are able to remember repetitions of vowels and transitions between adjacent syllables in the first month of life (Teinonen et al., 2009; Benavides-Varela et al., 2011, 2012). Other studies have, moreover, shown that starting at the age of 3 months, infants are also sensitive to more complex NADs between syllables (Friederici et al., 2011; Mueller et al., 2012; Friedrich et al., 2022). Behavioral indicators of successful NAD learning from artificial grammars have been reported from the first birthday (Gòmez and Maye, 2005; Lany and Gòmez, 2008; Culbertson et al., 2016), including the processing of word-internal phonological NAD relations (Gonzalez-Gomez and Nazzi, 2012, 2016). The detection of such structures in infants' native language has been attested behaviorally from around the age of 18 months (Santelmann and Jusczyk, 1998; Hoehle et al., 2006).

When it comes to the question of whether similar learning processes are also present in the non-linguistic auditory domain, behavioral studies provide in principle affirmative evidence, although with important stimulus-dependent variation. Saffran et al. (1999), for instance, showed that 8-month-olds learn transitional probabilities between tones equally easily as those between syllables. Yet, Marcus et al. (2007) observed that 7-month-old infants could learn a repetition rule instantiated with tones and animal sounds only when the rule was presented after sequences of speech sounds. At the neurophysiological level, event-related potential (ERP) studies within the first half year of life indicate that the ability to extract adjacent as well as non-adjacent regularities is present both for speech (Teinonen et al., 2009; Friederici et al., 2011; Mueller et al., 2012) and non-speech auditory stimuli (Kudo et al., 2011; Winkler et al., 2018). Yet, there is evidence that the sensitivity to specific types of stimuli presented during such learning tasks is subject to developmental changes across the first years of life. Several studies have shown that very young infants in the first half year of life learn dependencies that are not, or not so easily, learned at later ages (Dawson and Gerken, 2009; Mueller et al., 2012, 2019). A study using fNIRS showed different trajectories for learning NADs from linguistic and non-linguistic sequences (van der Kant et al., 2020). While NADs were learned at age 2, but not at age 3 when presented in the form of artificial linguistic stimuli, NADs were learned successfully at 3 years of age, but not yet at 2 years of age, when presented in the form of non-linguistic tone sequences (van der Kant et al., 2020). This apparent decline in the ability to extract linguistic NADs around age 3 was further confirmed in an ERP study using the same linguistic stimulus material (Paul et al., 2021). On the basis of these findings, one may speculate that the learning of NADs in the non-linguistic domain may be present across development and not decline as it does in the speech domain. Yet, the scarcity of developmental data does not allow for firm conclusions about the developmental trajectory of NAD learning in linguistic versus non-linguistic auditory domains.

While it may be the case that the learning of statistical regularities such as NADs is specifically honed to the speech signal during early development, there is ample evidence that basic auditory processes impact on the processing and learning of grammatical dependencies, as evidenced in studies both with impaired and unimpaired populations (Halliday and Bishop, 2006; Bishop, 2007; Arciuli and Simpson, 2012; Mueller et al., 2012; Kidd and Arciuli, 2016). More specifically, children with language impairments (e.g., developmental language disorders or developmental dyslexia) often display deficits also with respect to the processing of basic auditory parameters, as for example duration processing (Corriveau et al., 2007) or frequency discrimination (Hill et al., 2005; Halliday and Bishop, 2006). For example, a longitudinal study with children with and without family history of language impairment showed that early differences (from 6 to 48 months) in the maturation of auditory evoked potentials in response to non-linguistic auditory stimuli predicted language abilities at 3 and 4 years of age (Choudhury and Benasich, 2011). What is more, children and adults with developmental language disorders show difficulties in the domain of auditory statistical learning in particular (Evans et al., 2009; Kahta and Schiff, 2019). At the other end of the continuum, it has been shown that musically trained children, who possess particularly rich auditory experiences, outperform untrained children in some language and statistical learning tasks (Kraus and Chandrasekaran, 2010; Francois and Schön, 2011; Vasuki et al., 2017). Taken together, the reviewed evidence suggests that both speech and non-speech sequence learning are linked to the quality of auditory processing.

While many studies have examined how the processing of specific acoustic parameters, for example, frequency (Halliday and Bishop, 2006; Mueller et al., 2012; Halliday et al., 2014) or duration (Weber et al., 2005) are linked to language processing, it remains unclear whether the relation between the quality of auditory processing and higher order cognitive processing holds for non-linguistic sequence learning in a comparable way. On the one hand, it is possible that specific auditory parameters are predictive of language abilities due to their importance for coding linguistic information (for example, frequency information determining vowel quality, or duration parameters contributing to word stress). For example, in the study of Mueller et al. (2012), pitch perception was found to be related to the learning of NADs between syllables. As frequency is not merely an auditory parameter, but also involved in how different vowels are coded, the link between auditory perception and NAD learning in this study could be explained because both processes are based on frequency discrimination. In this line of argumentation, auditory parameters that do not code distinctive features important for dependency learning should thus not be related to sequence learning. On the other hand, it is conceivable that rather domain-general mechanisms, for instance bottom-up auditory attention (Kaya and Elhilali, 2014; Addleman and Jiang, 2019), form the common basis for the processing of both speech and non-speech structured sequential inputs. Such domain-general attentional processes might be necessary for both rather low-level auditory processing and higher-level speech-based and non-speech-based processing of auditory sequences. Individual differences with respect to these attentional processes may affect neural correlates of processing in both domains. Under such a domain-general account, a variety of auditory parameters that render stimuli salient and attract listeners' attention should be linked to sequence learning not only in the speech domain, but also in the non-speech domain. An example for such a parameter would be a change in intensity in sequences in which intensity is not relevant for the detection of a sequential regularity that has to be learned.

In the present study, we constructed NADs between non-linguistic pure tones and presented them in an oddball design, similarly to how Mueller et al. (2012, 2019) tested for syllable-based NAD learning. Using electroencephalography, we examined 3-year-old children's ability to differentiate frequently presented standard exemplars of these tone NADs from infrequent deviant patterns. Three-year-olds were deemed a particularly relevant research population, because a beginning decline in NAD learning for speech stimuli had previously been reported for this age group (Mueller et al., 2019; van der Kant et al., 2020; Paul et al., 2021). As we specifically aimed to explore a possible impact of auditory processing abilities on non-speech NAD learning, we presented auditory deviants, which were reduced in intensity, in addition to NAD deviants within the same continuous stimulus stream. We hypothesized (i) that 3-year-olds would still be able to learn non-linguistic NADs, given a possible language- or speech-specificity of the previously established developmental decline in NAD learning ability (van der Kant et al., 2020). In the present experimental design, successful learning can be inferred from a reliable mismatch response (MMR) in the ERP with either positive or negative polarity. We further expected (ii) that successful NAD learning would be indexed by similar ERP patterns as reported for the speech domain, because a negative MMR for violated NADs compared to standard sequences was found in Mueller et al. (2019), in a design comparable to the present study; and (iii) that this ERP effect would be modulated by the individual ability to detect acoustic changes in the stimuli (cf. Mueller et al., 2012). In the present study, we chose to test intensity changes in order to test the above outlined possibility that domain-general, stimulus-driven, attentional mechanisms may be sufficient to explain a potential link between auditory processing and NAD learning. We expected this link to be indicated by a predictive relationship between the amplitude of the intensity-related MMR and the amplitude of the NAD-related MMR. Such a relationship should be specific to auditory change discrimination and not just reflect unspecific interindividual differences in auditory processing *per se*, which have been identified as an important source of variance across different perceptual and cognitive domains in ERP research. Specifically for auditory stimuli, it has been shown that amplitudes of the evoked signals vary considerably across individuals (Melnik et al., 2017). To control for an impact of such unspecific interindividual variation, we also tested for a relation between amplitude variations in response to standard stimuli alone and the NAD learning effects. Such a control will allow for more compelling evidence in case of a relationship between the auditory discrimination effect and the NAD learning effect.

Materials and methods

Participants

Our study was approved by the Ethics Committee of the University of Leipzig and conformed to the guidelines of the Declaration of Helsinki (World Medical Association, 2013). Before the experiment, parents and children were verbally informed about the test procedure and the accompanying caregiver gave written informed consent. All participants had normal hearing and no history of neurological disorders, were born between the 38th and 42nd week of gestation and were German monolinguals.

Of the 49 infants who participated in the study, data from 23 children (15 female) were excluded from analysis for the following reasons: children did not want to participate in the EEG measurement or wanted to stop it before the experiment ended ($n = 7$), data loss because of experimenter error ($n = 1$), children had been diagnosed with or were at risk for a developmental disorder ($n = 2$), high artifact rate ($n = 13$). The remaining 26 participants (7 female) had a mean age of 36.8 months (SD = 0.4).

Experimental design and stimuli

The study made use of a modified oddball paradigm with triplets of sine tones. As illustrated in **Figure 1**, standard triplets conformed to an $A \times B$ grammar of item-specific NADs, in which the first tone A of a triplet reliably predicted the third tone B, while the intervening tone × varied. The individual tones covered a frequency range of 600–1,750 Hz in steps of 50 Hz and each tone had a duration of 100 ms. Two different $A \times B$ dependencies were used. The first type, $A_1 \times B_1$, comprised an NAD between a 600 and a 1,450 Hz tone, the second type, $A_2 \times B_2$, an NAD between a 1,750 and a 900 Hz tone. The middle × element was filled with the remaining set of 20 different tones (650, 700, 750, 800, 850, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 1,250, 1,300, 1,350, 1,400, 1,500, 1,550, 1,600, 1,650, and 1,700 Hz) in a pseudorandom manner. Within each $A \times B$ triplet, tones were separated by 50 ms pauses, and between triplets there was a pause of 700 ms. A total of 80% of trials were standard stimuli, 10% were NAD deviants in which the NAD between A and B was violated (e.g., A_1 was incorrectly paired with B_2), and another 10% of trials were acoustic deviants, which were characterized by an intensity decrease of 25% (reduction from 82.27 to 62.83 dB).

Stimuli were presented according to a randomization scheme ensuring that each type of $A \times B$ triplet occurred with equal frequency in a given series of standards and that the same $A \times B$ triplet was not repeated more than three times in a row. Between two deviants, sequences consisting of two, four, six, or eight standard stimuli occurred. When standard sequences were shorter than four standards, different deviant types occurred before and after in order to provide enough exemplars to re-establish the correct $A \times B$ dependency. In total, 818 stimuli were presented, with 658 standard stimuli conforming to the $A \times B$ dependency and 80 exemplars of NAD deviants and intensity deviants each.

Procedure

During stimulus presentation, children sat on their caregiver's lap in a sound-attenuated testing booth. Caregivers wore sound-attenuating earplugs. Stimuli were played via two loudspeakers at the same comfortable sound level across participants. During the experiment, children were presented with a silent animation movie in order to prevent excessive movements of body and head (for a similar procedure, see Männel and Friederici, 2011; Männel et al., 2013). The whole experiment took approximately 14 min.

EEG recording and analysis

The continuous electroencephalogram (EEG) was recorded from 23 Ag/AgCl monopolar electrodes fixed in an elastic cap placed on the child's head (EASYCAP GmbH, Germany) at the following standard positions corresponding to the international 10–20 system (Jasper, 1958): F7, F3, Fz, F4, F8, FC3, FC4, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, O1, O2, M1, and M2. Signal recorded from the electrode positions F9 and F10 served to calculate the horizontal electrooculogram. The vertical electrooculogram was recorded from the electrode FP2 and an additional electrode placed below the right eye. FP1 served as the ground electrode, and Cz as online reference. Electrode impedances were largely kept below 10 kΩ (at least below 20 kΩ). The EEG signal was amplified with a gain of 20 using a PORT-32/MREFA amplifier (Twente Medical Systems International B.V.) with an input impedance of 1,012 Ω, and digitized online at 500 Hz (AD converter with 22 bit, digital filter from DC to 125 Hz). The following pre-processing steps were performed using MATLAB (version R2018b; The MathWorks Inc., 2018) and the EEGLAB open source toolbox (version 14.1.1b; Delorme and Makeig, 2004). Offline, the EEG was re-referenced to the average of both mastoids (M1 and M2). The continuous EEG was epochized from −200 to 800 ms relative to the onset of the final tone in each triplet and band-pass filtered between 0.5 and 30 Hz (digital windowed sinc FIR-filter, −6 dB half-amplitude cut-off, cut-off frequencies of 0.6 and 29.99 Hz). For the correction of eye-movements, the data were band-pass filtered between 1 and 30 Hz (−6 dB, cut-off frequencies of 1.06 and 29.96 Hz) and submitted to an ICA. After removing the ICA components related to eye-movements based on subjective evaluation, the ICA weights from this dataset were applied to dataset with the 0.5–30 Hz band-pass filter. In an additional step, any remaining artifacts were rejected manually. ERPs were calculated for the epochs with a pre-stimulus baseline of 100 ms for each condition separately, i.e., for standards stimuli, intensity deviants and NAD deviants.

Only children with a minimum of 20 deviant trials (25%) per condition remaining were included in the statistical analysis. The resulting mean number of artifact-free trials was identical across deviant conditions (NAD deviants: $M = 36$; SD = 3; intensity deviants: $M = 36$; SD = 4; standards: $M = 216$; SD = 34). Finally, a 10-Hz low-pass filter was applied for data visualization only.

Statistical analysis

The statistical analyses of the overall effects of the NAD and intensity violations were conducted using the FieldTrip toolbox (Version 20170228; Oostenveld et al., 2011). Additional regression analyses were implemented using the statistical computing software R, Version 4.3.1 (R Core Team, 2022). As a first step, non-parametric, cluster-based permutation tests using dependent-sample t-tests were performed for each of the two experimental conditions, comparing ERP responses to the final tones of standard triplets to those in response to NAD deviants and intensity deviants, respectively. The entire epoch (0–800 ms) and all electrodes (except reference and EOG electrodes) were included in the analysis. In order to be included in a spatial cluster, a minimum

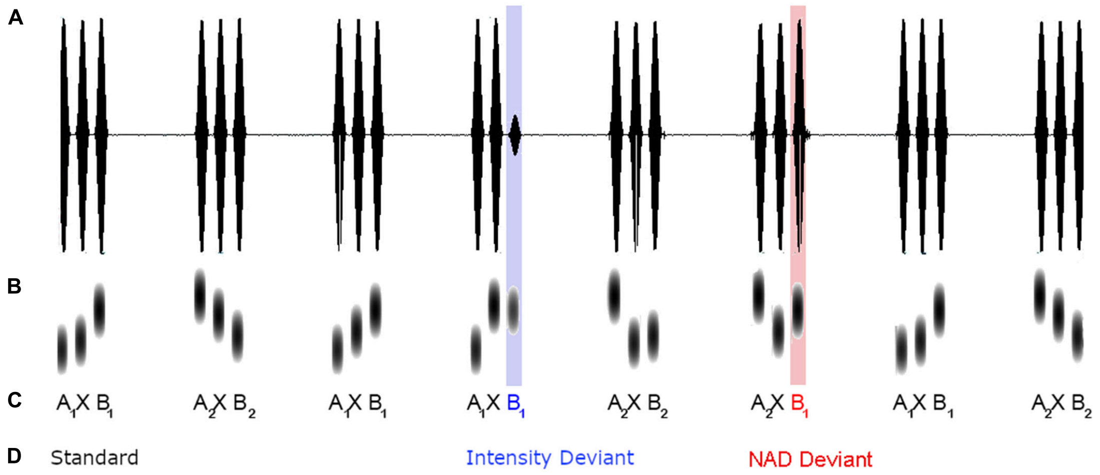


FIGURE 1
Schematic illustration of intensity **(A)**, frequency **(B)**, sequential structure (**C**, two types of NAD are indexed with the numbers 1 and 2 in the subscript) and condition of tone stimuli **(D)** embedded within an example of a presentation sequence.

of two significant neighboring electrodes was set, with spatial neighborhood defined based on the triangulation method. The statistical threshold for a specific time-sample to be included in a cluster was $p < 0.05$. The cluster-statistic permutation test (maxsum approach) was conducted using 1,000 draws from the permutation distribution via Monte-Carlo simulation and an alpha level of $p < 0.05$ (two-tailed).

As a second step, we performed step-wise regression analyses to test whether an identified NAD effect was related to the intensity effect. *Post hoc*, we extracted the mean amplitude values of the NAD and intensity effects for the indicated approximate time window of the dependency effect (195–329 ms) from seven frontal electrodes (F7, F3, Fz, F4, F8, FC3, and FC4) which were identified as part of the clusters in both NAD and intensity conditions. In order to ensure that any such relation was between NAD learning and intensity discrimination specifically and not grounded in interindividual differences with respect to amplitudes of auditory-related ERPs *per se* (which might affect both the NAD and intensity effects), we included the amplitude of the ERP response to standard stimuli as an additional predictor. This amplitude was extracted from the same time window as was used for the NAD and the intensity effect (195–329 ms).

Results

Figure 2 displays the ERP waveforms in response to the final tones of the triplets. The ERPs show a negative peak immediately after the onset of the final tone, which is followed by a short positive peak and a somewhat longer-lasting negativity. Before the baseline period (−100 to 0 ms), another negative peak is visible followed by a short positive peak, which can be related to the presentation of the second tone in the triplet. The short sequence of negative peaks accompanying the rapid presentation of single tones suggests that each tone elicits a characteristic ERP response consisting of obligatory auditory components appropriate for this age group (Ceponiene et al., 2002). Note, however, that these ERP responses to individual tones cannot be disentangled due to our short SOAs of 150 ms (from onset to onset). Importantly, the second tones, for which the responses are visible in the baseline, were similarly distributed from a set of 20 different tones across standard sequences as well as intensity and NAD deviant sequences.

NAD learning and intensity effects

Averaged ERPs across fronto-central electrode sites show a positive deflection for NAD deviants compared to standard stimuli (cf. Figure 2). The non-parametric cluster-based permutation test revealed this condition difference to be statistically significant (clusterstat $T = 2.186$, $p < 0.05$) within a time window ranging from approximately 195–326 ms relative to the onset of the last tone of the triplet (Figure 2A). For the comparison of standards versus intensity deviants, a similar, somewhat more broadly distributed positivity is visible at fronto-central electrodes. This condition difference is also confirmed statistically by a cluster-based permutation test (clusterstat $T = 12.933$, $p < 0.01$) (cf. Figure 2B) for an approximate time window of 142–479 ms relative to the onset of the last tone in the triplet.

Regression analysis

The step-wise regression revealed that only the intensity effect predicted the NAD effect, while the amplitude in response to the standard stimuli alone did not. Akaike information criterion (AIC) was used to determine whether a factor significantly contributed to the predictive power of the model. Results revealed that of the tested models, the model including the intensity effect as a sole predictor was an at least equal fit to explain the data (AIC 37.62) compared to the full model (AIC 39.44), which included both intensity effect and standard amplitudes, and a significantly better fit compared to the model with only the standard amplitude (AIC 42.2). In the final model, the amplitude of the intensity effect was

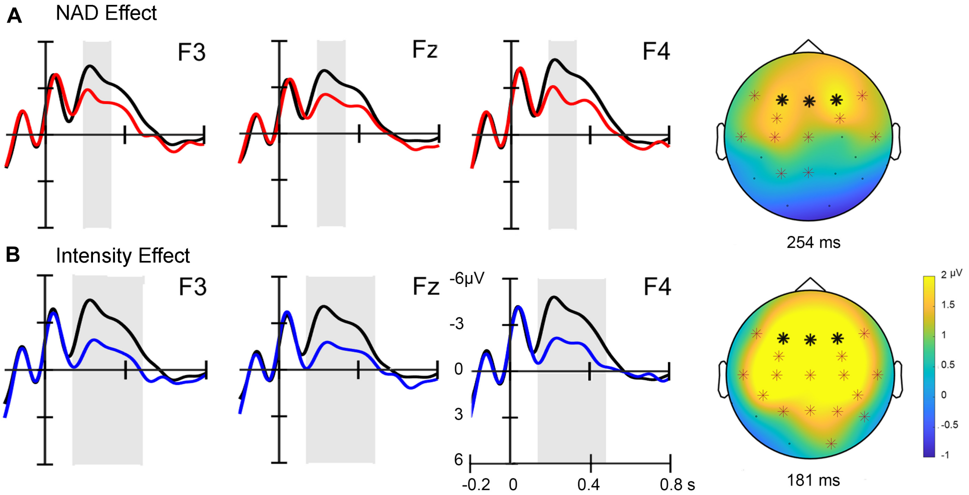


FIGURE 2

(A) Event-related potential (ERP) effects for NAD deviants (red lines) compared to standard stimuli (black lines) at three frontal electrode sites (marked with bold stars in the topographies). A topographic difference map on the right displays the positivity at its maximum with electrode sites included in the cluster marked with stars. **(B)** ERP effects for intensity deviants (blue lines) compared to standard stimuli (black lines) at three frontal electrode sites. A topographic difference map displays the positivity at its maximum with significant electrode sites marked. The sites of the electrodes included in the cluster are marked with stars.

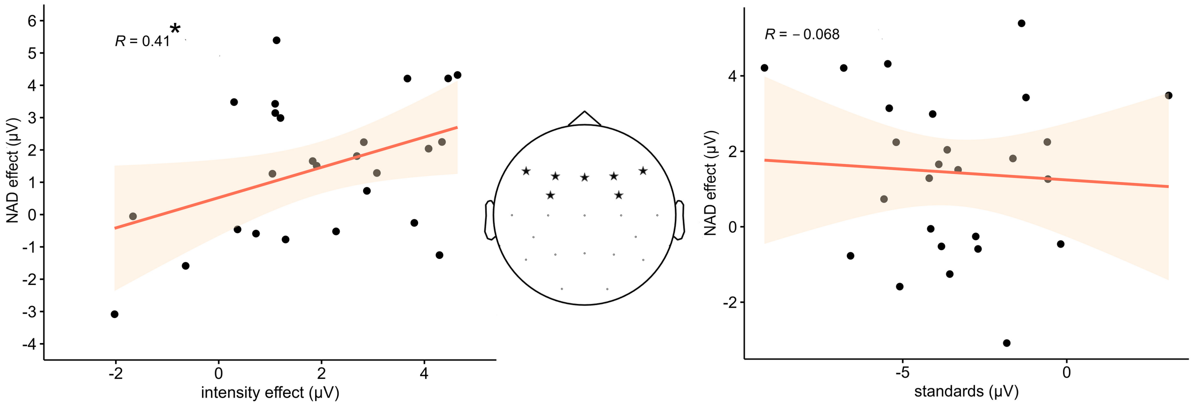


FIGURE 3

(Left) Statistically significant positive relation between the difference amplitude values of the intensity effect and the NAD effect in the ERP data. **(Middle)** Electrode sites from which the amplitude values were extracted. **(Right)** No statistically significant relation between the ERP amplitude of the standard stimuli and the NAD effect.

a significant predictor of the NAD effect amplitude [$F(1,24) = 4.76$, $p < 0.05$], with an explained variance of $R^2 = 0.13$. Figure 3 displays scatterplots of the results including Pearson correlation coefficients.

Discussion

The goal of the present study was to evaluate whether NAD learning could be observed in 3-year-old infants for non-linguistic tone sequences, similarly to the processing of linguistic sequences reported previously (Mueller et al., 2019). Additionally, the study probed the relationship between basic auditory processing ability in terms of intensity variation detection and NAD learning. Here, we found that learning of non-adjacent tone sequences is present in 3-year-olds. Yet, in contrast to our original hypothesis derived from findings of linguistic NAD learning at this age, successful learning and deviant detection was evidenced by a positive, not a negative, deflection in the ERP response. Further, the present data suggest that non-linguistic NAD learning is linked to intensity discrimination, similarly to how linguistic NAD learning has been linked to auditory pitch discrimination (Mueller et al., 2012). As a note of caution, given that our sample comprised more male than female participants, we cannot exclude that this bias influenced our findings.

The successful learning of non-linguistic NADs in the present study is consistent with the findings of van der Kant et al. (2020), who reported fNIRS evidence for the learning of NADs in tone sequences, but not in linguistic sequences, in 3-year-olds. Relatedly, two ERP studies reported that ERP indices of linguistic NAD learning decreased in amplitude between the first birthday and the age of 4 years (Mueller et al., 2019; Paul et al., 2021). Given that for the present study, no longitudinal data is available, it remains an open question whether the observed NAD learning effect for non-linguistic auditory stimuli would also decrease with increasing age. Initially, it may seem counterintuitive that sine tones, which are much less relevant and appealing for humans than speech (Vouloumanos and Werker, 2004), yield a more robust learning effect than species-specific communicative sounds. A whole range of experiments with younger infants has indeed suggested the opposite, namely an advantage of speech stimuli (and other communicative signals) over non-speech stimuli for the learning of repetition-based regularities (Ferguson and Lew-Williams, 2016; Rabagliati et al., 2019). We argue here that by the age of 3 years, language and its basic regularities may be so familiar already, and potentially even entrenched, that identifying and learning new language dependencies may require much more effort compared to the acquisition of new dependencies in a non-familiar sound system. Entrenchment, that is, the reduced plasticity of a system after it has settled in a stable state due to extensive exposure, has been brought up as a mechanism that applies to human statistical learning (Bulgarelli and Weiss, 2016; Siegelman et al., 2018) and as a potential explanation for some of the difficulties second language learners experience (MacWhinney, 2016). Thus, NADs in sine tone sequences might be easy to learn for 3-year-olds because they do not interfere with an entrenched rule system that is used for communication. Yet, whether previous experience plays a decisive role in explaining the auditory signal-specific differences modulating the efficiency of NAD learning in our study has to be tackled by further research.

Notably, we expected NAD learning/deviancy detection to be indexed by a negative-going ERP effect as our experiment was closely modeled after Mueller et al. (2019), who reported negative responses for linguistic NAD deviants compared to standards for both 2-year-olds and 4-year-olds. Yet, in our study, both deviant conditions yielded positive-going responses with similar fronto-central distributions. Although the polarity of these responses was unexpected for the current design, they still indexed children's successful detection of both acoustic and NAD changes. Moreover, positive-going responses for deviancy detection have previously been reported in similar oddball experiments with infants during their first year of life (Friederici et al., 2011; Mueller et al., 2012; Weyers et al., 2022). In addition to younger children often showing positivities and older children negativities in deviancy detection (e.g., Pihko et al., 1999; Cheng et al., 2013, 2015; Reh et al., 2021), several stimulus and experimental factors seem to influence the polarity of the deviancy response. For example, the type and acoustic distance of the tested sound contrasts (e.g., Morr et al., 2002; Cheng et al., 2013, 2015), experimental conditions (e.g., duration of inter-stimulus-interval; Cheour et al., 2002; Kushnerenko et al., 2002), and data-analysis decisions (e.g., Weber et al., 2004; He et al., 2007) have been reported to impact on the polarity of the deviancy response. For the current study, differences in the properties of non-linguistic and linguistic stimuli and the way NADs were encoded are particularly relevant. Here, previous infant studies reported positivities for deviancy detection of smaller stimulus differences and negativities for larger differences (Cheng et al., 2013, 2015). However, quantitative aspects of the difference between correct and incorrect NADs across tone and speech stimuli are difficult to judge. Furthermore, in studies with linguistic stimuli, positive responses have been proposed to reflect acoustic processing and negativities rather phonological processing (Rivera-Gaxiola et al., 2005; Garcia-Sierra et al., 2016; Ferjan Ramírez et al., 2017). Taken together, non-linguistic and linguistic stimuli are likely to trigger different processing levels in listeners' deviancy detection, resulting in positive responses for both acoustic and NAD changes in pure-tone sequences in the current study.

As hypothesized, we found a predictive relation between the ERP effect reflecting intensity discrimination and the ERP effect reflecting the detection of the NAD deviant. Similar links were reported previously for linguistic NAD learning in 3-month-olds, 4-year-olds, and adult participants, whereby larger responses for pitch discrimination were positively related to the responses for the detection of NAD violations (Mueller et al., 2012, 2019). These studies left unclear, however, whether this relation is specific to the frequency domain, or whether it is indicative of a more general effect linking auditory perception and sequence learning. In our study, the NADs were coded by pitch, while intensity was held constant across all standard stimuli. The feature intensity was thus not relevant for the extraction of the regularity. Nonetheless, participants' response for intensity discrimination was predictive of their response for NAD violation detection. Interindividual differences in the amplitude of the ERP response to the standard stimuli alone did not contribute to this link. This makes it unlikely that unspecific interindividual differences with respect to the measured ERP amplitudes (Melnik et al., 2017) explain the effects we found. The present results suggest that aspects of auditory processing ability, namely the ability to discriminate sounds based on their intensity, is predictive of the ability to detect NADs in tone sequences. We interpret this as indicative of a general link between auditory perception and stimulus-driven learning. What could it be that links the two? Since intensity discrimination and the detection of tone sequence patterns are not intrinsically related, we suggest that both are linked via a shared process, namely bottom-up auditory attention (Addleman and Jiang, 2019). It has been suggested that stimulus-driven auditory attention is affected by a variety of internal and external factors, which contribute to the salience of a stimulus over space and time. Specifically, pitch and intensity have shown to contribute to this process (Kaya and Elhilali, 2014; Addleman and Jiang, 2019). Following this, our data suggest that there are interindividual differences in how strongly children react to stimulus-driven saliency across different perceptual domains – namely intensity and pitch modulation over time (which could be one way to describe the NADs in the current study in which pitch patterned in complex ways across three items). The underlying link could thus be an attention-based mechanism, modulated by the sensitivity to respond to salient auditory stimuli, be it on the basis of the temporary state the child is in at the particular moment of testing, or a more general individual ability.

Conclusion

In sum, the current study provided ERP evidence for the ability of 3-year-old children to extract NADs from sine tone sequences. Although the specific ERP pattern (positivity) indicating learning success contrasts with evidence from similar studies from the linguistic domain, the evidence is consistent with the idea that NAD learning in early childhood is present across the linguistic and the non-linguistic domain. Further, we were able to extend the finding of a relation between auditory processing and NAD learning to the domain of non-linguistic tones. We propose that this is due to a shared attentional mechanism linking perception and NAD learning in the present design. Further studies will have to test the pervasiveness of such a potential linking mechanism across modalities and types of input.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Ethics Advisory Board of the Medical Faculty of the University of Leipzig. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

JM: Conceptualization, Data curation, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft. IW: Methodology, Writing – review & editing. AF: Conceptualization, Resources, Writing – review & editing. CM: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Open Access Funding provided by University of Vienna.

Conflict of interest

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

Publisher's note

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

References

Addleman, D. A., and Jiang, Y. V. (2019). Experience-driven auditory attention. *Trends Cogn. Sci.* 23, 927–937. doi: 10.1016/j.tics.2019.08.002

Arciuli, J., and Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cogn. Sci.* 36, 286–304. doi: 10.1111/j.1551-6709.2011.01200.x

Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A. H., Peretz, I., and Mehler, J. (2011). Memory in the neonate brain. *PLoS One* 6:e27497. doi: 10.1371/journal.pone.0027497

Benavides-Varela, S., Hochmann, J.-R., Macagno, F., Nespor, M., and Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17908–17913. doi: 10.1073/pnas.1205413109

Bendixen, A., Schröger, E., Ritter, W., and Winkler, I. (2012). Regularity extraction from non-adjacent sounds. *Front. Psychol.* 3:19473. doi: 10.3389/FPSYG.2012.00143/ABSTRACT

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going? *Psychol. Bull.* 133, 651–672. doi: 10.1037/0033-2909.133.4.651

Bregman, A. S., and Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J. Exp. Psychol.* 89, 244–249. doi: 10.1037/H0031163

Bulgarelli, F., and Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *J. Exp. Psychol.* 42, 1621–1631. doi: 10.1037/XLM0000263

Ceponiene, R., Rinne, T., and Näätänen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clin. Neurophysiol.* 113, 870–882. doi: 10.1016/S1388-2457(02)00078-0

Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., and Lee, C. Y. (2013). The development of mismatch responses to mandarin lexical tones in early infancy. *Dev. Neuropsychol.* 38, 281–300. doi: 10.1080/87565641.2013.799672

Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., and Lee, C. Y. (2015). Feature-specific transition from positive mismatch response to mismatch negativity in early infancy: Mismatch responses to vowels and initial consonants. *Int. J. Psychophysiol.* 96, 84–94. doi: 10.1016/J.IJPSYCHO.2015.03.007

Cheour, M., Martynova, O., Naatanen, R., Erkkola, R., Sillanpaa, M., Kero, P., et al. (2002). Speech sounds learned by sleeping newborns. *Nature* 415, 599–600. doi: 10.1038/415599b415599b

Choudhury, N., and Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clin. Neurophysiol.* 122, 320–338. doi: 10.1016/j.clinph.2010.05.035

Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol.* 31, 24–39. doi: 10.1037/0278-7393.31.1.24

Corriveau, K., Pasquini, E., and Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *J. Speech Lang. Hear. Res.* 50, 647–666. doi: 10.1044/1092-4388(2007/046)

Culbertson, J., Koulaguina, E., Gonzalez-Gomez, N., Legendre, G., and Nazzi, T. (2016). Developing knowledge of nonadjacent dependencies. *Dev. Psychol.* 52, 2174–2183. doi: 10.1037/dev0000246

Dawson, C., and Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-Month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition* 111, 378–382. doi: 10.1016/j.cognition.2009.02.010

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

Endress, A. D., and Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 19–35. doi: 10.1002/wcs.1376

Evans, J. L., Saffran, J. R., and Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *J. Speech Lang. Hear. Res.* 52, 321–335. doi: 10.1044/1092-4388(2009/07-0189

Ferguson, B., and Lew-Williams, C. (2016). Communicative signals support abstract rule learning by 7-month-old infants. *Sci. Rep.* 6:25434. doi: 10.1038/srep25434

Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., and Kuhl, P. K. (2017). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Dev. Sci.* 20:e12427. doi: 10.1111/DESC.12427

Francois, C., and Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* 21, 2357–2365. doi: 10.1093/CERCOR/BHR022

Friederici, A. D., Mueller, J. L., and Oberecker, R. (2011). Precursors to natural grammar learning: Preliminary evidence from 4-month-old infants. *PLoS One* 6:e17920. doi: 10.1371/journal.pone.0017920

Friedrich, M., Mölle, M., Born, J., and Friederici, A. D. (2022). Memory for nonadjacent dependencies in the first year of life and its relation to sleep. *Nat. Commun.* 13:7896. doi: 10.1038/S41467-022-35558-X

Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.* 19, 117–125. doi: 10.1016/j.tics.2014.12.010

Garcia-Sierra, A., Ramírez-Esparza, N., and Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *Int. J. Psychophysiol.* 110, 1–17. doi: 10.1016/J.IJPSYCHO.2016.10.004

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychon. Bull. Rev.* 16, 486–490. doi: 10.3758/PBR.16.3.486

Gòmez, R. L., and Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy* 7, 183–206.

Gonzalez-Gomez, N., and Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy* 17, 498–524. doi: 10.1111/j.1532-7078.2011.00104.x

Gonzalez-Gomez, N., and Nazzi, T. (2016). Delayed acquisition of non-adjacent vocalic distributional regularities. *J. Child Lang.* 43, 186–206. doi: 10.1017/S0305000915000112

Halliday, L. F., and Bishop, D. V. M. (2006). Auditory frequency discrimination in children with dyslexia. *J. Res. Read.* 29, 213–228. doi: 10.1111/j.1467-9817.2006.00286.x

Halliday, L. F., Barry, J. G., Hardiman, M. J., and Bishop, D. V. (2014). Late, not early mismatch responses to changes in frequency are reduced or deviant in children with dyslexia: An event-related potential study. *J. Neurodev. Disord.* 6:21. doi: 10.1186/1866-1955-6-21

He, C., Hotson, L., and Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *J. Cogn. Neurosci.* 19, 878–892. doi: 10.1162/JOCN.2007.19.5.878

Hill, P. R., Hogben, J. H., and Bishop, D. M. V. (2005). Auditory frequency discrimination in children with specific language impairment: A longitudinal study. *J. Speech Lang. Hear. Res.* 48, 1136–1146. doi: 10.1044/1092-4388(2005/080)

Hoehle, B., Schmitz, M., Santelmann, L. M., and Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Lang. Learn. Dev.* 2, 277–300. doi: 10.1207/s15473341lld0204_3

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.

Kahta, S., and Schiff, R. (2019). Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia* 25, 142–157. doi: 10.1002/DYS.1618

Kaya, E. M., and Elhilali, M. (2014). Investigating bottom-up auditory attention. *Front. Hum. Neurosci.* 8:85912. doi: 10.3389/fnhum.2014.00327

Kidd, E., and Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Dev.* 87, 184–193. doi: 10.1111/cdev.12461

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83, B35–B42. doi: 10.1016/S0010-0277(02)00004-5

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Kudo, N., Nonaka, Y., Mizuno, N., Mizuno, K., and Okanoya, K. (2011). On-line statistical segmentation of a non-speech auditory stream in neonates as demonstrated by event-related brain potentials. *Dev. Sci.* 14, 1100–1106. doi: 10.1111/j.1467-7687.2011.01056.x

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., and Naatanen, R. (2002). Maturation of the auditory change detection response in infants: A longitudinal ERP study. *Neuroreport* 13, 1843–1848.

Lany, J., and Gòmez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol. Sci.* 19, 1247–1252. doi: 10.1111/j.1467-9280.2008.02233.x

MacWhinney, B. (2016). "Entrenchment in second-language learning," in *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*, ed. H. Schmid (Washington, DC: American Psychological Association), 343–366. doi: 10.1037/15969-016

Männel, C., and Friederici, A. D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Dev. Sci.* 14, 786–798. doi: 10.1111/j.1467-7687.2010.01025.x

Männel, C., Schipke, C. S., and Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3-and 6-year-olds. *Dev. Cogn. Neurosci.* 5, 86–94. doi: 10.1016/j.dcn.2013.01.003

Marcus, G. F., Fernandes, K. J., and Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychol. Sci.* 18, 387–391. doi: 10.1111/j.1467-9280.2007.01910.x

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., et al. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data? *Front. Hum. Neurosci.* 11:245570. doi: 10.3389/FNHUM.2017.00150

Milne, A. E., Petkov, C. I., and Wilson, B. (2017). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience* 389, 104–117. doi: 10.1016/j.neuroscience.2017.06.059

Morr, M. L., Shafer, V. L., Kreuzer, J. A., and Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear Hear.* 23, 118–136.

Mueller, J. L., Friederici, A. D., and Männel, C. (2012). Auditory perception at the root of language learning. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15953–15958. doi: 10.1073/pnas.1204319109

Mueller, J. L., Friederici, A. D., and Männel, C. (2019). Developmental changes in automatic rule-learning mechanisms across early childhood. *Dev. Sci.* 22:e12700. doi: 10.1111/desc.12700

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011. doi: 10.1155/2011/156869

Paul, M., Männel, C., van der Kant, A., Mueller, J. L., Höhle, B., Wartenburger, I., et al. (2021). Gradual development of non-adjacent dependency learning during early childhood. *Dev. Cogn. Neurosci.* 50:100975. doi: 10.1016/J.DCN.2021.100975

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* 298, 604–607. doi: 10.1126/science.1072901

Pihko, E., Leppänen, P. H. T., Eklund, K. M., Cheour, M., Guttorm, T. K., and Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: I. Age effects. *Neuroreport* 10, 901–905. doi: 10.1097/00001756-199904060-00002

R Core Team (2022). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Rabagliati, H., Ferguson, B., and Lew-Williams, C. (2019). The profile of abstract rule learning in infancy: Meta-analytic and experimental evidence. *Dev. Sci.* 22:e12704. doi: 10.1111/desc.12704

Reh, R. K., Hensch, T. K., and Werker, J. F. (2021). Distributional learning of speech sound categories is gated by sensitive periods. *Cognition* 213:104653. doi: 10.1016/J.COGNITION.2021.104653

Rivera-Gaxiola, M., Silva-Pereyra, J., and Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7–and 11-month-old American infants. *Dev. Sci.* 8, 162–172. doi: 10.1111/J.1467-7687.2005.00403.X

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Santelmann, L. M., and Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition* 69, 105–134. doi: 10.1016/S0010-0277(98)00060-2

Siegelman, N., and Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *J. Mem. Lang.* 81, 105–120. doi: 10.1016/j.jml.2015.02.001

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., and Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition* 177, 198–213. doi: 10.1016/j.cognition.2018.04.011

Studer-Eichenberger, E., Studer-Eichenberger, F., and Koenig, T. (2016). Statistical learning, syllable processing, and speech production in healthy hearing and hearing-impaired preschool children. *Ear Hear.* 37, e57–e71. doi: 10.1097/AUD.0000000000000197

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10:21. doi: 10.1186/1471-2202-10-21

The MathWorks Inc. (2018). MATLAB version: 9.5 (R2018b). Natick, MA: The MathWorks Inc. Available online at: https://www.mathworks.com van der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., and Wartenburger, I. (2020). Linguistic and non-linguistic non-adjacent dependency learning in early development. *Dev. Cogn. Neurosci.* 45:100819. doi: 10.1016/J.DCN.2020.100819

Vasuki, P. R. M., Sharma, M., Ibrahim, R., and Arciuli, J. (2017). Statistical learning and auditory processing in children with music training: An ERP study. *Clin. Neurophysiol.* 128, 1270–1281. doi: 10.1016/j.clinph.2017.04.010

Vouloumanos, A., and Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Dev. Sci.* 7, 270–276. doi: 10.1111/J.1467-7687.2004.00345.X

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Brain Res. Cogn. Brain Res.* 18, 149–161.

Weber, C., Hahne, A., Friedrich, M., and Friederici, A. D. (2005). Reduced stress pattern discrimination in 5-month-olds as a marker of risk for later language impairment: Neurophysiologial evidence. *Cogn. Brain Res.* 25, 180–187.

Weyers, I., and Mueller, J. L. (2022). A special role of syllables, but not vowels or consonants, for nonadjacent dependency learning. *J. Cogn. Neurosci.* 34, 1467–1487. doi: 10.1162/JOCN_A_01874

Weyers, I., Männel, C., and Mueller, J. L. (2022). Constraints on infants' ability to extract non-adjacent dependencies from vowels and consonants. *Dev. Cogn. Neurosci.* 57:101149. doi: 10.1016/J.DCN.2022.101149

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., et al. (2018). Non-adjacent dependency learning in humans and other animals. *Top. Cogn. Sci.* 12, 843–858. doi: 10.1111/tops.12381

Winkler, M., Mueller, J. L., Friederici, A. D., and Männel, C. (2018). Infant cognition includes the potentially human-unique ability to encode embedding. *Sci. Adv.* 4:eaar8334. doi: 10.1126/sciadv.aar8334

World Medical Association (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310, 2191–2194. doi: 10.1001/jama.2013.281053

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/fnhum.2024.1379660

Exposure to bilingual or monolingual maternal speech during pregnancy affects the neurophysiological encoding of speech sounds in neonates differently

Natàlia Gorina-Careta[1,2,3†], Sonia Arenillas-Alcón[1,2,3†], Marta Puertollano[1,2,3], Alejandro Mondéjar-Segovia[1,2], Siham Ijjou-Kadiri[1,2], Jordi Costa-Faidella[1,2,3*], María Dolores Gómez-Roig[3,4] and Carles Escera[1,2,3*]

[1]Brainlab – Cognitive Neuroscience Research Group, Departament de Psicologia Clinica i Psicobiologia, Universitat de Barcelona, Barcelona, Spain, [2]Institut de Neurociències, Universitat de Barcelona, Barcelona, Spain, [3]Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Barcelona, Spain, [4]BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona, Barcelona, Spain

Introduction: Exposure to maternal speech during the prenatal period shapes speech perception and linguistic preferences, allowing neonates to recognize stories heard frequently *in utero* and demonstrating an enhanced preference for their mother's voice and native language. Yet, with a high prevalence of bilingualism worldwide, it remains an open question whether monolingual or bilingual maternal speech during pregnancy influence differently the fetus' neural mechanisms underlying speech sound encoding.

Methods: In the present study, the frequency-following response (FFR), an auditory evoked potential that reflects the complex spectrotemporal dynamics of speech sounds, was recorded to a two-vowel /oa/ stimulus in a sample of 129 healthy term neonates within 1 to 3 days after birth. Newborns were divided into two groups according to maternal language usage during the last trimester of gestation (monolingual; bilingual). Spectral amplitudes and spectral signal-to-noise ratios (SNR) at the stimulus fundamental (F_0) and first formant (F_1) frequencies of each vowel were, respectively, taken as measures of pitch and formant structure neural encoding.

Results: Our results reveal that while spectral amplitudes at F0 did not differ between groups, neonates from bilingual mothers exhibited a lower spectral SNR. Additionally, monolingually exposed neonates exhibited a higher spectral amplitude and SNR at F_1 frequencies.

Discussion: We interpret our results under the consideration that bilingual maternal speech, as compared to monolingual, is characterized by a greater complexity in the speech sound signal, rendering newborns from bilingual mothers more sensitive to a wider range of speech frequencies without generating a particularly strong response at any of them. Our results contribute to an expanding body of research indicating the influence of prenatal experiences on language acquisition and underscore the necessity of including prenatal language exposure in developmental studies on language acquisition, a variable often overlooked yet capable of influencing research outcomes.

Check for updates

OPEN ACCESS

EDITED BY
Judit Gervain,
Centre National de la Recherche Scientifique (CNRS), France

REVIEWED BY
Irene De La Cruz Pavía,
Université Paris Cité, France
Gaia Lucarini,
University of Padua, Italy

*CORRESPONDENCE
Carles Escera
✉ cescera@ub.edu
Jordi Costa-Faidella
✉ jcostafaidella@ub.edu

[†]These authors have contributed equally to this work

RECEIVED 31 January 2024
ACCEPTED 22 April 2024
PUBLISHED 22 May 2024

CITATION
Gorina-Careta N, Arenillas-Alcón S, Puertollano M, Mondéjar-Segovia A, Ijjou-Kadiri S, Costa-Faidella J, Gómez-Roig MD and Escera C (2024) Exposure to bilingual or monolingual maternal speech during pregnancy affects the neurophysiological encoding of speech sounds in neonates differently. *Front. Hum. Neurosci.* 18:1379660. doi: 10.3389/fnhum.2024.1379660

KEYWORDS

speech brainstem responses, bilingualism, newborns, early language acquisition, frequency-following response (FFR), prenatal exposure

Introduction

The process of language acquisition has long been a point of uncertainty in research exploring the roots of human language. Researchers have conducted extensive investigations to understand the initial state and process of language acquisition, providing insights into how environmental and genetic factors interact to fashion language and cognitive function, and the mechanisms underlying brain plasticity (Weaver et al., 2004; Werker and Tees, 2005; Barkat et al., 2011; Werker and Hensch, 2015). It is now widely accepted that both genetic and experiential factors contribute to language acquisition (Werker and Curtin, 2005; Gervain and Mehler, 2010), and researchers are interested in understanding how these factors interact during human development.

Infants at birth already exhibit advanced speech perception and language learning abilities. Newborns manifest a preference for speech over non-speech sounds (Vouloumanos and Werker, 2007), can discriminate between different languages based on their speech rhythms (Ramus et al., 2000), detect word boundaries (Christophe et al., 2001), discriminate words with different patterns of stress (Sansavini et al., 1997), or even distinguish consonant sounds (Cabrera and Gervain, 2020) and encode voice pitch in an adult-like manner (Arenillas-Alcón et al., 2021). These findings support the role of a genetically driven cerebral organization towards processing specific speech characteristics.

However, the prenatal period is not devoid of language experience and the study of its influence on the newborn's speech and language encoding capacities is receiving increasing attention. Hearing becomes functional and undergoes most of its development around the 26th to 28th week of gestation, allowing the fetus to perceive the maternal speech signal (Ruben, 1995; Moore and Linthicum, 2007; Granier-Deferre et al., 2011; May et al., 2011; Anbuhl et al., 2016). Although the exact characteristics of the acoustic signal reaching the fetus are not fully understood, intrauterine recordings from animal models and simulations suggest that the maternal womb acts as a low-pass filter, attenuating around 30 dB for frequencies over 600–1,000 Hz (Gerhardt and Abrams, 2000). The low-frequency components of speech that are transmitted through the uterus include pitch, slow aspects of rhythm and some phonetic information (Moon and Fifer, 2000; May et al., 2011). Evidence indicates that prenatal exposure to speech, despite attenuated by the filtering properties of the womb, shapes speech perception and linguistic preferences of newborns, as shown by studies revealing that neonates can recognize a story heard frequently *in utero* (DeCasper and Spence, 1986), prefer the voice of their mother (DeCasper and Fifer, 1980) and prefer their native language (Moon et al., 1993). Additionally, prenatal learning extends beyond these common preferences. Recent findings indicate that infants acquire specific knowledge of the prosody (Gervain, 2018) and prefer the rhythmic patterns of the language they were exposed to while *in utero* (Mariani et al., 2023), indicating a very early specialization for their native language.

Yet, with reported rates of bilingualism of around 65% in Europe (Luk, 2017), an open question remains on the influence of prenatal exposure to more than one language on neural plasticity. Over the past 20 years, mounting evidence has suggested that both exposure to a bilingual acoustic environment and learning several languages affects not only language acquisition but a wide range of developmental processes including perception, cognition and brain development (Byers-Heinlein et al., 2019). Prior research has highlighted that early exposure to language influences infants' acquisition of speech sounds, indicating that, at birth, infants are able to discriminate all phonetic contrasts. As infants age, their perceptual systems are tuned to collapse over phonetic contrasts not found in the input language or languages, such that their ability to distinguish between phonetic elements becomes increasingly specific to their native language(s) (Kuhl et al., 2006; Saffran et al., 2006; Gervain and Werker, 2008; Kovács and Mehler, 2009; Bosch and Sebastián-Gallés, 2010). Moreover, cross-language interactions modulate almost every level of language processing, including speech perception, phonological, vocabulary and semantic development [for comprehensive review, refer to Hammer et al. (2014) and Kroll et al. (2012)]. Furthermore, some bilinguals switch from one language to the other within the same sentence, demonstrating greater demands on cognitive control than monolinguals to navigate the potential cross-language competition considering that language production is equivalent (Kovács and Mehler, 2009).

Speaking two languages daily also has consequences for the way in which higher cognitive processes operate and results in more precocious development of inhibition and attentional abilities (Costa et al., 2008; Kovács and Mehler, 2009; for review see Barac et al., 2014; Bialystok, 2017). There is evidence for functional and structural brain changes associated with bilingualism, even after brief periods of second-language learning (for extensive review see Li et al., 2014). Bilingual infants show different brain responses to native and non-native speech sounds than monolingual infants (Conboy and Kuhl, 2011). Bilingualism also affects the structure of both grey (Ressel et al., 2012) and white matter (Kuhl et al., 2016) in adults. The observed advantages in cognitive control and attentional abilities, as well as the pattern of structural differences, are modulated by the age of second language acquisition, whether the two languages were acquired simultaneously from birth or sequentially later in life and the interaction between languages (Kroll et al., 2012; Barac et al., 2014; Li et al., 2014).

As bilingual mothers speak using two different sets of phonemic categories and even use two slightly different voice pitch ranges (e.g., Ordin and Mennen, 2017), *in-utero* bilingual environments are characterized by a greater complexity of the reaching speech signal than monolingual ones. Interestingly, neonates exposed prenatally to a bilingual environment can discriminate their two native languages already at birth and exhibit equal preferences for both (Byers-Heinlein et al., 2010). Thus, it appears clear that linguistic experiences while *in utero* play a significant role in shaping the early development of speech processing. However, how different prenatal maternal linguistic exposure influences the neural mechanisms underlying speech sound processing at birth is currently unknown.

A large body of evidence has supported the study of the neural encoding of speech sounds through electrophysiological recordings. In particular, the frequency-following response (FFR) can provide insights into the underlying neural mechanisms associated with prenatal language experience, shedding light on how early linguistic exposure shapes the speech-encoding capacities of newborns. The FFR is an auditory evoked potential elicited by periodic complex sounds that reflects neural synchronization with the auditory eliciting signal along the ascending auditory pathway (Skoe and Kraus, 2010; Krizman and Kraus, 2019), providing an accurate snapshot of the neural encoding of speech sounds. FFR recordings have thus become a useful tool to investigate the ability to distinguish between the pitch of different speakers' voices and the ability to encode the fine spectrotemporal details that distinguish different voiced speech sounds (Gorina-Careta et al., 2022). The interest in the neonatal FFR arises from its potential to serve as a predictive measure for future language development (Schochat et al., 2017), since alterations in FFR patterns in children have been associated with difficulties in reading and learning, dyslexia, impairments in phonological awareness and even autism (King et al., 2002; Banai et al., 2009; Chandrasekaran et al., 2009; Basu et al., 2010; Hornickel et al., 2012; Lam et al., 2017; Otto-Meyer et al., 2018; Font-Alaminos et al., 2020; Rosenthal, 2020). Interestingly, the FFR reflects the impact of a wide range of auditory experiences in children and adults, including training interventions, musical practice and bilingualism (Russo et al., 2005; Song et al., 2008; Kraus and Chandrasekaran, 2010; Krizman et al., 2012, 2015; Carcagno and Plack, 2017; Skoe et al., 2017; Gorina-Careta et al., 2019). In adults it has been observed that bilingual experience enhances the neural responses to the fundamental frequency of sounds (Krizman et al., 2015; Skoe et al., 2017), as well as the subcortical representation of pitch-relevant information (Krizman et al., 2012) and neural consistency, which correlated with both a better attentional control and language proficiency (Krizman et al., 2014). In neonates, FFR recordings have also been used to study the effects at birth of prenatal fetal auditory experiences such as music exposure (Arenillas-Alcón et al., 2023), but the influence of prenatal maternal bilingual speech remains unexplored.

In the present study, we aimed to examine the influence of maternal bilingual linguistic exposure *in-utero* in speech sound encoding at birth. To that end, we recorded FFRs from newborns who had been exposed to either a monolingual or a bilingual fetal environment during the last trimester of gestation and analyzed their capacity to encode voice pitch and vocalic formant structure information.

Methods

Participants

A sample of 131 newborns (mean age after birth = 38.32 ± 23.8 h) was recruited from *SJD Barcelona Children's Hospital* in Barcelona (Spain) and divided into two groups based on a short retrospective questionnaire delivered to the babies' mothers. Mothers were asked if they communicated using more than one language during the last 3 months of pregnancy and were instructed to report which languages they communicated in, provided they accounted for a minimum of 20% language usage time. Based on the collected responses, a total of 53 newborns were assigned to the group exposed to a monolingual fetal acoustic environment (MON; 27 females; mean gestational age = 39.93 ± 1.03 weeks; mean birth weight = 3,321 ± 272 g). A total of 76 newborns were assigned to the bilingual-exposed group (BIL; 33 females; mean gestational age = 39.71 ± 0.99 weeks; mean birth weight = 3,328 ± 327 g) after excluding two newborns, as their mothers were multilingual in Spanish, Catalan and English, being the third language used ≧20% of the time. Regarding the languages spoken by the bilingual mothers, all except one were Spanish—Other language and most of them were Spanish-Catalan bilinguals (77.3%). The other languages spoken were Arabic (6/75), English (1/75), Galician (1/75), German (1/75), Italian (2/75), Portuguese (2/75), Guaraní (2/75) and Romanian (2/75). On the other hand, newborns in the monolingual group were either exposed to Spanish (90.6%) or Catalan (9.4%).

No significant differences were found across groups in gestational age ($U_{(127)} = 1868.500$, $p = 0.370$), birth weight ($t_{(127)} = -0.116$, $p = 0.908$) and sex ($\chi^2 = 0.710$, $p = 0.399$). Maternal education level and musical exposure were assessed using a sociodemographic questionnaire (an English version of the sociodemographic questionnaire can be found in the Supplementary material). Groups did not differ in maternal educational level ($\chi^2 = 1.992$, $p = 0.574$), a key confounding factor associated with language acquisition and development (Hoff, 2003; Rowe, 2008) closely tied to the linguistic environment a fetus is exposed to. We also ascertained that groups did not differ in prenatal musical exposure [$\chi^2 = 0.025$, $p = 0.874$; see Arenillas-Alcón et al. (2023) for details], as it exerts a significant impact on speech encoding capacities at birth (Partanen et al., 2013b, 2022; Arenillas-Alcón et al., 2023).

All neonates obtained Apgar scores higher than 8 at 1 and 5 min of life and passed adequately the universal newborn hearing screening (UNHS) before the recruitment. According to the recommendations of the Joint Committee on Infant Hearing (2019), newborns born from high-risk gestations, after obstetric pathologies or any other kind of risk factors related to hearing impairment were excluded from the recruitment.

Additionally, as performed in previous research from our laboratory (Ribas-Prats et al., 2019, 2021, 2023; Arenillas-Alcón et al., 2021, 2023), both groups of newborns received a standard click-evoked auditory brainstem response (ABR) test to ensure the integrity of the auditory pathway. A click-stimulus, with a duration of 100 μs, was employed during the test, presented at a rate of 19.30 Hz with an intensity of 60 dB sound pressure level (SPL) until a total of 4000 artifact-free repetitions were collected. A prerequisite for participation in the experiment for all newborns was the successful identification of the wave V peak. This study was approved by the Ethical Committee of Clinical Research of the Sant Joan de Déu Foundation (Approval ID: PIC-53-17), and required the mothers to fill out a sociodemographic questionnaire and to sign an informed consent prior to the participation, in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimulus

Neonatal FFRs were collected to a two-vowel stimulus with a rising pitch ending (/oa/; Arenillas-Alcón et al., 2021). The /oa/ stimulus was created in *Praat* (Boersma and Weenink, 2020) and had a total length of 250 ms divided into three different sections, according its fundamental frequency (F_0) and its formant content (/o/ vowel section: 0–80 ms, F_0 = 113 Hz, F_1 = 452 Hz, F_2 = 791 Hz; /oa/ formant transition section = 80–90 ms; /a/ vowel steady section = 90–160 ms, F_0 = 113 Hz, F_1 = 678 Hz, F_2 = 1,017 Hz; /a/ vowel rising section = 160–250 ms, F_0 = 113–154 Hz, F_1 = 678 Hz, F_2 = 1,017 Hz; Figure 1A).

The stimulus was designed with optimal parameters to study the frequency-following response, specially taking into account that due to the low-pass filter characteristics of the womb, fetuses are isolated from the mid and high frequency acoustic content of external sounds that characterizes most of the temporal fine structure of speech. The /oa/ stimulus used includes a pitch variation and two vowel sections with different formant structure based on relatively lower frequency harmonic components and suitable durations for accurate spectral analyses, which enable a proper assessment of speech sound temporal envelope and temporal fine structure encoding (Krizman and Kraus, 2019; Arenillas-Alcón et al., 2021). The relatively low F_0 frequency, typical of a male speaker, was chosen to ensure a reliable measure of the neural representation of sound pitch (Krizman and Kraus, 2019) and the phonetic contrasts (/o/; /a/) belong to the phonetic repertoire of both Spanish and Catalan languages.

The /oa/ stimulus was presented at a rate of 3.39 Hz in alternating polarities and delivered monaurally to the right ear at 60 dB SPL of intensity with an earphone connected to a Flexicoupler disposable adaptor (Natus Medical Incorporated, San Carlos, CA).

Procedure and data acquisition

After the successful completion of the UNHS, neonates were tested at the hospital room while they were sleeping in their bassinet. Three disposable Ag/AgCl electrodes were placed in a vertical montage configuration (active at Fpz, ground at forehead, reference at the right mastoid, ipsilateral to the auditory stimulation; as shown in Figure 1B), ensuring impedances below 7 kΩ. The presentation of click and speech stimuli was done by using a *SmartEP* platform connected to a *Duet* amplifier, which incorporated the *cABR* and the *Advanced Hearing Research* modules (Intelligent Hearing Systems, Miami, FL, United States).

The experimental procedure involved the recording of two blocks of click stimuli, followed by four blocks of 1000 artifact-free responses to the /oa/ stimulus. Any electrical activity surpassing ±30 µV threshold was automatically rejected until a total of 4,000 presentations was collected. The total mean duration of the recording session was approximately 25 min [2 click blocks × 2,000 repetitions × 51.81 ms SOA + 4 /oa/ blocks × 1,000 repetitions × 295 ms of stimulus-onset asynchrony (SOA)] including the duration of rejected sweeps. The continuous EEG signal was acquired at a sampling rate of 13,333 Hz with an online bandpass filter with cutoff frequencies from 30 to 1,500 Hz and online epoched from −40.95 ms (pre-stimulus period) to 249.975 ms.

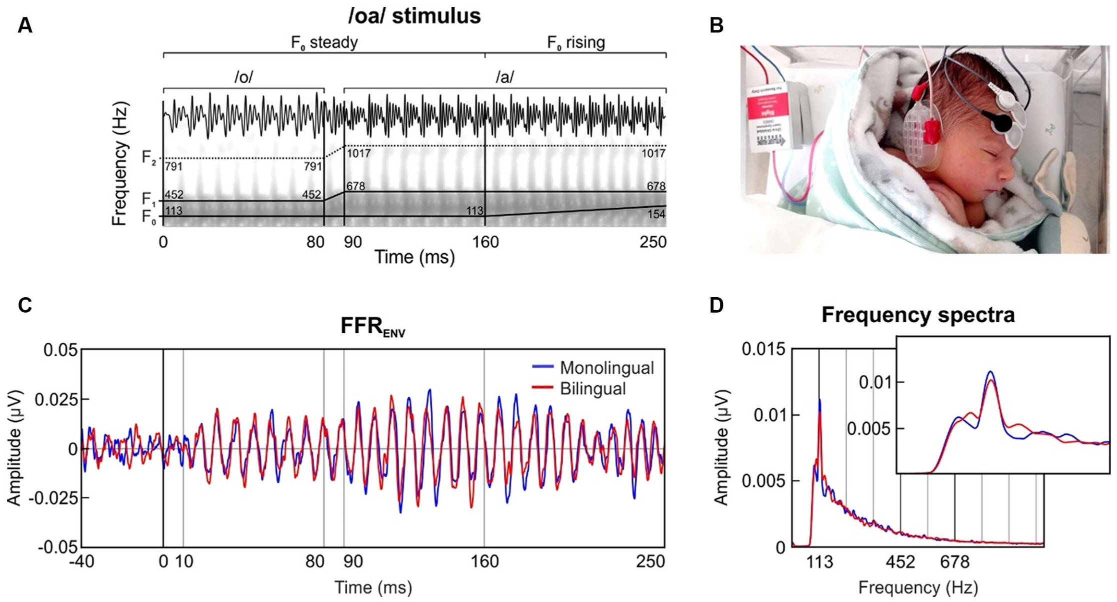


FIGURE 1
(A) Temporal and spectral representation of the two-vowel auditory stimulus /oa/, with traces indicating its fundamental frequency (F_0) and formant structure (F_1, F_2). **(B)** Recording setup of the three disposable electrodes placed in a vertical montage (active located at Fpz, ground at forehead, references at the right mastoid). Baby's photograph reproduced with the written consent of the neonate's parents. **(C)** Grand-averaged waveform of the FFR_{ENV} in the time domain, retrieved separately for the group exposed to monolingual (blue) and bilingual (red) fetal acoustic environment. **(D)** Frequency spectra of the FFR_{ENV} extracted from the steady pitch section of the stimulus (10–160 ms). The inset zooms in a narrower frequency band to illustrate the effect around the F_0 peak.

Data processing and analysis

Data epochs were bandpass filtered offline from 80 to 1,500 Hz and averaged separately per stimulus polarity. To highlight the encoding of the stimulus fundamental frequency (F_0) and to reduce the contribution of cochlear microphonics, neural responses to the two opposite stimulus polarities were added [(Condensation + Rarefaction)/2], obtaining the envelope-following response (FFR_{ENV}). Further, to emphasize the FFR components associated with the encoding of the stimulus temporal fine structure, such as the first formant (F_1), while reducing the impact of envelope-related activity, the neural responses to alternating polarities were subtracted [(Condensation − Rarefaction)/2], yielding the temporal fine structure-following response (FFR_{TFS}; Aiken and Picton, 2008; Krizman and Kraus, 2019). Considering the stimulus formant content, we focused our analyses exclusively on the spectral peaks that corresponded to F_1 frequencies, as F_2 frequencies fall at the limits of the spectral resolution of the FFR, resulting in elicited neural responses relatively weak and challenging to be accurately observed in newborns (Gorina-Careta et al., 2022). Detailed information regarding the analyzed parameters from the neonatal FFR can be found below. All parameters were computed using custom scripts in Matlab R2019b (The Mathworks Inc., 2019), developed in our laboratory and previously employed in similar analyses in former studies (Arenillas-Alcón et al., 2021).

Neural lag

Neural lag served as an indicator of the neural transmission delay within the auditory system, and was assessed to estimate the time passed from cochlear stimulus reception to the onset of neural phase-locking (Jeng et al., 2010; Liu et al., 2015; Ribas-Prats et al., 2019, 2021, 2023; Arenillas-Alcón et al., 2021, 2023). To calculate the neural lag, a cross-correlation analysis was computed between the auditory stimulus and the neural response. The neural lag was determined by identifying the time lag corresponding to the highest cross-correlation value within a time window of 3–13 ms.

Pre-stimulus root mean square (RMS) amplitude

The RMS of the pre-stimulus period was employed as a measure of the general magnitude of neural activity over time, and to dismiss electrophysiological disparities in the pre-stimulus region (Liu et al., 2015; White-Schwoch et al., 2015; Ribas-Prats et al., 2019, 2021, 2023; Arenillas-Alcón et al., 2023). This measure was computed by squaring each data point within the pre-stimulus region of the neural response (from −40 to 0 ms), calculating the mean of the squared values and subsequently obtaining the square root of the resulting average.

Voice pitch encoding from FFR_{ENV}

Spectral amplitude at F_0

Spectral amplitude at F_0 (113 Hz) was used as a quantitative measure of the neural phase-locking strength at the specific frequency of interest (White-Schwoch et al., 2015; Ribas-Prats et al., 2019, 2021, 2023; Arenillas-Alcón et al., 2021, 2023). It was computed by applying a fast Fourier transform (FFT; Cooley and Tukey, 1965) to obtain the frequency spectrum of the neural response during the steady pitch section of the stimulus (10–160 ms), and then calculating the average amplitude within a ± 5 Hz window centered around the peak of the stimulus F_0.

Signal-to-noise ratio at F_0

Signal-to-noise ratio (SNR) at F_0 was analyzed to obtain an estimation of the relative spectral magnitude of the response, taking into account not only to the amplitude value at the F_0 frequency peak (113 Hz) but also the noise levels at the surrounding frequencies. Therefore, the SNR was calculated by dividing the mean spectral amplitude within a ± 5 Hz frequency window centered at the peak of the frequency of interest (113 Hz) by the averaged mean amplitude within two additional 28 Hz wide frequency windows (flanks), centered at ±19 Hz from the frequency of interest (80–108 Hz and 118–146 Hz).

Formant structure encoding from FFR_{TFS}

Spectral amplitudes at F_1 peaks

To assess spectral amplitudes at the specific spectral peaks regarding the stimulus F_1 frequencies (452 Hz [/o/] and 678 Hz [/a/]), the neural responses corresponding to the /o/ section (10–80 ms time window) and the /a/ steady section (90–160 ms time window) were individually analyzed and the respective amplitudes within a ± 5 Hz window centered at the peak frequencies corresponding to the vowel formant centers were extracted. The transition from /o/ vowel to /a/ vowel was not analyzed due to its short duration (10 ms).

Signal-to-noise ratio at F_1

To compute the relative spectral magnitude of the response at the stimulus F_1 frequencies considering noise levels, SNRs at spectral peaks that correspond to the stimulus F_1 frequencies (452 Hz and 678 Hz) were calculated separately on the /o/ and the /a/−steady sections. To do so, the SNR was calculated by dividing the mean spectral amplitude within a ± 5 Hz frequency window centered at the peak of the frequency of interest (452 or 678 Hz) by the averaged mean amplitude within two additional 28 Hz wide frequency windows (flanks), centered at ±26 Hz from the frequency of interest (for 452 Hz peak: 402–430 Hz and 474–502 Hz; for 678 Hz peak: 628–656 Hz and 700–728 Hz).

Statistical analysis

Statistical analyses were conducted using Jamovi 2.3.26 (The Jamovi Project, 2023). Descriptive statistics were calculated, including the mean, standard deviation (SD), median, first quartile (Q_1), third quartile (Q_3), interquartile range (IQR), and minimum and maximum values, for each computed parameter within the two groups of newborns (MON; BIL).

To analyze the effects of prenatal bilingual exposure on neural transmission delay, pre-stimulus root mean square amplitude and voice pitch encoding depending on the normality of the data, two-tailed independent samples *t*-tests or Mann–Whitney U tests were conducted to evaluate significant differences between groups, with Cohen's *d* being reported as the effect size. Kolmogorov–Smirnov test was used to assess the normal distribution of the samples.

The effects of prenatal bilingual exposure on formant structure encoding were analyzed with two repeated–measures ANOVAs with the factor Stimulus Section (/o/ section; /a/ section) as within-subjects factor and the factor Group (Monolingual; Bilingual) as between-subjects factor for each of the two formant amplitudes (452 and 678 Hz) separately. The Greenhouse–Geisser correction was applied when the assumption of sphericity was violated. Additional two-tailed independent samples Mann–Whitney U post-hoc tests were performed to examine the direction of the effects. Results were considered statistically significant when $p < 0.05$.

Results

Frequency following responses (FFR) elicited by a two-vowel speech stimulus /oa/ (Figure 1A) were collected from a total sample of 129 newborns divided into two groups according to their prenatal fetal exposure to monolingual (MON) or bilingual (BIL) maternal speech. To comprehensively evaluate the neonates' ability to encode the pitch and vowel formant structure of speech sounds, the neural responses to the fundamental frequency (F_0) and the vowels' first formant (F_1) were analyzed considering the distinct sound characteristics of the different stimulus sections. All detailed descriptive statistics from the parameters analyzed can be found in Supplementary Table S1.

Neural transmission delay

No significant differences were found across groups in neural lag ($U_{(127)} = 1950.500$, $p = 0.763$, Rank-biserial correlation = 0.032).

Pre-stimulus root mean square (RMS) amplitude

There were no statistically significant differences observed between the groups with regards to the background neural activity preceding the auditory stimulation ($U_{(127)} = 1914.000$, $p = 0.634$, Rank-biserial correlation = 0.050).

Voice pitch encoding (FFR_{ENV})

The grand-averaged FFR_{ENV} waveform for each group is illustrated in Figure 1C. To assess the robustness of the voice pitch representation, we analyzed the steady section (10–160 ms) of the /oa/ stimulus with a steady fundamental frequency (F_0) of 113 Hz.

The grand-averaged spectral representation of the neonatal FFR extracted from each group is depicted in Figure 1D. No differences were found across groups in spectral amplitude at F_0 computed using the steady pitch section of the stimulus ($U_{(127)} = 1736.000$, $p = 0.184$, Rank-biserial correlation = 0.138).

Yet, the statistical analyses performed on the F_0 SNR, which represents the F_0 relative spectral amplitude in relation with the spectral amplitude of the neighboring frequencies, revealed significant differences between groups, indicating that newborns exposed to a monolingual prenatal fetal environment exhibited significantly larger SNR values as compared to the bilingual exposed neonates ($U_{(127)} = 1508.000$, $p = 0.016$, Rank-biserial correlation = 0.251).

Formant structure encoding (FFR_{TFS})

The grand-averaged FFR_{TFS} waveform for each group is shown in Figure 2A. To evaluate the newborns' ability to encode the formant structure of speech sounds, the /oa/ stimulus included two sections with the same voice pitch but different fine-structure. Specifically, the /o/ section (10–80 ms) was characterized by a center formant frequency (F_1) of 452 Hz, and the /a/ steady section (90–160 ms) by a F_1 frequency of 678 Hz. Spectral amplitudes were retrieved from the FFR_{TFS} separately from neural responses during the /o/ section and the /a/ steady-pitch section, selecting the spectral peaks corresponding to stimulus F_1 frequencies.

The grand-averages of the FFR_{TFS} spectral amplitudes during the /o/ section are illustrated in Figure 2B for each group separately, while the spectral representations during the /a/ steady section are depicted in Figure 2C. F_1 spectral amplitudes during the /o/ section and the /a/ steady section are depicted in Figure 3 for each group at each formant center frequency (452 Hz, 678 Hz) separately.

When analyzing the effects of a prenatal maternal bilingual language exposure in formant spectral amplitude at 452 Hz (Figure 3, left panel), which corresponds to the F_1 center frequency of the /o/ vowel, a main effect of group revealed significantly greater spectral amplitudes in the MON group as compared to the BIL (group main effect; $F_{(1,127)} = 4.939$, $p = 0.028$, ηp2 = 0.037). Moreover, a significantly larger spectral amplitude was observed during the /o/ section vs. /a/ steady section (stimulus section main effect; $F_{(1,127)} = 7.580$, $p = 0.007$, ηp2 = 0.056), thus indicating a proper encoding of the vowel /o/ in its corresponding stimulus section. Interestingly, a significant interaction of group per stimulus section was identified as well (interaction; $F_{(1,127)} = 5.809$, $p = 0.017$, ηp2 = 0.044), demonstrating that MON neonates showed significantly larger spectral amplitudes during the /o/ section at its corresponding formant frequency than BIL.

Similar results were observed when analyzing the effects of a prenatal maternal bilingual language exposure in the formant encoding at 678 Hz (Figure 3, right panel), which corresponds to the F_1 center frequency of the /a/ vowel. A main effect of group revealed significantly greater spectral amplitudes in the MON group as compared to the BIL (group main effect; $F_{(1,127)} = 5.01$, $p = 0.027$, ηp2 = 0.038). Moreover, a significantly larger spectral amplitude at 678 Hz during the /a/ steady section vs. /o/ section was observed (stimulus section main effect; $F_{(1,127)} = 10.93$, $p = 0.001$, ηp2 = 0.079), thus indicating a proper encoding of the /a/ vowel in its corresponding stimulus section. Interestingly, a significant interaction of group per stimulus section was also identified (interaction; $F_{(1,127)} = 5.812$, $p = 0.017$, ηp2 = 0.044), demonstrating that the MON group exhibited higher spectral amplitudes during the /a/ steady section at its corresponding frequency than the BIL.

The same pattern of results was obtained when comparing the relative spectral amplitude of the response at the stimulus F_1 frequencies taking into account the neural response to the neighboring frequencies. When analyzing the effects of a fetal maternal bilingual language exposure in SNR at 452 Hz, which corresponds to the F_1 of the /o/ vowel, a main effect of group revealed significantly greater spectral amplitudes in the MON group as compared to the BIL (group

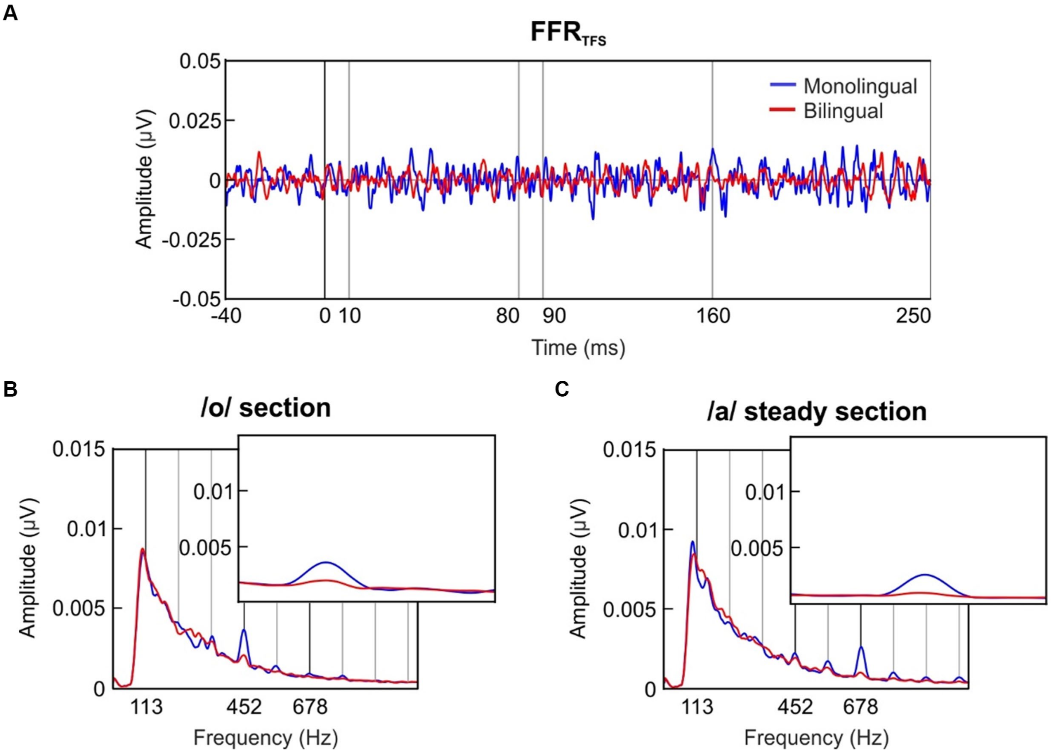


FIGURE 2
Formant structure encoding. **(A)** Grand-averaged waveform of the FFR$_{TFS}$ in the time domain, retrieved separately for the group exposed to a monolingual fetal acoustic environment (blue) and the bilingual-exposed group (red). **(B)** Frequency spectra of the FFR$_{TFS}$ extracted from the /o/ section of the stimulus (10–80 ms). The inset zooms in a narrower frequency band to illustrate the effect around the /o/ F_1 peak (452 Hz) during the /o/ section. **(C)** Frequency spectra of the FFR$_{TFS}$ extracted from the /a/ steady section of the stimulus (90–160 ms). The inset zooms in a narrower frequency band to illustrate the effect around the /a/ F_1 peak (678 Hz) during the /a/ steady section.

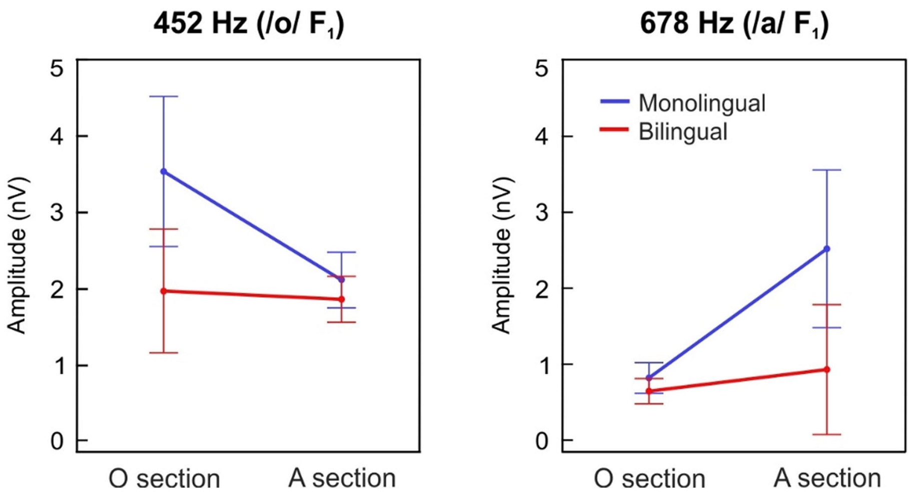


FIGURE 3
Spectral amplitudes at the first formant (F_1). F_1 spectral amplitudes at 452 Hz (left) and 678 Hz (right) during the /o/ section (10–80 ms) and the /a/ steady section (90–160 ms), plotted in blue and red lines for the monolingual and the bilingual-exposed newborns, respectively. Error bars represent 95% confidence intervals.

main effect; $F_{(1,127)} = 8.301$, $p = 0.005$, ηp2 = 0.061). Moreover, a significantly larger spectral amplitude was observed during the /o/ section vs. /a/ steady section (stimulus section main effect; $F_{(1,127)} = 7.517$, $p = 0.007$, ηp2 = 0.056). A significant interaction of group per stimulus section was identified as well (interaction; $F_{(1,127)} = 7.304$, $p = 0.008$, ηp2 = 0.054).

Similar effects were observed when analyzing the effects of a prenatal bilingual environment in the formant SNR at 678 Hz, which corresponds to the frequency of the /a/ vowel. A main effect of group revealed significantly greater spectral amplitudes in the MON group as compared to the BIL (group main effect; $F_{(1,127)} = 7.127$, $p = 0.009$, $\eta p2 = 0.053$). Moreover, a significantly larger spectral amplitude at 678 Hz during the /a/ steady section vs. /o/ section was observed (stimulus section main effect; $F_{(1,127)} = 22.072$, $p < 0.001$, $\eta p2 = 0.148$). Finally, a significant interaction of group per stimulus section was also identified (interaction; $F_{(1,127)} = 10.330$, $p = 0.002$, $\eta p2 = 0.075$).

Discussion

The present study investigated the impact of maternal bilingual speech during pregnancy on the neural encoding of speech pitch and vowel formant structure in neonates. A total sample of 129 healthy-term newborns was divided into two groups according to their monolingual or bilingual prenatal exposure during the last trimester of gestation, as reported by their mothers through a questionnaire. FFRs elicited to a two-vowel speech stimulus /oa/ (Arenillas-Alcón et al., 2021) were recorded to assess the neural responses to the stimulus' fundamental frequency ($F_0 = 113$ Hz; related to voice pitch encoding) and the first formant of each vowel (/o/ $F_1 = 452$ Hz; /a/ $F_1 = 678$ Hz; related to vowel formant structure encoding). Our results revealed that the neural representation of pitch, as indexed by the spectral amplitude of the FFR_{ENV} at the stimulus F_0, did not differ between monolingual and bilingual exposure groups, but monolingually exposed neonates exhibited a higher signal-to-noise ratio (SNR) at the F_0 spectral peak, suggesting the contribution of a higher spectral noise at neighboring frequencies in the bilingual group. Additionally, monolingually exposed neonates exhibited larger spectral amplitudes and SNRs of the FFR_{TFS} at the formant peak frequencies (F_1) of the speech stimulus used, indicating a stronger encoding of vocalic structure. Furthermore, no significant group differences were observed in neural lag and pre-stimulus root mean square (RMS) amplitude, implying comparable neural transmission delays and absence of a distinct overall neural activity prior to the auditory stimulation. Together, these findings provide novel insights into the effects of prenatal language exposure on the neural encoding of speech sounds at birth.

Pitch is a crucial attribute in the perception of periodic speech sounds, as it conveys prosodic information, facilitates speaker recognition and speech segmentation, accelerates phoneme acquisition in tonal languages, helps with language comprehension in noisy environments and even contributes to the perception of the emotional state in a conversation (Musacchia et al., 2007; Benavides-Varela et al., 2012; Partanen et al., 2013a; Plack et al., 2014; Gervain, 2018; Cabrera and Gervain, 2020; Arenillas-Alcón et al., 2021; Ribas-Prats et al., 2021). The fact that neural mechanisms underlying voice pitch encoding are already mature at birth (Jeng et al., 2011; Ribas-Prats et al., 2019; Cabrera and Gervain, 2020; Arenillas-Alcón et al., 2021) suggests that pitch may play a crucial role in the very first stages of language acquisition (Jeng et al., 2016). Going a step further, pitch could provide a neural synchrony channel onto which separate neural representations of other speech features would anchor as parts of an ensemble that would, ultimately, give rise to a coherent percept (Eggermont, 2001).

Previous studies demonstrated that pitch and pitch contour discrimination drastically improve with training (e.g., Carcagno and Plack, 2017). In this regard, growing up in a bilingual environment, which is characterized as more demanding, dynamic, phonologically rich and requiring heightened attention to all linguistic input, is related to a strengthened neural representation of pitch (Krizman et al., 2012, 2015). Different languages have distinct overall height pitch levels. For example, Catalan was observed to have a higher pitch compared to Spanish (Marquina Zazura, 2011); Polish was found to have a higher pitch compared to American English (Majewski et al., 1972); Mandarin, a higher pitch than English (Keating and Kuo, 2012); Japanese, a higher pitch than Dutch (Van Bezooijen, 1995); or Slavic languages, a higher pitch than Germanic ones (Andreeva et al., 2014). Further, speakers of two phonologically similar dialects exhibit differences in their height pitch levels (e.g., two different dialects of Mandarin; Deutsch et al., 2009).

Yet, pitch height is not the only element that contributes significantly to the distinctiveness of a particular language. The intonational patterns, which are the rising and falling patterns of pitch that convey meaning and contribute to the rhythm of speech, may differ between the different languages. When a speaker switches between languages they naturally adjust the specific contours, pitch ranges, and other prosodic features to conform to the norms of the target language, and many linguistic features such as intonation, may affect the mean fundamental frequency of speech (Järvinen et al., 2013). This adjustment helps maintaining communicative clarity and aligns with the phonetic characteristics of the language being spoken (Mary and Yegnanarayana, 2008; Passoni et al., 2022).

With continued exposure to these complex linguistic contexts, the auditory system gradually becomes finely tuned to process sound more efficiently (Krizman et al., 2012). Thus, individuals with years of exposure and interaction with bilingual environments develop enhanced flexibility and speech-encoding abilities. Most notably, previous studies have shown that bilingual individuals, particularly females, exhibit different pitch frequency ranges depending on the language they speak (Ordin and Mennen, 2017). As both pitch and the intonational patterns of the languages are different, and the prosodic elements of speech which include pitch contours, rhythm, and stress (Moon and Fifer, 2000) are acoustic features reliably transmitted through the womb (Gerhardt and Abrams, 2000; May et al., 2011), bilingual mothers provide their children with a higher pitch variability *in utero*.

Considering the reviewed literature, if the developing auditory system of a fetus, who underwent approximately 3 months of noninteractional exposure to degraded speech, responded to acoustic exposure as the mature one, we would expect newborns from bilingual mothers to exhibit a higher neural encoding of voice pitch. But our results showed otherwise. We found no differences across groups in FFR_{ENV} spectral amplitudes at F_0, which aligns with the idea that pitch processing mechanisms are already mature at birth. Yet, we observed a decreased SNR at the F_0 in newborns who were prenatally exposed to a bilingual environment. We attempt to reconcile our seemingly contradicting results by hypothesizing that the higher spectral amplitudes found in bilingually exposed neonates at F_0 neighboring frequencies reflect an increased sensitivity to a wider range of pitch frequencies without yet generating a particularly strong response at any of them.

This view aligns with research on perceptual phonetic development, especially when growing in bilingual environments. Previous studies demonstrated that experience with language shapes infants' abilities to process speech sounds and, with age, the newborn's ability to differentiate phonetic distinctions becomes more language-specific (Kuhl et al., 2006; Saffran et al., 2006; Gervain and Werker, 2008; Bosch and Sebastián-Gallés, 2010). At birth all infants possess the ability to perceive all sound distinctions used in languages as they are sensitive to the basic rhythmic differences between languages (Nazzi et al., 1998; Byers-Heinlein et al., 2010). Around 3–4 months of age infants are sensitive to rhythmic differences between languages that go beyond their belonging to the three basic rhythmic classes (Bosch and Sebastián-Gallés, 2010; Molnar et al., 2014) and by the age of 6 months monolingual infants' ability to perceive speech becomes tailored to their native language. Infants exposed to two languages are also able to discriminate the sound contrasts of both their languages, but this occurs only at the end of their first year (Bosch and Sebastián-Gallés, 2003; Sundara et al., 2008; for review see Hammer et al., 2014).

Yet, the early prenatal impact of language goes beyond language discrimination. As reviewed in the introduction, newborns prefer their mother's voice over other female voices (DeCasper and Fifer, 1980), their communicative cries reflect the prosody of the language they heard *in utero* (Mampe et al., 2009) and can recognize stories heard during pregnancy (DeCasper and Spence, 1986). Moreover, previous studies also demonstrated that differences in prenatal language exposure modulate perceptual grouping biases at birth (Abboub et al., 2016) and suggest that hearing pitch contrasts before birth may influence pitch-based grouping preferences and may lead to a stable bias at birth. Thus, despite the discrimination (or no discrimination) of languages at birth, prenatal language exposure modulates the processing of speech sounds. Our findings align with the suggested hypothesis that being bilingual confers a greater perceptual flexibility (Abboub et al., 2016), as we observed in bilingually exposed newborns an increased sensitivity to a wider range of pitch frequencies.

Our results also reveal a modulation of the neural encoding of vowel formants (F_1) depending on prenatal linguistic exposure. In particular, monolingual-exposed neonates exhibited higher spectral amplitudes at the corresponding formant frequencies of the stimulus' /o/ and steady–/a/ vowels. In a previous study, we found that while the neural encoding of pitch was adult-like at birth, formant encoding was still immature (Arenillas-Alcón et al., 2021). As vowel formant center frequencies are language specific and stable regardless of voice pitch variation, which also presents slight modulations in monolingual individuals during natural speaking, the auditory system of a monolingual-exposed fetus receives a more consistent phonetic repertoire than that of a bilingual-exposed. This would possibly lead to a more effective and accurate encoding of the specific language vowel sound characteristics at birth. Simply put, monolingual newborns seem to have an advantage in processing the specific sounds of their mother tongue, a finding previously attributed to postnatal linguistic exposure (Kuhl, 2010). Our findings thus highlight the greater variability of acoustic speech inputs to which the fetus of bilingual mothers would be exposed and therefore suggest the need for bilinguals to develop a different phonological representation for each of the languages (Sebastian-Gallés et al., 2006). Further investigation into the developmental trajectories of auditory processing in different populations of newborns, with different prenatal auditory experiences, and using language-specific phonetic contrasts (e.g., Catalan contrasts such as /e - ɛ/), which are especially difficult –when not impossible– to detect for Spanish-monolinguals (Pallier et al., 1997, 2001), may shed more light on this issue.

Despite being confident about our results due to the abovementioned reasons, we are fully aware of a number of limitations of our study: language exposure was assessed by a short (approx. 5 min answer time), retrospective questionnaire provided at the time of delivery, with a spoken description of the content of the questionnaire. This poses, at least, two factors not adequately controlled. First, the actual frequency in which mothers spoke any of the two languages, as we rely only on their reports referring to the last trimester of pregnancy. Furthermore, although a minimum period of usage time had to occur to be considered as valid, the questionnaire did not address the exact amount of language usage within a day. Future studies should address these limitations, for instance, by collecting large amounts of data from a maternal diary of language usage during the last trimester of pregnancy and include an additional language abilities test (such as LEAP-Q; Marian et al., 2007) to evaluate the putative link between F_0 encoding abilities in newborns and maternal language usage percentage.

Overall, our findings emphasize the potential importance of prenatal linguistic exposure in shaping the neural mechanisms underlying language acquisition and highlight the sensitivity of the FFR in capturing these subtle changes. The results add to a growing body of research that suggests a role for prenatal fetal experiences in modeling language acquisition (Moon et al., 2012; Partanen et al., 2013b; Gervain, 2015, 2018; Arenillas-Alcón et al., 2023). Furthermore, they also highlight the importance of considering prenatal language exposure in developmental studies about language acquisition, a factor that is not routinely measured and reported, and that may contribute to divergent findings.

Conclusion

The present study contributes significant insights into the impact of prenatal bilingual exposure on the neural encoding of speech sounds at birth, thereby increasing our knowledge of the early stages of language acquisition. The observed differences in the encoding of voice pitch and formant structure depending on prenatal linguistic exposure highlight the remarkable plasticity and learning potential of the human brain even before birth, emphasizing the complex interaction between genetic and environmental factors in shaping our cognitive abilities and linguistic development.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Ethical Committee of Clinical Research of the Sant Joan de Déu Foundation (Approval ID: PIC-53-17). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

NG-C: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. SA-A: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. MP: Investigation, Writing – review & editing, Methodology. AM-S: Writing – review & editing, Methodology, Investigation. SI-K: Writing – review & editing, Methodology, Investigation. JC-F: Writing – review & editing, Supervision, Methodology, Conceptualization. MG-R: Writing – review & editing, Resources, Funding acquisition. CE: Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Spanish Ministry of Science and Innovation PGC2018-094765-B-I00 project (MCIN/AEI/10.13039/5011000 11033/ FEDER "Una manera de hacer Europa"); the project PID2021-122255NB-100 supported by MCIN/AEI/10.13039/501100011033/ FEDER, UE; the María de Maeztu Center of Excellence CEX2021-001159-M (supported by MCIN/AEI/10.13039/501100011033); the 2021SGR-00356 Consolidated Research Group of the Catalan Government, and the ICREA Acadèmia Distinguished Professorship awarded to CE.

Acknowledgments

The authors would like to thank to all families who selflessly participated in this study and made it possible.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2024.1379660/full#supplementary-material

References

Abboub, N., Nazzi, T., and Gervain, J. (2016). Prosodic grouping at birth. *Brain Lang.* 162, 46–59. doi: 10.1016/j.bandl.2016.08.002

Aiken, S. J., and Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. doi: 10.1016/j.heares.2008.08.004

Anbuhl, K. L., Uhler, K. M., Werner, L. A., and Tollin, D. J. (2016). "Early development of the human auditory system" in *Fetal and neonatal physiology*. eds. R. A. Polin, S. H. Abman, D. Rowitch and W. E. Benitz. *5th* ed (Amsterdam: Elsevier), 1396–1410.

Andreeva, B., Demenko, G., Möbius, B., Zimmerer, F., Jügler, J., and Jastrzebska, M. (2014). Differences of pitch profiles in Germanic and Slavic languages. In *Proceedings of Interspeech*. ISCA.

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., and Escera, C. (2021). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Sci. Rep.* 11:6660. doi: 10.1038/s41598-021-85799-x

Arenillas-Alcón, S., Ribas-Prats, T., Puertollano, M., Mondéjar-Segovia, A., Gómez-Roig, M. D., Costa-Faidella, J., et al. (2023). Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: evidence from neonatal frequency-following responses. *Dev. Sci.* 26:e13362. doi: 10.1111/desc.13362

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (2009). Reading and subcortical auditory function. *Cereb. Cortex* 19, 2699–2707. doi: 10.1093/cercor/bhp024

Barac, R., Bialystok, E., Castro, D. C., and Sanchez, M. (2014). The cognitive development of young dual language learners: a critical review. *Early Child. Res. Q.* 29, 699–714. doi: 10.1016/j.ecresq.2014.02.003

Barkat, T., Polley, D., and Hensch, T. (2011). A critical period for auditory thalamocortical connectivity. *Nat. Neurosci.* 14, 1189–1194. doi: 10.1038/nn.2882

Basu, M., Krishnan, A., and Weber-Fox, C. (2010). Brainstem correlates of temporal auditory processing in children with specific language impairment. *Dev. Sci.* 13, 77–91. doi: 10.1111/j.1467-7687.2009.00849.x

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., and Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proc. Natl. Acad. Sci. USA* 109, 17908–17913. doi: 10.1073/pnas.1205413109

Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychol. Bull.* 143, 233–262. doi: 10.1037/bul0000099

Boersma, P., and Weenink, D. (2020). *Praat: Doing phonetics by computer* (version 6.1.09). Available at: http://www.praat.org/

Bosch, L., and Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech* 46, 217–243. doi: 10.1177/00238309030460020801

Bosch, L., and Sebastián-Gallés, N. (2010). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* 2, 29–49. doi: 10.1207/S15327078IN0201_3

Byers-Heinlein, K., Burns, T. C., and Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychol. Sci.* 21, 343–348. doi: 10.1177/0956797609360758

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., and Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra Psychol.* 5:37. doi: 10.1525/collabra.233

Cabrera, L., and Gervain, J. (2020). Speech perception at birth: the brain encodes fast and slow temporal information. *Sci. Adv.* 6:eaba 7830. doi: 10.1126/sciadv.aba7830

Carcagno, S., and Plack, C. J. (2017). "Short-term learning and memory: training and perceptual learning" in *The frequency-following response: A window into human communication, vol. 61*. eds. N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay and A. N. Popper (London: Springer Nature), 75–100.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., and Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64, 311–319. doi: 10.1016/j.neuron.2009.10.006

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

Conboy, B. T., and Kuhl, P. K. (2011). Impact of second-language experience in infancy: brain measures of first- and second-language speech perception. *Dev. Sci.* 14, 242–248. doi: 10.1111/j.1467-7687.2010.00973.x

Cooley, J. W., and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* 19, 297–301. doi: 10.1090/S0025-5718-1965-0178586-1

Costa, A., Hernández, M., and Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition* 106, 59–86. doi: 10.1016/j.cognition.2006.12.013

DeCasper, A. J., and Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voice. *Science* 208, 1174–1176. doi: 10.1126/science.7375928

DeCasper, A. J., and Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behav. Dev.* 9, 133–150. doi: 10.1016/0163-6383(86)90025-1

Deutsch, D., Le, J., Shen, J., and Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *J. Acoust. Soc. Am.* 125, EL208–EL213. doi: 10.1121/1.3113892

Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code. *Hear. Res.* 157, 1–42. doi: 10.1016/S0378-5955(01)00259-3

Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., et al. (2020). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biol. Psychol.* 149:107807. doi: 10.1016/j.biopsycho.2019.107807

Gerhardt, K. J., and Abrams, R. M. (2000). Fetal exposures to sound and vibroacoustic stimulation. *J. Perinatol.* 20, S21–S30. doi: 10.1038/sj.jp.7200446

Gervain, J. (2015). Plasticity in early language acquisition: the effects of prenatal and early childhood experience. *Curr. Opin. Neurobiol.* 35, 13–20. doi: 10.1016/j.conb.2015.05.004

Gervain, J. (2018). The role of prenatal experience in language development. *Curr. Opin. Behav. Sci.* 21, 62–67. doi: 10.1016/j.cobeha.2018.02.004

Gervain, J., and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu. Rev. Psychol.* 61, 191–218. doi: 10.1146/annurev.psych.093008.100408

Gervain, J., and Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Lang. Linguist. Compass* 2, 1149–1170. doi: 10.1111/j.1749-818X.2008.00089.x

Gorina-Careta, N., Ribas-Prats, T., Arenillas-Alcón, S., Puertollano, M., Gómez-Roig, M. D., and Escera, C. (2022). Neonatal frequency-following responses: a methodological framework for clinical applications. *Semin. Hear.* 43, 162–176. doi: 10.1055/s-0042-1756162

Gorina-Careta, N., Ribas-Prats, T., Costa-Faidella, J., and Escera, C. (2019). "Auditory frequency-following responses" in *Encyclopedia of computational neuroscience*. eds. D. Jaeger and R. Jung (New York, NY: Springer), 1–13.

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., and Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Dev. Sci.* 14, 336–352. doi: 10.1111/j.1467-7687.2010.00978.x

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., and Sandilos, L. E. (2014). The language and literacy development of young dual language learners: a critical review. *Early Child. Res. Q.* 29, 715–733. doi: 10.1016/j.ecresq.2014.05.008

Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., and Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *Neuro Report* 23, 6–9. doi: 10.1097/WNR.0b013e32834d2ffd

Järvinen, K., Laukkanen, A.-M., and Aaltonen, O. (2013). Speaking a foreign language and its effect on F0. *Logoped. Phoniatr. Vocol.* 38, 47–51. doi: 10.3109/14015439.2012.687764

Jeng, F. C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., et al. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear Hear.* 32, 699–707. doi: 10.1097/AUD.0b013e31821cc0df

Jeng, F. C., Lin, C.-D., and Wang, T.-C. (2016). Subcortical neural representation to mandarin pitch contours in American and Chinese newborns. *J. Acoust. Soc. Am.* 139, EL190–EL195. doi: 10.1121/1.4953998

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.-D., et al. (2010). Early maturation of frequency-following responses to voice pitch in infants with normal hearing. *Percept. Mot. Skills* 111, 765–784. doi: 10.2466/10.22.24.PMS.111.6.765-784

Joint Committee on Infant Hearing (2019). Year 2019 position statement: principles and guidelines for early hearing detection and intervention programs. *Pediatrics* 106, 798–817. doi: 10.1542/peds.106.4.798

Keating, P., and Kuo, G. (2012). Comparison of speaking fundamental frequency in English and mandarin. *J. Acoust. Soc. Am.* 132, 1050–1060. doi: 10.1121/1.4730893

King, C., Warrier, C. M., Hayes, E., and Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci. Lett.* 319, 111–115. doi: 10.1016/S0304-3940(01)02556-3

Kovács, Á. M., and Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science* 325, 611–612. doi: 10.1126/science.1173947

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Krizman, J., and Kraus, N. (2019). Analyzing the FFR: a tutorial for decoding the richness of auditory function. *Hear. Res.* 382, 107779–107174. doi: 10.1016/j.heares.2019.107779

Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *PNAS* 109, 7877–7881. doi: 10.1073/pnas.1201575109

Krizman, J., Skoe, E., Marian, V., and Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain Lang.* 128, 34–40. doi: 10.1016/j.bandl.2013.11.006

Krizman, J., Slater, J., Skoe, E., Marian, V., and Kraus, N. (2015). Neural processing of speech in children is influenced by bilingual experience. *Neurosci. Lett.* 0, 48–53. doi: 10.1016/j.neulet.2014.11.011.Neural

Kroll, J. F., Dussias, P. E., Bogulski, C. A., and Valdes Kroff, J. R. (2012). Chapter seven – Juggling two languages in one mind: what bilinguals tell us about language processing and its consequences for cognition. *Psychol. Learn. Motiv.* 56, 229–262. doi: 10.1016/B978-0-12-394393-4.00007-8

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron* 67, 713–727. doi: 10.1016/j.neuron.2010.08.038

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Stevenson, J., Corrigan, N. M., Van den Bosch, J. J. F., Deniz Can, D., and Richards, T. (2016). Neuroimaging of the bilingual brain: structural brain correlates of listening and speaking in a second language. *Brain Lang.* 162, 1–9. doi: 10.1016/j.bandl.2016.07.004

Lam, S. S.-Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., and Kraus, N. (2017). Neural stability: a reflection of automaticity in reading. *Neuropsychologia* 103, 162–167. doi: 10.1016/j.neuropsychologia.2017.07.023

Li, P., Legault, J., and Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex* 58, 301–324. doi: 10.1016/j.cortex.2014.05.001

Liu, F., Maggu, A. R., Lau, J. C. Y., and Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Front. Hum. Neurosci.* 8, 1–19. doi: 10.3389/fnhum.2014.01029

Luk, G. (2017). "Bilingualism" in *The Cambridge encyclopedia of child development*. eds. B. Hopkins, E. Geangu and S. Linkenauger. *2nd* ed (Cambridge: Cambridge University Press), 385–391.

Majewski, W., Hollien, H., and Zalewski, J. (1972). Speaking fundamental frequency of polish adult males. *Phonetica* 25, 119–125. doi: 10.1159/000259375

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Curr. Biol.* 19, 1994–1997. doi: 10.1016/j.cub.2009.09.064

Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)

Mariani, B., Nicoletti, G., Barzon, G., Ortiz-Barajas, M. C., Shukla, M., Guevara, R., et al. (2023). Prenatal experience with language shapes the brain. *Sci. Adv.* 9:eadj3524. doi: 10.1126/sciadv.adj3524

Marquina Zazura, M. (2011). *Estudio acústico de la variación inter e intralocutor en la frecuencia fundamental de hablantes bilingües de catalán y de castellano* [Universitat Autònoma de Barcelona]. Available at: https://ddd.uab.cat/record/77033

Mary, L., and Yegnanarayana, B. (2008). Extraction and representation of prosodic features for language and speaker recognition. *Speech Comm.* 50, 782–796. doi: 10.1016/j.specom.2008.04.010

May, L., Byers-Heinlein, K., Gervain, J., and Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Front. Psychol.* 2, 1–9. doi: 10.3389/fpsyg.2011.00222

Molnar, M., Lallier, M., and Carreiras, M. (2014). The amount of language exposure determines nonlinguistic tone grouping biases in infants from a bilingual environment. *Lang. Learn.* 64, 45–64. doi: 10.1111/lang.12069

Moon, C., Cooper, R. P., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500. doi: 10.1016/0163-6383(93)80007-U

Moon, C., and Fifer, W. (2000). Evidence of transnatal auditory learning. *J. Perinatol.* 20, S37–S44. doi: 10.1038/sj.jp.7200448

Moon, C., Lagercrantz, H., and Kuhl, P. K. (2012). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatr.* 102, 156–160. doi: 10.1111/apa.12098

Moore, J. K., and Linthicum, F. H. (2007). The human auditory system: a timeline of development. *Int. J. Audiol.* 46, 460–478. doi: 10.1080/14992020701383019

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. USA* 104, 15894–15898. doi: 10.1073/pnas.0701498104

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756–766. doi: 10.1037/0096-1523.24.3.756

Ordin, M., and Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *J. Speech Lang. Hear. Res.* 60, 1493–1506. doi: 10.1044/2016_JSLHR-S-16-0315

Otto-Meyer, S., Krizman, J., White-Schwoch, T., and Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Exp. Brain Res.* 236, 733–743. doi: 10.1007/s00221-017-5164-4

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in vowel acquisition. *Cognition* 64, B9–B17. doi: 10.1016/s0010-0277(97)00030-9

Pallier, C., Colomé, A., and Sebastian-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychol. Sci.* 12, 445–449. doi: 10.1111/1467-9280.00383

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., and Huotilainen, M. (2013a). Learning-induced neural plasticity of speech processing before birth. *PNAS* 110, 15145–15150. doi: 10.1073/pnas.1302159110

Partanen, E., Kujala, T., Tervaniemi, M., and Huotilainen, M. (2013b). Prenatal music exposure induces long-term neural effects. *PLoS One* 8:e78946. doi: 10.1371/journal.pone.0078946

Partanen, E., Mårtensson, G., Hugoson, P., Huotilainen, M., Fellman, V., and Ådén, U. (2022). Auditory processing of the brain is enhanced by parental singing for preterm infants. *Front. Neurosci.* 16:772008. doi: 10.3389/fnins.2022.772008

Passoni, E., De Leeuw, E., and Levon, E. (2022). Bilinguals produce pitch range differently in their two languages to convey social meaning. *Lang. Speech* 65, 1071–1095. doi: 10.1177/00238309221105210

Plack, C. J., Barker, D., and Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hear. Res.* 307, 53–64. doi: 10.1016/j.heares.2013.07.020

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

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., et al. (2012). An effect of bilingualism on the auditory cortex. *J. Neurosci.* 32, 16597–16601. doi: 10.1523/JNEUROSCI.1996-12.2012

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., et al. (2019). The frequency-following response (FFR) to speech stimuli: a normative dataset in healthy newborns. *Hear. Res.* 371, 28–39. doi: 10.1016/j.heares.2018.11.001

Ribas-Prats, T., Arenillas-Alcón, S., Pérez-Cruz, M., Costa-Faidella, J., Gómez-Roig, M. D., and Escera, C. (2023). Speech-encoding deficits in neonates born large-for-gestational age as revealed with the frequency-following response. *Ear Hear.* 44, 829–841. doi: 10.1097/AUD.0000000000001330

Ribas-Prats, T., Arenillas-Alcón, S., Lip-Sosa, D. L., Costa-Faidella, J., Mazarico, E., Gómez-Roig, M. D., et al. (2021). Deficient neural encoding of speech sounds in term neonates born after fetal growth restriction. *Dev. Sci.* 25:e13189. doi: 10.1111/desc.13189

Rosenthal, M. A. (2020). A systematic review of the voice-tagging hypothesis of speech-in-noise perception. *Neuropsychologia* 136:107256. doi: 10.1016/j.neuropsychologia.2019.107256

Rowe, M. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *J. Child Lang.* 35, 185–205. doi: 10.1017/S0305000907008343

Ruben, R. J. (1995). The ontogeny of human hearing. *Int. J. Pediatr. Otorhinolaryngol.* 32, S199–S204. doi: 10.1016/0165-5876(94)01159-U

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., and Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behav. Brain Res.* 156, 95–103. doi: 10.1016/j.bbr.2004.05.012

Saffran, J. R., Werker, J. F., and Werner, L. A. (2006). "The Infant's auditory world: hearing, speech, and the beginnings of language" in *Handbook of child psychology: cognition, perception, and language*. eds. D. Kuhn, R. S. Siegler, W. Damon and R. M. Lerner (Hoboken, NJ: John Wiley & Sons, Inc), 58–108.

Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33, 3–11. doi: 10.1037/0012-1649.33.1.3

Schochat, E., Rocha-Muniz, C. N., and Filippini, R. (2017). "Understanding auditory processing disorder through the FFR" in *The frequency-following response: A window into human communication*. eds. N. Kraus, S. Anderson, T. White-Schwoch, R. Fay and A. Popper (London: Springer International Publishing), 225–250.

Sebastian-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., and Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *J. Cogn. Neurosci.* 18, 1277–1291. doi: 10.1162/jocn.2006.18.8.1277

Skoe, E., Burakiewicz, E., Figueiredo, M., and Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience* 349, 278–290. doi: 10.1016/j.neuroscience.2017.02.049

Skoe, E., and Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324. doi: 10.1097/AUD.0b013e3181cdb272

Song, J. H., Skoe, E., Wong, P. C. M., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902. doi: 10.1162/jocn.2008.20131

Sundara, M., Polka, L., and Molnar, M. (2008). Development of coronal stop perception: bilingual infants keep pace with their monolingual peers. *Cognition* 108, 232–242. doi: 10.1016/j.cognition.2007.12.013

The Jamovi Project. (2023). *Jamovi 2.3*. Available at: https://www.jamovi.org.

The Mathworks Inc. (2019). *MATLAB R2019b,* Natick, Massachusetts.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Lang. Speech* 38, 253–265. doi: 10.1177/002383099503800303

Vouloumanos, A., and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* 10, 159–164. doi: 10.1111/j.1467-7687.2007.00549.x

Weaver, I., Cervoni, N., Champagne, F., D'Alessio, A., Sharma, S., Seckl, J., et al. (2004). Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7, 847–854. doi: 10.1038/nn1276

Werker, J. F., and Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234. doi: 10.1080/15475441.2005.9684216

Werker, J. F., and Hensch, T. (2015). Critical periods in speech perception: new directions. *Annu. Rev. Psychol.* 66, 173–196. doi: 10.1146/annurev-psych-010814-015104

Werker, J. F., and Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev. Psychobiol.* 46, 233–251. doi: 10.1002/dev.20060

White-Schwoch, T., Davies, E. C., Thompson, E. C., Woodruff Carr, K., Nicol, T., Bradlow, A. R., et al. (2015). Auditory-neurophysiological responses to speech during early childhood: effects of background noise. *Hear. Res.* 328, 34–47. doi: 10.1016/j.heares.2015.06.009

frontiers | Frontiers in Human Neuroscience

TYPE Review
PUBLISHED 24 May 2024
DOI 10.3389/fnhum.2024.1384116

OPEN ACCESS

EDITED BY
Marcela Pena,
Pontificia Universidad Católica de Chile, Chile

REVIEWED BY
Hélène Cochet,
Université Toulouse – Jean Jaurès, France
Sachiko Kiyama,
Tohoku University, Japan

*CORRESPONDENCE
Bálint Forgács
✉ forgacs.balint@ppk.elte.hu

RECEIVED 08 February 2024
ACCEPTED 02 May 2024
PUBLISHED 24 May 2024

CITATION
Forgács B (2024) Meaning as mentalization.
Front. Hum. Neurosci. 18:1384116.
doi: 10.3389/fnhum.2024.1384116

Meaning as mentalization

Bálint Forgács[1,2]*

[1]Department of Experimental and Neurocognitive Psychology, Freie Universität Berlin, Berlin, Germany, [2]Department of Cognitive Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

The way we establish meaning has been a profound question not only in language research but in developmental science as well. The relation between linguistic form and content has been loosened up in recent pragmatic approaches to communication, showing that code-based models of language comprehension must be augmented by context-sensitive, pragmatic-inferential mechanisms to recover the speaker's intended meaning. Language acquisition has traditionally been thought to involve building a mental lexicon and extracting syntactic rules from noisy linguistic input, while communicative-pragmatic inferences have also been argued to be indispensable. Recent research findings exploring the electrophysiological indicator of semantic processing, the N400, have raised serious questions about the traditional separation between semantic decoding and pragmatic inferential processes. The N400 appears to be sensitive to mentalization—the ability to attribute beliefs to social partners—already from its developmental onset. This finding raises the possibility that mentalization may not simply contribute to pragmatic inferences that enrich linguistic decoding processes but that the semantic system may be functioning in a fundamentally mentalistic manner. The present review first summarizes the key contributions of pragmatic models of communication to language comprehension. Then, it provides an overview of how communicative intentions are interpreted in developmental theories of communication, with a special emphasis on mentalization. Next, it discusses the sensitivity of infants to the information-transmitting potential of language, their ability to pick up its code-like features, and their capacity to track language comprehension of social partners using mentalization. In conclusion, I argue that the recovery of meaning during linguistic communication is not adequately modeled as a process of code-based semantic retrieval complemented by pragmatic inferences. Instead, the semantic system may establish meaning, as intended, during language comprehension and acquisition through mentalistic attribution of content to communicative partners.

KEYWORDS

language comprehension, social cognition, semantic processing, mentalization, theory-of-mind, N400, language acquisition, pragmatics

1 Introduction

This study presents a new perspective on how content is transmitted during linguistic communication by proposing a novel theory of how meaning is established in the human mind. It offers an explanation for language acquisition and word learning from the perspective of mentalization—that is, the attribution of intentions, beliefs, and desires to social partners (Premack and Woodruff, 1978; Leslie, 1987). The foundation for this novel model of processing, establishing, and acquiring semantic content is based on a series of neurocognitive experiments with infants and adults. However, before introducing these studies, the broader question of the interplay between human communication, social cognition, and language comprehension will be addressed. First, I will take a closer look at the changes in thinking regarding the transmission of linguistic meaning, from the code model to pragmatic theories of communication. Next, I will explore the role of communicative intentions and how they enable language comprehension and acquisition. Then, I will discuss how social cognition is involved in utilizing language as an information transmission device and how infants employ mentalization to track the comprehension of communicative partners. Finally, I will argue that semantic processing involves the attribution of mental content through mentalization and that such mentalistic meaning-making drives and enables language acquisition and word learning.

The main claim of this study is that contrary to standard models of language comprehension, linguistic meaning does not emerge from decoding information by looking up semantic content in a mental lexicon and placing it in syntactic frames, nor from the applying rule- or relevance-based social-pragmatic inference mechanisms. While lexical retrieval and pragmatic enrichment play important roles in language processing, I argue that comprehension of meaning, as intended, fundamentally relies on attributing mental content to communicative agents as belief states. While some approaches recognize the importance of mentalization in communication, they limit its role to setting up communicative interactions (Tomasello, 2008) or reference resolution (Bloom, 2000). Both assign social cognition a key role, which is to link mental representations (of the physical world) to word forms. What distinguishes the current approach is that meaning is not identified externally in the physical world, with the help of social cognition, but internally in the mental world of communicative partners as the content of attributions of beliefs about the world.

2 Form and content in language

The question of how meaning is established, transmitted, and acquired is a matter of heated debate not only in linguistics but also in psychological science. In the 1950s, the dominant structuralist view on language was challenged from multiple directions. The basic assumption of this tradition was the equivalence between form and content. In contrast, the new approaches pointed out that the comprehension of linguistic meaning is only partly based on interpreting language as a code, and external factors such as communicative intentions, context, and social cognition may also play key roles.

The founder of structuralism, de Saussure, noted the arbitrariness of the connection between signifiers (form) and signified (content). His Linearity Principle promised that analyzing the sequences of signifiers would provide a systematic explanation of content (de Saussure, 1966). It also complemented Frege's Compositionality Principle, which proposes that linguistic meaning is based on a systematic derivation of the truth-value of linguistic propositions or sentences, viewed as functions of the grammatical combination of words (Frege, 1948). Structuralism and the idea of unity between form and content remain highly influential in developmental psychology. It often serves as the hidden axiomatic assumption behind the acquisition of word-to-world mappings during word learning and the source of the expectation that language acquisition is a gradual, step-by-step process that proceeds from phonology through word learning to grammar.

The idealization of *form as content* served as the foundation for Shannon and Weaver's information theory, which provided a mathematical formalization of communication (Shannon and Weaver, 1949). It is based on the code model, which is still the textbook model of human communication (Blackburn, 2007). This model proposes that an information source, or sender, encodes its message via a transmitter, which then sends the signal through a channel. Upon receipt, the receiver reconstructs the message by decoding it from the signal. Modeled after the telegraph, communication is formalized here as the challenge of recovering the message from signals received through a noisy transmission channel, while it takes it for granted that messages are clearly defined chunks of information and unambiguous in their content once decoded.

In the 1950s, a series of theories challenged the structuralist tradition, although not the code model itself. Wittgenstein pointed out that the relationship between form and content may be far looser than previously assumed (Wittgenstein, 1953). Chomsky proposed that linguistic meaning is, in fact, recovered from syntactic deep structures rather than from the surface forms of word sequences (Chomsky, 1957, 2015, 2017). Both of these ideas became highly influential (Pléh, 2024). Around the same time, arguments developed by two philosophers of language led to the establishment of the field of linguistic pragmatics. Austin suggested in 1955 that certain kinds of sentences function as speech acts (e.g., "thank you" or "excuse me") that we do not evaluate based on their literal truth-value but in terms of their social force. We recognize them as not describing reality but rather bringing about some intended change in the world (Austin, 1962; Searle, 1969). Intentions were introduced as central to communication, but as Reboul and Moeschler (1998) pointed out, this approach remained unsuccessful because it tried to account for intended meaning in terms of linguistic conventions and formulae. Paul Grice was the first to suggest that communicative intentions play a key role in conveying meaning via non-linguistic inferences (Grice, 1957). He differentiated between an indicative sense of the word "meaning" (e.g., clouds may foreshadow rain) and a "meant by" sense (e.g., if someone has their head in the clouds, it expresses they live in a fantasy). Form and content are decoupled thereby: the meaning of utterances cannot be recovered by simply decoding the lexical contents of word combinations without considering speakers' intentions in the communicative context. Grice did not elaborate on the structure or content of intentional mental states but showed that they are not only indispensable for communication but also independent of the code structure of language. When we hear the word "tiger," we do not automatically run away assuming that the word refers to an actual tiger present in the here and now: we understand it as a communicative act, not of "indicating" but of "meaning by."

Grice nevertheless anchored his theory in the code model when he suggested that "what was said" needs to be decoded first, based on its literal meaning and truth-value. Only afterward may one engage in the inferential mechanisms necessary to recover the implicatures, the intended meanings (Grice, 1975). Thereby, sentences are decoupled from utterances: sentences are still processed as code-based signals, but utterances become pragmatic-inferential interpretations. The precondition for these inferences is that both parties are interested in holding a conversation based on truth and trust—a concept for which Grice proposed the Cooperative Principle. Speakers may strategically violate either of four maxims (quality, quantity, relevance, and manner), the normative rules of conversations, to prompt hearers to engage in the inferential reverse engineering of the intended meaning. Here, two sets of rules exist: one for the linguistic code and another for the maxims, akin to constitutive and regulative rules, with the former establishing the framework and the latter navigating it (Black, 1962). Reboul and Moeschler (1998) argue that Grice's theory, while underspecified to be truly cognitive, was a game changer that eventually led to the birth of experimental pragmatics (Noveck and Sperber, 2004; Noveck and Reboul, 2008; Bambini and Bara, 2010; Noveck, 2018).

Form and content are even more strongly decoupled in Relevance Theory (RT) (Sperber and Wilson, 1986; Wilson and Sperber, 2004). RT breaks with the idea that linguistic form—"what was said"—can be recovered solely based on decoding. Sperber and Wilson suggest that the language module provides an initial interpretation by constructing a logical form that serves as the premise for the inferential mechanisms interpreting utterances. However, pragmatic inferences play a role already in uncovering what was said (the explicatures), not only in what was meant (the implicatures). They also reject Grice's Cooperative Principle and keep only one of the maxims: relevance. They suggest that relevance seeking is a general mechanism of cognition that aims to maximize effects by minimizing efforts, and it drives language comprehension as well. Instead of cooperation and normative rules, they introduce the concept of cognitive environment, which, along with the logical form, constitutes the inputs of the pragmatic inference machine. It includes physical and perceptual information as well as common knowledge, common history, and common ground. Although most examples provided by Sperber and Wilson involve mental states, on-line mentalization serves merely as an optional input for pragmatic enrichment (Mazzarella and Noveck, 2021). Alas, pragmatic-inferential mechanisms do not necessarily require cooperation or even the attribution of intentional states. Social cognition manifests in RT as ostensive communicative signals, which are attention-capturing acts that trigger relevance-seeking processes during communicative interactions.

Taken together, there has been a gradual but tectonic shift that has transformed thinking about how language conveys meaning: away from the structuralist tradition of meaning carried by linguistic forms in a code-like manner and toward meaning inferred as intended (Figure 1). While some theorists still argue that pragmatic processes enter language comprehension only when things go wrong, and until that point, language works like a code (Millikan, 2005), there now seems to be broad agreement that language is not interpreted directly through decoding. Even though the nature of inferences is hotly debated, they also all appear to be rule-based mechanisms of social cognition that do not involve the attribution of intentional or mental states to communicative partners to establish intended meaning. The cooperation-based Neo-Griceans tradition (Grice, 1975; Levinson, 2000; Horn, 2006;

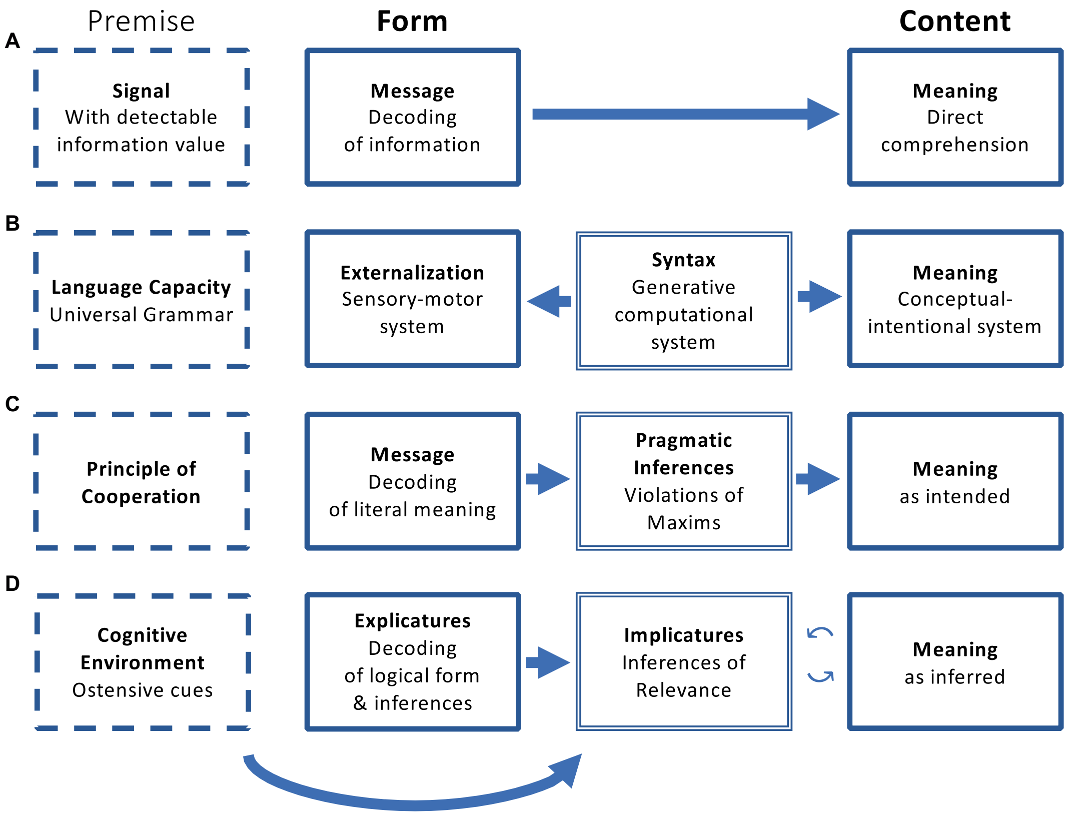


FIGURE 1
The fundamental shift in how meaning is thought to be conveyed by language, as per **(A)** the code model (Shannon and Weaver, 1949); **(B)** the generativist syntax-first approach (Chomsky, 2017); the pragmatic models of **(C)** the Gricean and Neo-Gricean (Grice, 1975); and **(D)** the Post-Gricean traditions (Sperber and Wilson, 1986). Note that "decoding" typically implies the processing of both semantic and syntactic information. Additionally, pragmatic inferences may involve some kind of social cognition (cooperation and/or ostension), but communicative intentions are thought to be derived with no attribution of mental/belief/intentional states to communicative partners.

Goodman and Frank, 2016) argues that hearers infer the intended meaning of speakers in a serial fashion, first decoding the literal meaning of spoken language, then looking for violations (of norms/maxims or rationality/utility). The ostension-based Post-Gricean Relevance Theory camp (Sperber and Wilson, 1986; Noveck and Sperber, 2004; Noveck, 2018) suggests that hearers seek relevance while considering the linguistic input and the cognitive environment in parallel to develop implicatures. The two models agree on the central role of social cognition and communicative intentions, yet neither has put forward mechanisms that were based on mentalization, which has generated a longstanding debate about whether communication involves mentalization or not (Pléh, 2000; Bosco et al., 2018). Moreover, the term "communicative intention" is often used ambiguously in at least two different senses.

3 Communicative intentions in human communication

The notion of intentionality traces back to Brentano's reintroduction of scholastic ideas. He suggested that the hallmark of psychological or mental states, such as beliefs and desires, is a kind of "aboutness": mental events, as opposed to physical objects, are directed toward entities beyond themselves (Brentano, 2009). The question of higher-order intentionality of mental states, such as beliefs about others' beliefs, burst into the scientific discourse in psychology with the debate about whether chimpanzees have a Theory-of-Mind (ToM) (Bennett, 1978; Dennett, 1978; Premack and Woodruff, 1978). ToM is the ability to attribute psychological states with intentionality to social partners (Jacob, 2023). In his highly influential works, Dennett proposed that humans take the "intentional stance" to predict and explain the behavior of social agents (including humans, animals, and even machines) by attributing intentional mental states, beliefs, and desires to them (Dennett, 1987).

The concept of communicative intentions has been used somewhat differently in the field of pragmatics, where it has been proposed that they may simply be recognized based on ostensive-behavioral signals without necessarily ascribing mental content to others (Sperber and Wilson, 1986). It has also been suggested that the earmark of human communication is not simply that it is intentional but rather that it is overtly intentional (Scott-Phillips, 2015). Non-human primates may be able to communicate intentionally, perhaps inferentially, but not truly ostensively (Warren and Call, 2022). Overtly intentional ostensive communicative signals call the attention of their addressee to the communicative act itself. The communicative transmission commences when an agent's communicative intention is recognized by the addressee; the communicative intention is fulfilled when a second, informative intention is also recognized. The recognition of an informative intention is equal to comprehension and its fulfillment to believing the content. In other words, the recognition of the communicative intention opens a unique kind of communication channel, suspended from the present here and now, which allows for the transmission of information with no reference to the physical environment. In primate communication, signals may be sent informatively, even intentionally (e.g., when producing a predator alarm call to warn conspecifics), but in lack of highlighting and recognizing the communicative intention, it does not seem possible to exchange communicative signals about objects (or dangers) beyond perception. It is a species-specific feat of human communication that communicative transmissions can be about entities beyond the local physical surroundings and the present moment. Signals following the recognition of communicative intentions are suspended from perceptual reality, yet they still seem to possess a kind of aboutness akin to intentional mental states.

3.1 Communicative intentions in language acquisition

The critical importance of the social environment in the emergence of language was put forward by Vygotsky (1978) and Bruner (1983). In their studies, which laid the foundations for a social constructivist view on language, the social world is discussed as a special kind of environment, distinct from the physical world. It is an indispensable yet external context for learning and development, not an internal matter of mind and cognition (Pléh, 2024). Additionally, their basic assumption, like many others', has been that communication requires a code-based signaling system, that is, language. The idea that human communication is built on social cognition, as it is not mere information transmission, reverses the above order: communication may need to precede language acquisition. There are two dominant views regarding the developmental origins of human-specific communication and the role social cognition plays in it. Tomasello (2008), by and large, follows the (Neo-) Gricean tradition in emphasizing cooperation as the basis of communication and language in his joint attention framework; Gergely and Csibra's Natural Pedagogy theory fits well with RT and the Post-Gricean perspective when proposing that ostensive cues play a central role in communicative interactions even in preverbal infants (Csibra and Gergely, 2009; Gergely and Csibra, 2013).

3.1.1 Tomasello's shared intentionality infrastructure

To "break into the code" of language (as Tomasello puts it) to decipher its syntactic and semantic structures, children need to be able to communicate in some way from the outset. Tomasello argues that this initial form of communication is founded upon a dedicated "cognitive infrastructure" of "shared intentionality." Pointing and pantomiming are not based on preestablished conventional codes, yet infants can use them communicatively. These preverbal communicative acts presuppose sensitivity to cooperation (missing in our closest primate relatives) and are based on shared intentionality, an understanding that *we*, as social partners, may intend things together. Having a joint goal is not merely two agents having the same goal simultaneously, like aiming to reach the same destination, but like walking together. Humans' propensity for cooperation to pursue joint goals provides the context in which communicative acts such as pointing or pantomiming acquire a shared meaning. This joint context, the common ground (Clark, 1996), is indispensable for interpreting actions as communicative gestures. It is established by what Tomasello calls "joint attention," when participants are aware that they simultaneously attend to the same object. Tomasello specifies three basic human communicative intentions enabled by the above cooperative infrastructure: requesting, informing, and sharing. To fulfill such intentions, participants need to reason not simply practically (i.e., rationally) but cooperatively, relying on their partners' helping attitude.

This kind of human cooperation that emerges around 9–12 months of age in human ontogeny, Tomasello argues, requires "recursive intention-reading and mind-reading abilities." Phylogenetically, these abilities are supposed to originate from the ability to establish joint goals, which led to the emergence of joint attention and eventually enabled the establishing of common ground. At the same time, Tomasello also argues that children under 4 years of age possess only rudimentary mentalization abilities akin to the ToM of great apes (Tomasello, 2018)—which is nevertheless still sufficient to support the recursive mind-reading necessary for the earliest forms of human communication. Shared intentionality is thus a feat of cooperation, not of ToM. Others have also argued that some level of mentalization may be indispensable for the attribution of goals (Csibra and Gergely, 2007) or attention (Elekes and Király, 2021), but it is not clear how sophisticated these ToM representations need to be to support "shared goals" and "joint attention." In Tomasello's view, recursive mindreading is present already in 9-month-olds, but it does not seem to play a significant role in the transmission of information, as the shared intentionality infrastructure and attention-checking may be sufficient to sustain the earliest forms of human communicative interactions.

Such a framing, while placing cooperation at the center, turns Grice's account upside down. Instead of first decoding content and then enriching it with cooperation-based pragmatic inferences, it is now sensitivity to cooperation that allows for information transmission. It is the common ground that allows for interpreting, in Tomasello's terms, the "natural" signals of pointing and pantomiming and eventually the "conventional" linguistic signals. Language, however, is just a ritualization of communicative interactions, not qualitatively different from non-conventional forms of cooperative communication. Tomasello points out that the first words appear in and emerge from cooperative routines (Bruner, 1983) when infants understand the intentional structure of the shared goal and begin to reason cooperatively (Tomasello, 2008).

3.1.2 The ostensive signals of Natural Pedagogy

Gergely and Csibra's Natural Pedagogy theory sidesteps the issue of mindreading when it proposes that perceptually identifiable ostensive-behavioral signals may account for human-specific communicative interactions (Csibra and Gergely, 2009). In contrast to RT, the primary role of ostensive signals is not to support inferential communication but to enable cultural learning (Gergely and Csibra, 2005). They allow learners to identify and recognize the communicative, pedagogical intention of knowledgeable partners to transmit culturally relevant knowledge (Gergely and Csibra, 2013). Here, ostensive signals do not simply function to capture attention, as in RT; rather, particular attention-grabbing signals are employed to induce cultural learning because they are ostensive (Gergely, 2010). There is evidence for at least three, perhaps innately specified ostensive signals that may be recognized as indicating communicative intentions already around 4–5 months of age (Grossmann et al., 2007; Parise et al., 2008; Parise and Csibra, 2013): eye-contact, contingent (turn-taking) reactivity, and infant-directed speech. While code-based signals, in general, are poor means of information transmission without pragmatic inferences, ostensive signals can be utilized in a code-like manner to induce the recognition of communicative intentions (Csibra, 2010).

Natural Pedagogy puts forward a mechanism for establishing common ground not based on cooperation but on evolved, trust-based procedures that pick out certain behavioral cues as ostensive signals. These signals do not merely activate attentional resources but also initiate species-unique cultural learning strategies. When addressed ostensively, infants assume that the information transmitted is generalizable, socially and culturally shared, and constitutes normative knowledge (Gergely et al., 2002; Yoon et al., 2008; Futó et al., 2010; Hernik and Csibra, 2015). One outstanding example of cultural learning in natural pedagogical situations is the acquisition of word labels, which are indeed socially shared and mostly refer to kinds. Taken together, Natural Pedagogy is an evolved system that serves social knowledge transmission underlying cultural learning just as well as human communication.

3.2 Mentalization and communicative intentions?

Although the Neo-Gricean and the Post-Gricean models are not mutually exclusive, they emphasize different aspects of communicative interactions and assume different sufficiency and necessity conditions for establishing common ground. Both models are primarily interested in explaining how to identify communicative intentions, with less emphasis on how information is actually transmitted (i.e., the informative intention). While communicative intentions may be recognized either through a code-like signaling system (ostension) or through the motivation for cooperation (joint attention), neither model argues for the necessity of mentalization, even though both involve attending to a social partner and identifying the partner's focus of attention as a referent.

Another person's attention or goal (i.e., the referent of a communicative interaction), however, cannot be but an attribution of attention (Elekes and Király, 2021) or of a goal (Csibra and Gergely, 2007). It may be argued that recognizing the attention or the goals of social partners relies on some sort of non-mentalistic mechanism, for example, on teleology (Gergely and Csibra, 2003). However, teleology does not seem to suffice to explain communication non-mentalistically because in communicative interactions, referent objects are not the goals of agents but the goals of the communication itself. Moreover, information transmission takes place only after minds establish common ground through joint attention and/or ostensive signals. Yet, minds never literally "join"; thus, common ground can emerge only in the minds of the two interlocutors separately. The two parties may mutually assume that the other has an identical belief about the state of affairs as themselves, but this can only be an ascription of a mental state to the other party. The point is that the machineries proposed for setting up communicative interactions by identifying communicative intentions are fundamentally attention-directing and attention-checking systems. These systems are aimed at identifying what the partner is calling the infant's attention to, which is the partner's own focus of attention, to establish that both parties have the same referent in mind. Perhaps due to the ambiguity of the term "communicative intentions," a considerable group of researchers seems to believe it is obvious that language, like communication, involves mentalization, while another large group appears to hold it is rather obvious that neither communication nor language does.

3.3 Timing and difficulty of inferring communicative intentions?

There is another intriguing contradiction regarding pragmatic inferences, as highlighted by Bohn and Frank (2019). There is a long line of research arguing that infants reason skillfully about intentions already during language acquisition (Nelson, 1973; Bates, 1976; Bloom, 2000; Tomasello, 2003). Another line of research reports the difficulties children experience in deriving pragmatic inferences to interpret intended meanings (Papafragou and Musolino, 2003; Huang and Snedeker, 2009). Bohn and Frank propose to resolve the contradiction by defining communication as social cognition and reasoning about the goals of communicative partners. The Rational Speech Act framework suggests that pragmatic reasoning integrates all the elements of inferential communication, which are in place early on and foster the gradual development of language comprehension (Bohn and Frank, 2019).

The apparent contradiction may stem from the ambiguity of the term "communicative intentions." It may refer to the social-cognitive inferential mechanisms employed to identify the *intent to communicate*, which may be present from early infancy onward on the one hand. It may also refer to the inferential mechanisms applied to the transmitted information content to recover *meaning as intended*, which may be challenging even for kids, on the other. While the former is the utilization of pragmatic-inferential mechanisms to set up a communicative infrastructure for the upcoming information (*cf.* Tomasello), the latter involves enriching the already available information with pragmatic inferences (*cf.* Grice).

The above two kinds of ambiguity surrounding the term "communicative intention"—(1) whether it is mentalistic or not and (2) whether interpreting it appropriately is easy (already in infancy) or difficult (even in late childhood) —seem to be orthogonal (Table 1). Some argue that inferring communicative intentions is essential for word learning, is available already in infancy, and involves mentalization (Bloom, 2000; Papafragou, 2002; Tomasello, 2003; Thompson, 2014). Others assume that word learning is based on inferential mechanisms, but it does not require mentalization (Sperber and Wilson, 1986; Csibra and Gergely, 2009; Bohn and Frank, 2019). Some show how non-mentalistic pragmatic inferences are challenging for kids when interpreting, for example, scalar inferences (Papafragou and Musolino, 2003; Huang and Snedeker, 2009; Noveck, 2018), contrastive inferences (Kronmüller et al., 2014), or logical terms (Noveck and Chevaux, 2002; Pouscoulous and Noveck, 2009). Finally, the position that pragmatic inferences develop slowly but co-develop with and involve mentalization has also been put forward (Rubio-Fernandez, 2021). The debate boils down to two key questions: (1) can communicative intentions be truly non-mentalistic (i.e., inferred based on code-like signals or regulative rules) and (2) what level of mentalization may be available in infancy (i.e., to what extent language acquisition may or should be tied to it)? Both of these questions are going to be addressed later on.

Taken together, there seems to be an agreement that social cognition, in the form of pragmatic inferences, plays a key role both in setting up communicative interactions and in deriving the implied meaning of utterances behind words and sentences. The first mechanism appears to precede the transmission of the linguistic code (i.e., setting up a communicative interaction, either via cooperation and joint attention or via ostensive communicative signals), while the second one follows it (i.e., enriching it inferentially to interpret it, either via checking for violations of maxims in a cooperative framework or via carrying out a relevance-based calculus). The first one seems to pertain to "communicative intentions" in the narrow sense: a scaffold of informative intentions. The second one uses the term in the broad sense: inferring meaning as intended. It may be argued that the latter actually concerns informative intentions, but this is not entirely clear from the literature. "Intended meaning" is typically referred to as what was intended to be conveyed (i.e., communicated), not what one was intended to be informed of. The term "information" may be the culprit here: it could mean either the content (utterance) or the form (sentence)—perhaps due to the remarkable influence of the code model. Notably, the question of how social cognition modulates the information content, specifically the link between linguistic form and semantic content, seems to have gathered limited attention. Yet, the linguistic signal is the most variable and rich source of input entering pragmatic inferential mechanisms.

4 Language as an information transmission device

While human communication may be viewed as a form of social cognition based on pragmatic inferences that enable language acquisition and comprehension (Tomasello, 2008; Bohn and Frank, 2019), it may also be viewed as a tool for information transmission (Shannon and Weaver, 1949; Tauzin and Gergely, 2018, 2019). Of course, these two views are not mutually exclusive, and, in some sense, they represent two sides of the same coin. Nevertheless, they still represent fundamentally different views on the role language plays in communication. In the former view, meaning emerges primarily from the common ground and the structure of the social interaction (Clark, 1996; Tomasello, 2008; Bohn and Frank, 2019), while in the latter, it arises from the properties and the variability of the signal, perhaps in interaction with mental states (Tauzin and Gergely, 2018).

The signal variability approach gains particular relevance due to the richness of the contents that may be transmitted in spoken language, from storytelling to discussions of shared memories. The communicative goal of verbal interactions is often far beyond the social situation and maybe more intricate than the basic intentions of requesting, informing, or sharing (Tomasello, 2008). Narrative stories rely heavily on information transmission to set up the communicational situation itself. It follows that the information content—the intended meaning—may be at least partly recoverable from the mental state of communicative partners rather than solely

TABLE 1 The various interpretations and uses of the term "communicative intentions" in the literature.

Communicative intentions	Mentalistic	Non-mentalistic
Easy to infer	Bloom (2000), Papafragou (2002), Thompson (2014), and Tomasello (2003)	Bohn and Frank (2019), Csibra and Gergely (2009), and Sperber and Wilson (1986)
Difficult to infer	Rubio-Fernandez (2021)	Huang and Snedeker (2009), Noveck (2018), Noveck and Chevaux (2002), Papafragou and Musolino (2003), and Pouscoulous and Noveck (2009)

from the common ground or the situational features of the cognitive environment. Linguistic forms may have reached the unmatched level of signal complexity precisely because they may be more about what the partner could have in mind and less about the social interactions of relatively limited complexity, especially those in which non-human and young human apes typically engage.

4.1 Sensitivity to the code-like features of language in neonates

From an information transmission point of view, it may not be surprising that language, as a stimulus class, enjoys a special status in human ontogeny. Well before birth, prenatal humans begin to pick up the prosodic properties of their native tongue (Abboub et al., 2016), and right at birth, they are sensitive to a range of physical-acoustic and phonological features of language (Werker and Tees, 1999). Newborns prefer speech to matched non-speech, forward-going speech to backward speech, their mother's voice to other female voices, and their native language to unfamiliar languages; they can also differentiate between languages based on rhythmic properties even if they have never heard them before and can detect word boundaries, discriminate lexical stress, and even distinguish function words from content words based on acoustic characteristics (reviewed by Gervain and Mehler, 2010).

During language acquisition, infants use several of these acoustic properties, including rhythm (Goswami, 2022) and statistical distributional patterns (Kujala et al., 2023), to "break into the code" of language by approaching it from its information-transmitting potential. The acoustic-phonological-prosodic properties of language lend themselves readily to being deciphered as a code. First, the set of sounds the human vocal tract can produce is limited, making it computationally manageable. Second, phonology is not only governed by rules but also carries information about higher-level syntactic operations, both of which involve code-based computational structures. Even newborns rely on prosody and statistical learning of transitional probabilities to identify word boundaries (Fló et al., 2019). To segment the continuous speech stream into potentially meaningful units, they also employ various dedicated mechanisms to learn about the segments themselves (Fló et al., 2022). They also utilize innate pattern recognition mechanisms to pick up repetition structures (i.e., pseudowords with ABB structure, e.g., "mi-zu-zu," as opposed to ABC random structures, e.g., "mi-zu-ka") (Gervain et al., 2008). The repetitions may be engaging for them because they reveal a rule that may indicate syntactic structures, thereby providing a better opportunity to learn (about) language. Such sensitivity to rule-like structures extends to musical tones as well; however, only pseudowords, not tones, activate the left inferior frontal regions (Nallet et al., 2023). These regions include Broca's area, which is responsible for processing the structural properties of language in adults (Musso et al., 2003; Liakakis et al., 2011; Friederici, 2012) and responds to language at birth (Peña et al., 2003; Perani et al., 2011). Newborns detect not only repetition structures but also their sequential position (ABB vs. AAB), and these two properties seem to be the two fundamental building blocks of any code-based system (Gervain et al., 2012). These findings strongly suggest that language is a unique signal for humans, engaging dedicated mechanisms, from statistical learning to pattern recognition, to identify word-like units and grammar-like rules right from birth. Newborns appear to be very well equipped to unpack the structural properties of the code system humans use to transmit information.

4.2 Communicative self-referentiality in the speech signal

In the process of acquiring linguistic meaning, the only communicative cue that has been suggested to signal communicative intentions and is linguistic in nature is infant-directed speech (IDS) (Csibra and Gergely, 2009; Gergely, 2010); also called "motherese," its characteristic prosodic pattern includes higher and broader pitch, greater amplitude variation, and slower speed than typical adult speech (Csibra, 2010). Although there seems to be some cultural variation (Cristia, 2023), sensitivity to IDS appears to be innate, present at birth (Cooper and Aslin, 1990), and universal (Fernald, 1992). The perceptual features of motherese appear to open the gateway toward the content of speech: IDS may simultaneously carry information about communicative and informative intentions. In fact, Sperber and Wilson (1986) consider language to be an ostensive cue in and of itself. Every communicative act carries its own relevance by definition, and a linguistic utterance is clearly communicative—at least for adults. Infants may exploit the prosodic layer of IDS for communicative intentions to gain access to its contents (i.e., informative intentions). IDS modulates electrophysiological responses to faces in 4-month-olds, perhaps because it generates communicative expectations (Sirri et al., 2020). It is interpreted as an ostensive cue by 5–6-month-olds, just as eye-gaze (Senju and Csibra, 2008; Parise and Csibra, 2013; Lloyd-Fox et al., 2015). It also facilitates 7-month-olds' cortical tracking of speech (Kalashnikova et al., 2018). IDS appears to serve as a self-referential linguistic inroad toward linguistic meaning as it induces a communicative interpretation of the linguistic code. As an ostensive signal, it creates an expectation that incoming information refers to kinds and not individuals (Gergely and Csibra, 2013), which clearly aids word learning since, with the exception of proper names, words refer to categories.

4.3 Recognizing the communicative function of language

Infants also appear to realize early on that language may carry information. In a series of remarkable experiments, Vouloumanos and colleagues (Martin et al., 2012; Vouloumanos et al., 2012; Vouloumanos, 2018) showed that infants expect speech—but not coughing or humming—to transmit information about intentions (i.e., objects preferences). Even 6-month-olds showed this expectation even when they had no chance to understand the transmission because it was non-sensical or foreign to them (Vouloumanos et al., 2014). Eleven-month-olds may prefer to interact with native speakers because they expect them to share information (Begus et al., 2016). Intriguingly, signal properties may modulate such expectations.

Using the so-called Flatfish paradigm, Tauzin and Gergely demonstrated that 10.5-month-old infants identify beeping entities as agents with preferences, but only if they exchange varying tone signals—not if they parrot each other using identical signals (Tauzin and Gergely, 2019). The authors argue that, from an information theoretical perspective, there are two elementary building blocks of communication: (1) there are two agents taking turns exchanging signals, and (2) the signals vary in an optimal way, with a high-but-imperfect level of contingency. The exchanged signals need to be similar enough to form a correspondence yet different enough to carry added information value (Tauzin and Gergely, 2021). At 13 months of age, infants assume a flatfish to update its falsely held belief about the location of an object only after an optimally variable signal exchange with another flatfish (Tauzin and Gergely, 2018). Communication, as information transmission, may directly modify mental state attribution through signal variation, irrespective of social-pragmatic inferences.

According to a recent study, humans may be sensitive to the information transmission value of language already at birth. When presented with grammatically structured ABB pseudowords, but now as an exchange between a female and a male voice, newborns showed increased activity near Broca's area when the pseudowords were different tokens (female: kamumu; male: dekiki) compared to when they were identical (male: bulili; female: bulili) (Forgács et al., 2022a). These findings demonstrate, first, that neonates can identify the possibility of information transmission in communicative interactions, even when they may have no idea about its contents. Notably, they can do so even when they are not participants in the interaction and without relying on social cognition. If turn-taking were interesting in and of itself, there should have been no difference between the identical and variable signal exchanges. Second, the activation of Broca's area suggests that it is the language-processing region of the brain that responds to the possibility of information transmission. Processing the potential for information transmission may be a core feature of human language.

The sensitivity to information value is independent of any particular semantic content or the structure of social interactions. In contrast, it is markedly missing from pragmatic models of human communication (Sperber and Wilson, 1986; Grice, 1989; Tomasello, 2008; Goodman and Frank, 2016). These models assume that information is transmitted as a code and then enriched and inferentially unpacked by social cognition. Yet, humans seem to be sensitive to information transmission even without knowing any code. Humans at birth appear to possess a structured representational template of an informative intention embedded within a communicative intention, allowing them to identify communicative intentions even when the embedded representational slot for the informative intention remains empty and even in the absence of social cues directed toward them. Of course, the above study did not provide direct evidence of a second-order representation or the recognition of a communicative intention or the relationship between the two; thus, these interpretations remain just as hypothetical as they are for infants (Csibra, 2010). Nevertheless, the underlying information estimation mechanism may be a third route for identifying communicative intentions alongside joint attention and ostension. Note that the ostensive cue of turn-taking is based on tracking proximal contingencies in interactions infants are part of, while the information estimation route capitalizes on distal contingencies.

Newborns' sensitivity to information structure implies that humans may assume the existence of a code with content that can be sent and received, which may be just as important to language acquisition and processing as syntax and social cognition. This possibility is in sharp contrast with both Chomsky's and Tomasello's proposals. Chomsky suggests that syntax is the core feature of cognition in the form of recursion (Chomsky, 1965; Hauser et al., 2002) or merge (Berwick and Chomsky, 2011). However, its externalization, spoken language and communication are of no particular interest. Tomasello's (2008) work implies that information can be transmitted communicatively only once the cognitive infrastructure for shared intentionality emerges. The notion that information transmission may be identified based on signal variability hints that humans may be able to enter the suspended space of human communication beyond the here and now, right from the very beginning of life—and language acquisition. Information may be identified without awareness of any form, content, or social context. But how may the actual meaning of language be figured out?

5 The emergence of linguistic meaning

Humans arrive in our world with an impressive cognitive arsenal to acquire language. They are well-prepared to unpack the code-like features and structures of phonology and syntax. They are endowed with inferential tools of social cognition to engage in human communication and have some understanding of information transmission to recognize communicative intentions. For information content not only to be identified but also to be learned, that is, for linguistic forms to be connected to conceptual knowledge, meaning needs to emerge within communicative interactions.

Social cognition has been proposed to play an important role in communication, even in the animal kingdom (Fitch et al., 2010), and to aid language acquisition throughout human development. There is a rich literature on how gaze-following (Brooks and Meltzoff, 2008, 2013), indicative of attention (Baldwin and Markman, 1989; Baldwin, 1993), or perspective-taking (Nadig and Sedivy, 2002; Nilsen and Graham, 2009; Khu et al., 2018) enables reference resolution; on how infants are able to exploit ostensive eye-gaze and pointing (Behne et al., 2005), iconic gestures (Bohn et al., 2019), and ostensive cues in communicative situations (Egyed et al., 2013); or on how the ability to use gaze, pointing, and other communicative gestures fosters later referential language production (Carpenter et al., 1998). The list is long, with excellent reviews (Clark and Amaral, 2010; Bohn and Frank, 2019) and meta-analyses available (Lewis et al., 2016; Bergmann et al., 2018). More radical forms of pragmatic-constructivist theories of language acquisition suggest that instead of word-referent mappings (Bloom, 2000), meaning is based on usage (Tomasello, 2003) or that usage may even start without meaning (Nelson, 2009). The broad agreement in developmental science is that social cognition, in the form of pragmatic inferences, plays a fundamental role in language acquisition.

In this social-pragmatic line of research, meaning is traced back to communicative intentions but is inferred from the social context, not attributed to the communicative partner. Communicative intentions are supposed to be formed by *assuming* goals or attention, not by *attributing* intentionality. Moreover, most, if not all, pragmatic inference mechanisms involving gaze-following, pointing, perspective taking, or gestures mainly, if not exclusively, target reference resolution—the content of communicative exchanges. Reference resolution is the point at which the promiscuously used term "communicative intention" switches from its sense of "intending to behave communicatively" to its other sense of "intending to express a particular meaning." Meaning, as identified by attention and goal tracking mechanisms, is assumed to be linked to a referent in the outside world in the form of an object (nouns), an action (verbs), or a property (adjectives). The role of social cognition is to narrow the communicative interaction to the appropriate property of the physical environment, but it does not have much to do with meaning *per se*. While it is controversial whether communicative intentions involve mentalization, at least in the strict sense of attributing false beliefs, whether informative intentions—i.e., referential information transmission—may have anything to do with mindreading is not even considered.

Moreover, it is not entirely clear when and where meaning, as comprehension, emerges during communication. In the Neo-Gricean tradition, inferences are applied only after a literal meaning is decoded. The informative intention is treated as the code and the communicative intention as the pragmatic inference; thus, meaning emerges after the second step. In the Post-Gricean approach, once communicative intentions are recognized, inferences are employed to develop both explicatures and implicatures. It is the recognition of the informative intention that yields an accurate interpretation of meaning as inferred. Thereby, communicative intentions could contribute to meaning in two ways. Either in the broad sense by inferring the intended meaning (i.e., deriving implicatures) at a late stage. Or in the narrow sense, intending to communicate at an early stage. In the latter case, however, meaning (i.e., implicature) is computed at the level of embedded informative intentions.

Whether any of the pragmatic inferences employed to derive meaning involve mentalization remains unresolved. First, for the Neo-Griceans, ToM may contribute to communication either before language processing proper (*cf.* Tomasello's shared intentionality) or after decoding, during the pragmatic inference stage (*cf.* Grice's enrichment). Even though the Rational Speech Act theory suggests a fully integrated mechanism (Bohn et al., 2021), it is still based on literal meaning, which presupposes encapsulated decoding (Goodman and Frank, 2016; Bohn and Frank, 2019). It also remains uncommitted as to whether mentalization contributes to the integrated pragmatic inferences that yield meaning. For the Post-Griceans, mentalization is an optional input, along the logical frame, for pragmatic inferences (Mazzarella and Noveck, 2021). The initial decoding is thus sufficient for identifying encyclopedic entries but insufficient to convey meaning. Nonetheless, since mentalization is optional, it does not seem necessary for meaning. Taken together, the question of whether there can be word learning without the attribution of mental states remains unanswered. Pragmatic theories argue for the decisive role of pragmatic inferences, either before (to identify the intention to communicate) or after words are decoded (to reason about their possible content), but they downplay or omit the role of mental states in the comprehension of meaning, despite building their arguments on intentions, communicative in nature.

5.1 Word learning: meaning as the merger form and content?

The idea that meaning emerges by establishing word-to-world mappings (Waxman and Lidz, 2007) via linking objects to sounds can be traced back at least to John Locke (Locke, 1975). In fact, it may be a unique feat of our species that a single system, rather than two separate ones, handles both conceptual representations and communication (Miller, 1990). The way these connections are established is still debated, however. The classical view of associations (Hume, 1978; Sloutsky et al., 2017), a form of statistical learning (Smith and Yu, 2008), has been seriously questioned on the grounds of social-pragmatic cognition (Tomasello, 2003; Bohn and Frank, 2019) and by placing intentions at the center stage (Macnamara, 1972; Bloom, 2000).

One outstanding challenge in explaining word learning is the question of referentiality. Referentiality is the idea that words single out and point to things in the world. However, they do so not at the level of individuals—and based on associations—but at the level of kinds (Waxman and Gelman, 2009). This definition is a minimalist one because proper names pick out individuals, but it suffices for most words. According to a series of well-crafted studies, when objects are labeled consistently, with pseudowords rather than tones, 3-month-olds form categories based on sets of objects and generalize membership to previously unseen novel members (Perszyk and Waxman, 2018). On the other side of the same coin, words refer also in the sense that they pick out objects. It has been shown that 4-month-olds follow the gaze direction of an actor faster to locate an object if the actor utters a pseudoword beforehand—backward speech, no vocalization, or looking at the infant instead of the side of the screen where the object is to appear do not do the trick (Marno et al., 2015). These findings show that very young infants can link linguistic signals to conceptual categories and expect these signals to indicate objects.

The first word infants seem to grasp is their own name, at least by 5 months of age (Grossmann et al., 2010; Parise et al., 2010). They do not take long to have at least some understanding of at least some—food-related and body-part—words by 6 months of age (Bergelson and Swingley, 2012, 2015; Tincoff and Jusczyk, 2012). Even these first words are organized in a semantically structured manner (Bergelson and Aslin, 2017), although word frequency and cross-linguistic differences may play a role here (Kartushina and Mayor, 2019; Steil et al., 2021). These findings refuted the long-held idea that during the first year of life, infants primarily learn the phonology of their native language(s) and that word learning proper begins only around their first birthday (Bloom, 2000; Kuhl, 2011).

Word comprehension undergoes qualitative changes during the first year, nevertheless. An electrophysiological indicator of semantic processing, the so-called N400 event-related potential (ERP) (Kutas and Hillyard, 1983; Kutas and Federmeier, 2011), can be elicited in infants by mislabeling objects (Friedrich and Friederici, 2004; Parise and Csibra, 2012). It appears as early as 6 months of age but only during the encoding phase of novel object-label pairings; a day later, infants show only a so-called N200-N500 phonological familiarity effect (Friedrich and Friederici, 2011). These results reveal that word forms are processed and semantic memory structures are in place but function at a limited capacity in 6-month-olds. Even at 9 months of age, the semantic system requires some support to produce an N400, such as words being produced by the infants' caregiver instead of by an experimenter (Parise and Csibra, 2012) or infants being familiarized with word labels in the lab (Junge et al., 2012). Only the top third high word producers of 12-month-olds exhibit the N400, and it can be reliably evoked only in 14-month-olds (Friedrich and Friederici, 2005, 2008; Forgács et al., 2019). A turning point in word learning at 14 months of age is underscored by the dramatic increase in infants' performance in Bergelson and Swingley's (2012) data as well. Werker and colleagues also demonstrated that only 14-month-olds, but not 12-month-olds, can link objects with labels during habituation training (Werker et al., 1998). A boost in the acquisition of abstract words has also been reported in this age group (Bergelson and Swingley, 2013), as well as a more sophisticated understanding of common ground (Moll et al., 2008). These shifts occur right before the onset of the supposed vocabulary spurt, an intense, albeit debated, expansion of the mental lexicon (Bloom, 2000; McMurray, 2007).

To expand their vocabulary, kids are thought to employ a number of dedicated learning strategies—not simply general inductive mechanisms. They rely on word learning constraints (Markman, 1990), such as the whole-object assumption, the taxonomic assumption—from 18 months of age (Markman and Hutchinson, 1984)—and the mutual exclusivity assumption—from as early as 12 months of age (Pomiechowska et al., 2021). They also utilize semantic (Pinker, 1984) and syntactic bootstrapping mechanisms (Brown, 1958; Gleitman, 1990), whereby they infer the meaning of words based on the meaning of the surrounding words in the former and by their syntactic role in sentences in the latter case. Taken together, the semantic system, which is thought to store the meaning of words, seems to be operational from 6 months of age and fully functional by 14 months of age. Word learning is thought to be aided by social cognition, which is thought to be external to the semantic system and pertaining mostly to pragmatic interpretative mechanisms.

5.2 Mentalization in the interpretation of the meaning of language

Just as with the diverse use of the term "communicative intentions," there is a continuum among researchers who advocate for the role of mentalization in acquiring the meaning of words and those arguing against it. Those who believe that mentalization is crucial on the road toward linguistic meaning—beyond the recognition of communicative intentions—mostly aim to account for referent resolution (Bloom, 2000; Tomasello, 2008). Tomasello's (2008) line of reasoning practically seeks to resolve referential ambiguity: recursive mind reading is necessary for appreciating shared goals, which creates joint attention, giving rise to common ground, which in turn allows for identifying the content of pointing, pantomiming, or words. Even those who do not explicitly argue for mentalization in referent resolution rely on some form of attention-guiding mechanism (Baldwin, 1991). Ostensive cues play a very similar role when they serve to establish a cultural learning interaction and, thereby, a unique interpretative context by guiding attention toward objects (for nouns), actions (for verbs), or functions/properties (for adjectives) (Csibra and Gergely, 2009). However, attention tracking may not be a good substitute for mentalization, as it still requires an attribution (Elekes and Király, 2021). One important motivation for leaving out mentalization from language acquisition has been uncertainty about whether ToM is available before 4 years of age. Pragmatics may have seemed a safe place to introduce mentalization in language acquisition because it fitted well with an unspoken, linear developmental order and the sequential conceptions of online language processing inherited from the serial comprehension models of Grice and Chomsky.

5.3 Developmental psychology's debate: language for ToM or ToM for language?

When the question of ToM was first raised in cognitive science (Premack and Woodruff, 1978), it soon became a tool to explain autism spectrum disorder (ASD) (Baron-Cohen et al., 1985; Frith and Happé, 1994). Autism had previously been treated mainly as a language deficit, but the new argument was that ASD children are unable to learn to use language in a socially appropriate manner because of a lack of a well-functioning ToM module and concomitant reduction in social motivations that curtail the necessary linguistic input. In an interesting twist, this idea was reversed while researchers scrambled to explain the classic explicit ToM tasks such as the Sally-Anne or Maxi task (Wimmer and Perner, 1983). The argument shifted to the idea that it was language development that enabled ToM (de Villiers and de Villiers, 2000), although the possibility of bidirectional influences was also offered (de Villiers, 2007). Some proposed the necessity of semantic development: as conceptual enrichment unfolds hand-in-hand with word learning (Gopnik and Meltzoff, 1998), ToM becomes available through learning mental words such as "think" or "believe" (Olson, 1988; Brooks and Meltzoff, 2015). Others emphasized that the emergence of ToM depends on grammatical structures (de Villiers, 1998; de Villiers and Pyers, 2002; Hale and Tager-Flusberg, 2003). Paralleling the finding that the acquisition of mental words is aided by complement clauses ("thinking or believing *that*") (Papafragou et al., 2007), mental state attribution is made possible by learning the syntactic structure for embedding propositions into propositions, in a meta-representational format ("Maxi thinks that »the chocolate is in the cupboard«"). Again, others have argued for the role of pragmatics (Harris et al., 2005; Frank, 2018; Rubio-Fernandez, 2021). A meta-analysis found that language indeed exerts a considerable influence on ToM: syntax and semantics, alongside receptive vocabulary size, memory for complements, and general language ability, were all positively associated with it (Milligan et al., 2007).

This direction of thinking has taken for granted, however, that ToM becomes available only once kids are able to pass explicit ToM tasks around 4 years of age (Wimmer and Perner, 1983). In such paradigms—the Maxi, the Sally-Anne, or the Smarties task (Perner et al., 1987)—children are explicitly asked about the mental contents of social partners (e.g., "What does Sally think, where are her marbles?"). Perhaps it is no wonder that language competence and ToM abilities have consistently been found to be interrelated.

When it emerged that preverbal infants exhibited ToM abilities (Scott and Baillargeon, 2017) as early as 6–8 months of age (Kovács et al., 2010; Southgate and Vernetti, 2014; Kampis et al., 2015), the idea that various language abilities lay the foundations for ToM was seriously challenged. Explicit tasks may not be tapping into mentalization *per se* but could instead run into some communicational-pragmatic burden (Helming et al., 2014). The implicit ToM results have been swiftly questioned (Ruffman, 2014) on methodological grounds (Poulin-Dubois et al., 2018), or they were explained away, either entirely (Heyes, 2014), or by suggesting that infant mentalization is inferior to that of adults. It was suggested to be ape-like and not suitable for coordinating perspectives (Tomasello, 2018) or that it is perceptual and "low-level," restricted to some sort of object tracking and physical perspective-taking system (Apperly and Butterfill, 2009; Low et al., 2016).

Nevertheless, a growing body of findings is proving to be increasingly difficult to explain without assuming adult-like meta-representational ToM in infancy. Observations that false beliefs can be ascribed to social partners without knowing their actual mental content (Kovács et al., 2021; Kampis and Kovács, 2022) suggest that infants attribute structured belief files (Kovács, 2016). Moreover, the ToM of 14-month-olds is capable of handling semantic representations that are in the appropriate "high-level" representational format for beliefs proper (Forgács et al., 2019, 2020). Based on these findings, it is well possible that ToM contributes to or enables language development rather than the other way around.

5.4 The social N400: is semantic processing mentalistic?

Recent findings on the so-called social N400 effect have profoundly challenged the received knowledge on the neurocognitive organization of language processing and its relation to social cognition (Rueschemeyer et al., 2015; Westley et al., 2017; Jouravlev et al., 2019; Hinchcliffe et al., 2020). When participants were required to track the comprehension of a confederate while reading semantically incongruous sentences together ("The boy had *gills*"), they exhibited an N400. Surprisingly, this occurred even when they heard context sentences beforehand ("In the boy's dream, he could breathe under water"), which should have attenuated the N400 by providing interpretative context. The intriguing finding is that a social effect, which should have engaged pragmatic mechanisms, elicited a semantic response.

In a paradigm designed to directly manipulate the belief state of a communicative partner during language comprehension, even 14-month-olds produced a social N400 in response to the miscomprehension of a social partner (Forgács et al., 2019). In a puppet theater experiment, infants were presented with familiar objects that were always correctly labeled from their perspective but sometimes incorrectly labeled from the perspective of an observer. The observer, seated on the other side of the stage, had visual access to objects only when an occluder was lowered. First, an object was placed in front of infants (e.g., a cup), which was revealed to the observer as well; however, when the occluder moved back up, the observer turned away, and the first object was replaced by a second one (e.g., a car), unbeknownst to the observer. When the observer turned back, the second object was labeled ("car"), which was congruent for infants but incongruent with the false belief of the observer. Despite experiencing no semantic processing demands, infants produced an N400 (Forgács et al., 2019, 2020). Thus, the ERP indicator of language comprehension responded to a mentalistic manipulation, not simply a social one. These findings are relevant for ToM research because they demonstrate that false beliefs can be attributed as semantic content, which is compatible with a propositional meta-representational format. Conversely, ToM may be at full capacity already in infancy (Leslie, 1994). The findings are also remarkable from the perspective of experimental pragmatics because they show that not simply social cognition but specifically mentalization can impact language comprehension, not only at the level of pragmatic-inferential mechanisms but also at the level of semantic processing.

The social N400 appears to have two constituents: the *false belief N400* and the *social presence N400* (Forgács et al., 2022b). When presented with congruent and incongruent object-labeling events, adults showed an enhanced N400 response not only to incongruity but also to the mere presence of another person, in contrast to when they were alone. The typical N400 seems to be best explained as a semantic memory retrieval effort (Kutas and Federmeier, 2011; Brouwer et al., 2012; Urbach et al., 2020), which is evoked at all times but reduced when semantic predictions are met. Thus, the social presence effect can be understood as a lesser reduction of the N400 when someone is simply present. This may be due to a broader range of semantic elements remaining activated, which is likely to enhance potentially ensuing social interactions. The *false belief N400* can be elicited in adults as well, over and above the *social presence N400*. In the *false belief N400* paradigm, an observer is always present. However, an additional N400 effect is evoked only if participants are explicitly instructed to follow the comprehension of the other person (Forgács et al., 2022b)—just as in the information asymmetry social N400 experiments (Jouravlev et al., 2019). In sum, semantic processing seems to involve two mentalistic components: a spontaneous one, the *social presence N400*, and a strategic one apparent only following instructions, the *false belief N400*.

The *social presence N400* is evident already at 14 months of age, right at the developmental onset of the N400. Nonetheless, the effect appears only in response to incongruent labels, not to congruent ones (Forgács et al., submitted). It seems that infants ration their limited cognitive capacities to engage in semantic mentalization only when incongruent labels potentially incur divergent perspectives and false beliefs rather than when congruent labels require the attribution of true beliefs. In such cases, it may be sufficient to assume a shared, normative belief (Király et al., 2018). It may be argued that no attribution of beliefs is necessary for the social presence effect and that attributing perception may suffice. This may be true, but it is based on the assumption that the N400 is an indicator of perceptual processing. While it has been argued that the language system is fundamentally a reflex-like perceptual system (Fodor, 1983), the more broadly accepted view is that semantic mechanisms pertain to the conceptual system in some way. Additionally, the attribution of perception could also be viewed as a form of mentalization, as it involves ascribing an experience as a mental state.

The mentalistic social N400 is a riddle for pragmatic theories. For Neo-Griceans, social cognition, let alone mentalization, should not influence semantic mechanisms—only pragmatic inferences. For Post-Griceans, since inferential mechanisms may already be involved in developing explicatures, the impact of social cognition on semantic processing may not be unexpected. However, mentalization should be an input to the pragmatic module (together with the logical frame) and should not influence the initial lexical retrieval. The thought-provoking aspect of the mentalistic N400 is that none of these experiments were supposed to elicit an N400, as they did not pose any semantic processing demands *per se*. Instead, they well could have evoked ERPs associated either with ToM, including parietal or frontal responses (Liu et al., 2009; McCleery et al., 2011), or with pragmatics and contextual processing, such as the P600 (e.g., Van Berkum, 2009). Mentalization apparently impacted language comprehension not on a pragmatic but on a semantic level, which was not predicted by any pragmatic theories.

At a minimum, these results suggest that the ToM network may coordinate very closely with the language network (Paunov et al., 2019). The ToM network is a bilateral system, perhaps slightly more right-lateralized, with centers at the temporoparietal junction (TPJ) and the middle prefrontal cortex (mPFC) (Frith and Frith, 2003; Saxe and Kanwisher, 2003). It is part of a broader network of social cognition (Schurz et al., 2020). The language network is a more left-lateralized system of temporal and frontal regions (Binder et al., 1997). It has been argued that these two networks work independently (Shain et al., 2023), despite some apparent overlaps, and have stronger connections within themselves than between each other (Paunov et al., 2019). However, there are concerns with explaining the social N400 based on an interaction between the two networks. It is a lexical input that should trigger the ToM network (instead of the language processor), which in turn should activate the semantic system soon enough to produce an N400, yet not for semantic retrieval but to represent the mental state of a social partner. Thus, not only was an N400 not expected in the social N400 experiments (in the absence of semantic processing demands), and no other ERPs were observed (to indicate the activation of the ToM network), but specifically an ERP associated with the semantic system responded to the language comprehension and miscomprehension of a social partner.

It is true that belief attribution was accompanied by frontal effects in infants' *false belief N400* experiments (Forgács et al., 2019, 2020). However, these effects were inconsistent between French and Hungarian infants, and frontal regions may be engaged during false belief processing for a variety of reasons beyond belief computations, from inhibitory control through response selection to resolving conflicting representations (Southgate, 2020). It is also true that the infant social presence study (Forgács et al., submitted) involved no false beliefs, only the tracking of another person's experience of a semantic incongruity, which nevertheless still seems to qualify as at least some form of belief attribution. The overall pattern of results suggests that the semantic system is engaged in processing ToM in the mentalistic N400 experiments. Such an interpretation does not preclude the possibility that the ToM and language networks are separate systems that work closely together (Paunov et al., 2019; Shain et al., 2023; Fedorenko et al., 2024). The semantic system could work mentalistically without subserving other ToM functions.

6 Meaning as mentalization

The main claim of this paper is that the semantic system may function in a mentalistic manner by storing, manipulating, and retrieving content based on belief attributions. The sensitivity of the N400 to mentalistic manipulations is not a curious detail but a functional characteristic of the semantic system. The idea is that the information transmitted via linguistic forms—the phonological-lexical input—triggers an unpacking mechanism of the belief the speaker intends to express based on semantic activations. Thus, interpreting utterances does not begin with merely looking up content in the database of the mental lexicon (as per the code model and the Neo-Griceans) or by generating a raw logical form that serves as an entry point for the mental encyclopedia (as per RT and the Post-Griceans). Instead, semantic content is the result of a memory retrieval of a likely intended sense based on the lexical evidence and the belief ascribed to the communicative partner as a probable piece of information (Figure 2).

Mentalization may or may not play a role in setting up communicative interactions by identifying communicative intentions (based on ostensive signals and/or engaging in joint attention) or in deriving pragmatic inferences (of social cognition and/or logical reasoning). However, it may be crucial exactly in between the two, when meaning is arrived at—where intentions may matter the most. The content linked to linguistic forms during language acquisition, as well as during everyday language comprehension, may be viewed as an attribution within the constraints of both the code-like features of language and the social cognitive dimensions of human communication. The structural properties of language and word forms may help limit the scope of the mentalistic attributions of intended content, while pragmatic inferences may help further specify and adjust it, if necessary. In contrast to Vygotsky's and Bruner's studies, where scaffolding by the social world fills the minds of children from the outside (Wood et al., 1976), the present approach proposes the reverse direction. The social aspect may work from the inside out in the form of social cognition, from the minds of children toward the minds of social partners to acquire meaning by attributing beliefs. Thus, semantic content may not be identified in the external world, as referents discovered during social interactions, but in the internal worlds of communicators, as hearers' best guesses for belief ascription.

The recognition of communicative intentions, in the sense of *intention to communicate*, maybe the entry point for ascribing beliefs to social partners. The mentalistic attribution of potential content may be the richer the more complex the code is, such that pointing is superseded by pantomiming, which is superseded by language proper, be it whistle, sign, or verbal language use. The recognition of communicative intentions may not simply aid reference resolution via attention guidance, after which relevant information can be transmitted regarding the world (of objects, actions, or properties). Instead, it may initiate the attribution of what the other person may have in mind (a particular object, action, or property). By the time linguistic information transmission commences, the referent of a spoken word may not be identified as a physical object but as the mental representation of the object attributed to the communicative partner.

It may be argued that no mentalization is required once joint attention or ostensive cues have done their job because infants may simply take the attended object to be the referent of the word to establish word-to-world mappings. They need no representation of the mental content of the communicative partner by the time information is transmitted. However, the content of the word would still be enormously difficult to determine based solely on the tracking of attention, goals, and physical objects, as highlighted by the "gavagai problem" (Quine, 1960). Markman's constraints may provide some aid on a pragmatic level, but they do not seem to solve the matter comprehensively, especially at the very early stages of world learning. The problem largely evaporates if one assumes that the referents we interact with, communicate, and talk about are not simply in the physical world but inside the minds of speakers, in contrast to traditional views on language acquisition.

The transmission of the signal may be exploited to narrow the range of possible mentalistic attributions, which specify communicative intentions, now in the sense of *meaning as intended*.

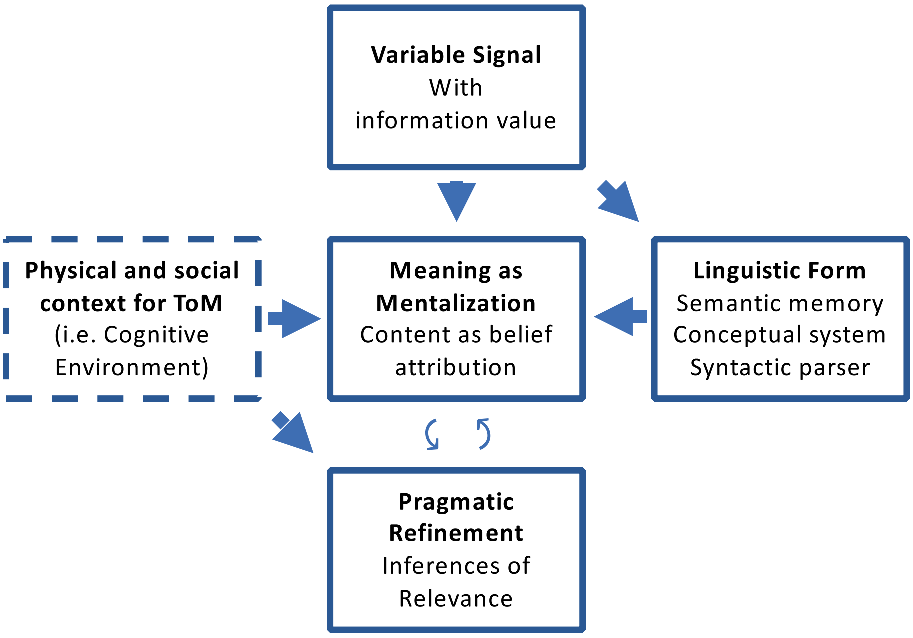


FIGURE 2
A novel model for establishing linguistic meaning through attributing mental content as intended meaning to social partners, based on the lexical input along with meta-communicative and other signals, the physical and social context, and the cognitive environment. Semantic content as a belief ascription may be updated based on logical and/or social inferential mechanisms during pragmatic enrichment. Mentalization may be optional for setting up communicative interactions and deriving pragmatic inferences. However, it seems indispensable for any theory involving (communicative) intentions that aims to explain meaning as the content conveyed in communication.

The informative intention could thus be viewed as the particular attributed belief. Such a mechanism could account for both the social acquisition of linguistic forms (from words to grammar) and the interpretational wiggle room language always seems to leave. Pragmatic inferences may further narrow the remaining ambiguity but may not necessarily involve mentalization. Social-contextual adjustments may be made optionally based on information available in the cognitive environment and/or the common ground, and sometimes updating the initial content attribution may be unavoidable, but not always.

The semantic system would still accumulate, store, and utilize statistical, taxonomic, and other structural regularities of the incoming signal to provide a better springboard for its main function of attributing meaning. As noted by Bruner (1990), linguistic meaning does not seem to be looked up from a data table but is rather reinvented from incoming raw materials in a creative process of "meaning making." Viewing the semantic system as a mentalistic system could bridge the gaps between word, sentential, and contextual meaning by treating them as the same kind of belief attribution by the language system, albeit with gradually increasing complexity. We may rely more on the code in particular routine situations, from formulaic language to other conventions, as proposed and perhaps overgeneralized by the Speech Act theorists (Austin, 1962; Searle, 1979) or Millikan's (2005) direct perception model. However, even such interpretative best guesses could be mentalistic in nature and not qualitatively different from semantic ToM efforts when communication does not unfold as predicted.

This view could account for the mentalistic social N400 findings without appealing to a peculiar interaction between the language and the ToM networks. The idea is that the language system was not recruited by the ToM network but worked independently, carrying out mentalistic functions. The current proposal argues that these experiments were not revealing exceptions but rather the modus operandi of the semantic system. The findings of Fedorenko and colleagues that classic ToM and language tasks do not engage the other network do not refute the idea that the language network may function mentalistically.

The finding that adults produce a false belief social N400 only when explicitly requested to do so (Forgács et al., 2022b), while infants show it spontaneously (Forgács et al., 2019), suggests that language learners may rely more heavily on belief attributions to identify intended meanings than adults. With accumulating conversational routine, adults may be less prone to invest additional neural resources in strategic mentalization beyond spontaneous mentalization. During language acquisition, the semantic system may be optimized toward a generic model of an idealized speaker. With the gradual expansion of lexical databases, linguistic conventions, and conversational routine, semantic mentalization may increasingly resort to normative attributions to a default speaker. By adulthood, only when interactions and conversations take unexpected turns may personalized mentalization retake the lead.

A possible objection to the semantic system always functioning mentalistically is that it would imply no difference between social and non-social language input. The present framework proposes that the amplitude of the N400, being a graded ERP, reflects varying neural processing demands not only in response to lexical retrieval but also to mentalization. The various technological innovations that allow linguistic input to be provided without a speaker actually being present in person (from writing systems to audio recordings) may hack into the proper cognitive domain of the semantic system.

Classical psycholinguistic experiments testing individuals alone may have tapped into a special case of language processing based on generic semantic attributions to a default speaker. When linguistic stimuli are encountered in the physical presence of a social partner, additional semantic attributions are spontaneously generated for the specific individual beyond the generic model. The system's functioning is further geared up when the other person experiences a false belief, and the conversation may be derailed. Attributions of meaning may be simpler if the interlocutors are closer to each other's idealized default speaker model. In a close language community, each individual's idealized speaker model is based on a highly similar body of language input, toward which the code structures are statistically optimized. The statistical structures of the semantic memory system that psycholinguistic experiments have described in great detail may reflect these statistical features, but the proper function of the system may still be determining what was meant by communicative partners.

It may also be argued that the social N400 is a result of social facilitation. While such an explanation may be possible for adults' social presence effect, it does not work for infants' because it appeared only in a semantically incongruent condition, which suggests its strategic employment. This explanation also cannot account for the *false belief N400* effect because it appeared in adults only after explicit instructions, indicating again a strategic element. Of course, future studies are necessary to further scrutinize and gather additional evidence in support of the theory.

7 Conclusion

Can meaning be understood as unintended? Is meaning an abstraction in the world or a psychological phenomenon in the mind? It seems paradoxical to argue that intentions, especially when they are communicative, are not attributed to social partners. One may reverse the question: how much of establishing intended meaning is *not* mentalistic? The present study proposes that, instead of relying on decoding and pragmatic mechanisms, meaning is directly interpreted as it is intended. Meaning may be the information mentalistically attributed as a belief to a communicative partner.

Author contributions

BF: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

I would like to thank Ildikó Király, Judit Gervain, Csaba Pléh, and György Gergely for the insightful discussions on the road toward developing the ideas presented in this study.

Conflict of interest

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

Publisher's note

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

References

Abboub, N., Nazzi, T., and Gervain, J. (2016). Prosodic grouping at birth. *Brain Lang.* 162, 46–59. doi: 10.1016/j.bandl.2016.08.002

Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923

Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Dev.* 62, 875–890. doi: 10.1111/j.1467-8624.1991.tb01577.x

Baldwin, D. A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Dev. Psychol.* 29, 832–843. doi: 10.1037/0012-1649.29.5.832

Baldwin, D. A., and Markman, E. M. (1989). Establishing word-object relations: a first step. *Child Dev.* 60, 381–398. doi: 10.2307/1130984

Bambini, V., and Bara, B. (2010). What is neuropragmatics? A brief note. Quaderni Del Laboratorio Di, 9. Available at: http://linguistica.sns.it/QLL/QLL10/Bambini-Bara.pdf

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Bates, E. (1976). *Language and context: The Acquisition of Pragmatics*. Cambridge, MA: Academic Press.

Begus, K., Gliga, T., and Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proc. Natl. Acad. Sci. USA* 113, 12397–12402. doi: 10.1073/pnas.1603261113

Behne, T., Carpenter, M., and Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Dev. Sci.* 8, 492–499. doi: 10.1111/j.1467-7687.2005.00440.x

Bennett, J. (1978). Some remarks about concepts. *Behav. Brain Sci.* 1, 557–560. doi: 10.1017/S0140525X00076573

Bergelson, E., and Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proc. Natl. Acad. Sci.* 114, 12916–12921. doi: 10.1073/pnas.1712966114

Bergelson, E., and Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proc. Natl. Acad. Sci.* 109, 3253–3258. doi: 10.1073/pnas.1113380109

Bergelson, E., and Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition* 127, 391–397. doi: 10.1016/j.cognition.2013.02.011

Bergelson, E., and Swingley, D. (2015). Early word comprehension in infants: replication and extension. *Lang. Learn. Dev.* 11, 369–380. doi: 10.1080/15475441.2014.979387

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., et al. (2018). Promoting replicability in developmental research through Meta-analyses: insights from language acquisition research. *Child Dev.* 89, 1996–2009. doi: 10.1111/cdev.13079

Berwick, R. C., and Chomsky, N. (2011). The biolinguistic program: the current state of its development. In ScullioA. M. Di and C. Boeckx (Eds.), *The biolinguistic Enterprise* (pp. 19–41). Oxford: Oxford University Press.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362. doi: 10.1523/jneurosci.17-01-00353.1997

Black, M. (1962). *Models and metaphors: Studies in language and Philosopy*. Ithaca, NY: Cornell University Press.

Blackburn, P. L. (2007). *The code model of communication*. A powerful metaphor in linguistic metatheory.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: The MIT Press.

Bohn, M., Call, J., and Tomasello, M. (2019). Natural reference: a phylo-and ontogenetic perspective on the comprehension of iconic gestures and vocalizations. *Dev. Sci.* 22, e12757–e12712. doi: 10.1111/desc.12757

Bohn, M., and Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Ann. Rev. Dev. Psychol.* 1, 223–249. doi: 10.1146/annurev-devpsych-121318-085037

Bohn, M., Tessler, M. H., Merrick, M., and Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nat. Hum. Behav.* 5, 1046–1054. doi: 10.1038/s41562-021-01145-1

Bosco, F. M., Tirassa, M., and Gabbatore, I. (2018). Why pragmatics and theory of mind do not (completely) overlap. *Front. Psychol.* 9, 1–7. doi: 10.3389/fpsyg.2018.01453

Brentano, F. (2009). *Psychology from an empirical standpoint*. New York: Taylor & Francis Original work published in 1874.

Brooks, R., and Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *J. Child Lang.* 35, 207–220. doi: 10.1017/S030500090700829X

Brooks, R., and Meltzoff, A. N. (2013). Gaze following: a mechanism for building social connections between infants and adults. *Mechan. Soc. Connect.* 2013, 167–183. doi: 10.1037/14250-010

Brooks, R., and Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: a longitudinal study connecting gaze following, language, and explicit theory of mind. *J. Exp. Child Psychol.* 130, 67–78. doi: 10.1016/j.jecp.2014.09.010

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

Brown, R. (1958). How shall a thing be called? *Psychol. Rev.* 65, 14–21. doi: 10.1037/h0041727

Bruner, J. S. (1983). *Child's talk: Learning to use language*. Oxford: Oxford University Press.

Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., and Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63:i. doi: 10.2307/1166214

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (2015). *What kind of creatures are we?* New York, NY: Columbia University Press.

Chomsky, N. (2017). The language capacity: architecture and evolution. *Psychon. Bull. Rev.* 24, 200–203. doi: 10.3758/s13423-016-1078-6

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, E. V., and Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Lang. Linguis. Compass* 4, 445–457. doi: 10.1111/j.1749-818X.2010.00214.x

Cooper, R. P., and Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev.* 61, 1584–1595. doi: 10.2307/1130766

Cristia, A. (2023). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Dev. Sci.* 26, e13265–e13215. doi: 10.1111/desc.13265

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind Lang.* 25, 141–168. doi: 10.1111/j.1468-0017.2009.01384.x

Csibra, G., and Gergely, G. (2007). 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol.* 124, 60–78. doi: 10.1016/j.actpsy.2006.09.007

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005

de Saussure, F. (1966). *Course in general linguistics*. New York, NY: McGraw-Hill Book Company Original work published in 1916.

de Villiers, J. G. (1998). On acquiring the structural representations for false complements. University of Massachusetts Occasional Papers in Linguistics, 24, 125–136.

de Villiers, J. G. (2007). The interface of language and theory of mind. *Lingua* 117, 1858–1878. doi: 10.1016/j.lingua.2006.11.006

de Villiers, J. G., and de Villiers, P. (2000). "Linguistic determinism and the understanding of false beliefs" in *Children's reasoning and the mind*. eds. P. Mitchell and K. Riggs (London: Psychology Press), 890.

de Villiers, J. G., and Pyers, J. E. (2002). Complements to cognition: a longitudinal study of the relationship between complex synthax and false-belief-understanding. *Cogn. Dev.* 17, 1037–1060. doi: 10.1016/S0885-2014(02)00073-4

Dennett, D. C. (1978). Beliefs about beliefs [P&W, SR&B]. *Behav. Brain Sci.* 1, 568–570. doi: 10.1017/S0140525X00076664

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Egyed, K., Király, I., and Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychol. Sci.* 24, 1348–1353. doi: 10.1177/0956797612471952

Elekes, F., and Király, I. (2021). Attention in naïve psychology. *Cognition* 206:104480. doi: 10.1016/j.cognition.2020.104480

Fedorenko, E., Ivanova, A. A., and Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* 25, 289–312. doi: 10.1038/s41583-024-00802-4

Fernald, A. (1992). "Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective" in *The adapted mind* (New York, NY: Oxford University Press), 391–428.

Fitch, W. T., Huber, L., and Bugnyar, T. (2010). Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron* 65, 795–814. doi: 10.1016/j.neuron.2010.03.011

Fló, A., Benjamin, L., Palu, M., and Lambertz, G. D. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Sci. Rep.* 12:4391. doi: 10.1038/s41598-022-08411-w

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., and Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Dev. Sci.* 22:e12802. doi: 10.1111/desc.12802

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT press.

Forgács, B., Gervain, J., Parise, E., Berkes, N., Szigeti, A., Bumin, F. B., et al. (submitted). The intriguingly social N400 of preverbal infants.

Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., et al. (2020). Electrophysiological investigation of infants' understanding of understanding. *Dev. Cogn. Neurosci.* 43:100783. doi: 10.1016/j.dcn.2020.100783

Forgács, B., Gervain, J., Parise, E., Gergely, G., Elek, L. P., Üllei-Kovács, Z., et al. (2022a). Semantic systems are mentalistically activated for and by social partners. *Sci. Rep.* 12:4866. doi: 10.1038/s41598-022-08306-w

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., and Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Dev. Sci.* 22:e12751. doi: 10.1111/desc.12751

Forgács, B., Tauzin, T., Gergely, G., and Gervain, J. (2022b). The newborn brain is sensitive to the communicative function of language. *Sci. Rep.* 12:1220. doi: 10.1038/s41598-022-05122-0

Frank, C. K. (2018). Reviving pragmatic theory of theory of mind. *AIMS Neurosci.* 5, 116–131. doi: 10.3934/Neuroscience.2018.2.116

Frege, G. (1948). Sense and reference. *Philos. Rev.* 57, 209–230. Original work published in 1892. doi: 10.2307/2181485

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* 16, 262–268. doi: 10.1016/j.tics.2012.04.001

Friedrich, M., and Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *J. Cogn. Neurosci.* 16, 1465–1477. doi: 10.1162/0898929042304705

Friedrich, M., and Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport* 16, 653–656. doi: 10.1097/00001756-200504250-00028

Friedrich, M., and Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport* 19, 1757–1761. doi: 10.1097/WNR.0b013e328318f014

Friedrich, M., and Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding–weak retention. *J. Cogn. Neurosci.* 23, 3228–3240. doi: 10.1162/jocn_a_00002

Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. Royal Soc. B: Biol. Sci.* 358, 459–473. doi: 10.1098/rstb.2002.1218

Frith, U., and Happé, F. (1994). Language and communication in autistic disorders. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 346, 97–104. doi: 10.1098/rstb.1994.0133

Futó, J., Téglás, E., Csibra, G., and Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition* 117, 1–8. doi: 10.1016/j.cognition.2010.06.003

Gergely, G. (2010). "Kinds of agents: the origins of understanding instrumental and communicative agency" in *The Wiley-Blackwell handbook of childhood cognitive development* (Hoboken, NJ: Wiley), 76–105.

Gergely, G., Bekkering, H., and Király, I. (2002). Rational imitation in preverbal infants. *Nature* 415:755. doi: 10.1038/415755a

Gergely, G., and Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends Cogn. Sci.* 7, 287–292. doi: 10.1016/S1364-6613(03)00128-1

Gergely, G., and Csibra, G. (2005). The social construction of the cultural mind: imitative learning as a mechanism of human pedagogy. *Interact. Stud.* 6, 463–481. doi: 10.1075/is.6.3.10ger

Gergely, G., and Csibra, G. (2013). "Natural Pedagogy" in *Navigating the social world: What infants, Chidren, and other species can teach us*. eds. M. R. Banaji and S. A. Gelman (Oxford: Oxford University Press), 127–132.

Gervain, J., Berent, I., and Werker, J. F. (2012). Binding at birth: the newborn brain detects identity relations and sequential position in speech. *J. Cogn. Neurosci.* 24, 564–574. doi: 10.1162/jocn_a_00157

Gervain, J., Macagno, F., Cogoi, S., Pena, M., and Mehler, J. (2008). The neonate brain detects speech structure. *Proc. Natl. Acad. Sci.* 105, 14222–14227. doi: 10.1073/pnas.0806530105

Gervain, J., and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu. Rev. Psychol.* 61, 191–218. doi: 10.1146/annurev.psych.093008.100408

Gleitman, L. (1990). "The structural sources of verbal meaning" in *Language Acquisition*, vol. 1, 3–55.

Goodman, N. D., and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci.* 20, 818–829. doi: 10.1016/j.tics.2016.08.005

Gopnik, A., and Meltzoff, A. N. (1998). "Words, thoughts, and theories" in *Words, thoughts, and theories* (Cambridge, MA: The MIT Press).

Goswami, U. (2022). Language acquisition and speech rhythm patterns: an auditory neuroscience perspective. *R. Soc. Open Sci.* 9:211855. doi: 10.1098/rsos.211855

Grice, H. P. (1957). Meaning. *The Philosophical Review* 66:377. doi: 10.2307/2182440

Grice, H. P. (1975). "Logic and conversation" in *Speech acts*. eds. P. Cole and J. L. Morgan (Cambridge, MA: Academic Press), 41–58.

Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.

Grossmann, T., Johnson, M. H., Farroni, T., and Csibra, G. (2007). Social perception in the infant brain: gamma oscillatory activity in response to eye gaze. *Soc. Cogn. Affect. Neurosci.* 2, 284–291. doi: 10.1093/scan/nsm025

Grossmann, T., Parise, E., and Friederici, A. D. (2010). The detection of communicative signals directed at the self ininfant prefrontal cortex. *Front. Hum. Neurosci.* 4, 1–5. doi: 10.3389/fnhum.2010.00201

Hale, C. M., and Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Dev. Sci.* 6, 346–359. doi: 10.1111/1467-7687.00289

Harris, P. L., De Rosnay, M., and Pons, F. (2005). Language and children's understanding of mental states. *Curr. Dir. Psychol. Sci.* 14, 69–73. doi: 10.1111/j.0963-7214.2005.00337.x

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The Faculty of Language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Helming, K. A., Strickland, B., and Jacob, P. (2014). Making sense of early false-belief understanding. *Trends Cogn. Sci.* 18, 167–170. doi: 10.1016/j.tics.2014.01.005

Hernik, M., and Csibra, G. (2015). Infants learn enduring functions of novel tools from action demonstrations. *J. Exp. Child Psychol.* 130, 176–192. doi: 10.1016/j.jecp.2014.10.004

Heyes, C. (2014). Submentalizing: I am not really Reading your mind. *Perspect. Psychol. Sci.* 9, 131–143. doi: 10.1177/1745691613518076

Hinchcliffe, C., Jiménez-Ortega, L., Muñoz, F., Hernández-Gutiérrez, D., Casado, P., Sánchez-García, J., et al. (2020). Language comprehension in the social brain: electrophysiological brain signals of social presence effects during syntactic and semantic sentence processing. *Cortex* 130, 413–425. doi: 10.1016/j.cortex.2020.03.029

Horn, L. R. (2006). The border wars: a neo-Gricean perspective. In HeusingerK. von and K. Turner (Eds.), *Where semantics meets pragmatics* (pp. 21–48). Leiden: Brill.

Huang, Y. T., and Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. *Dev. Psychol.* 45, 1723–1739. doi: 10.1037/a0016704

Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University Press Original work published in 1739.

Jacob, P. (2023). "Intentionality" in *The Stanford encyclopedia of philosophy (spring 202)*. eds. E. N. Zalta and U. Nodelman (Stanford: Stanford University).

Jouravlev, O., Schwartz, R., Ayyash, D., Mineroff, Z., Gibson, E., and Fedorenko, E. (2019). Tracking Colisteners' knowledge states during language comprehension. *Psychol. Sci.* 30, 3–19. doi: 10.1177/0956797618807674

Junge, C., Cutler, A., and Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia* 50, 3702–3712. doi: 10.1016/j.neuropsychologia.2012.10.012

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., and Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci. Rep.* 8, 13745–13748. doi: 10.1038/s41598-018-32150-6

Kampis, D., and Kovács, Á. M. (2022). Seeing the world from others' perspective: 14-month-olds show Altercentric modulation effects by others' beliefs. *Open Mind* 5, 189–207. doi: 10.1162/opmi_a_00050

Kampis, D., Parise, E., Csibra, G., and Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proc. R. Soc. B Biol. Sci.* 282:20151683. doi: 10.1098/rspb.2015.1683

Kartushina, N., and Mayor, J. (2019). Word knowledge in six-to nine-month-old Norwegian infants? Not without additional frequency cues. *R. Soc. Open Sci.* 6:180711. doi: 10.1098/rsos.180711

Khu, M., Chambers, C., and Graham, S. A. (2018). When You're happy and I know it: four-year-olds' emotional perspective taking during online language comprehension. *Child Dev.* 89, 2264–2281. doi: 10.1111/cdev.12855

Király, I., Oláh, K., Csibra, G., and Kovács, Á. M. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proc. Natl. Acad. Sci.* 115, 11477–11482. doi: 10.1073/pnas.1803505115

Kovács, Á. M. (2016). Belief files in theory of mind reasoning. *Rev. Philos. Psychol.* 7, 509–527. doi: 10.1007/s13164-015-0236-5

Kovács, Á. M., Téglás, E., and Csibra, G. (2021). Can infants adopt underspecified contents into attributed beliefs? Representational prerequisites of theory of mind. *Cognition* 213:104640. doi: 10.1016/j.cognition.2021.104640

Kovács, Á. M., Teglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792

Kronmüller, E., Moriseau, T., and Noveck, I. A. (2014). Show me the pragmatic contribution: a developmental investigation of contrastive inference. *J. Child Lang.* 41, 985–1014. doi: 10.1017/S0305000913000263

Kuhl, P. K. (2011). Who's talking? *Science* 333, 529–530. doi: 10.1126/science.1210277

Kujala, T., Partanen, E., Virtala, P., and Winkler, I. (2023). Prerequisites of language acquisition in the newborn brain. *Trends Neurosci.* 46, 726–737. doi: 10.1016/j.tins.2023.05.011

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cogn.* 11, 539–550. doi: 10.3758/BF03196991

Leslie, A. M. (1987). Pretense and representation: the origins of "theory of mind.". *Psychol. Rev.* 94, 412–426. doi: 10.1037/0033-295X.94.4.412

Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition* 50, 211–238. doi: 10.1016/0010-0277(94)90029-9

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., et al. (2016). A quantitative synthesis of early language acquisition using Meta-analysis. *PsyArXiv*, 1–24. doi: 10.31234/osf.io/htsjm

Liakakis, G., Nickel, J., and Seitz, R. J. (2011). Diversity of the inferior frontal gyrus-a meta-analysis of neuroimaging studies. *Behav. Brain Res.* 225, 341–347. doi: 10.1016/j.bbr.2011.06.022

Liu, D., Sabbagh, M. A., Gehring, W. J., and Wellman, H. M. (2009). Neural correlates of children's theory of mind development. *Child Dev.* 80, 318–326. doi: 10.1111/j.1467-8624.2009.01262.x

Lloyd-Fox, S., Széplaki-Köllőd, B., Yin, J., and Csibra, G. (2015). Are you talking to me? Neural activations in 6-month-old infants in response to being addressed during natural interactions. *Cortex* 70, 35–48. doi: 10.1016/j.cortex.2015.02.005

Locke, J. (1975). *An essay concerning human understanding*. Oxford: Oxford University Press Original work published in 1689.

Low, J., Apperly, I. A., Butterfill, S. A., and Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: a primer on the two-systems account. *Child Dev. Perspect.* 10, 184–189. doi: 10.1111/cdep.12183

Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychol. Rev.* 79, 1–13. doi: 10.1037/h0031901

Markman, E. M. (1990). Constraints children place on word meanings. *Cogn. Sci.* 14, 57–77. doi: 10.1016/0364-0213(90)90026-S

Markman, E. M., and Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations. *Cogn. Psychol.* 16, 1–27. doi: 10.1016/0010-0285(84)90002-1

Marno, H., Farroni, T., Vidal Dos Santos, Y., Ekramnia, M., Nespor, M., and Mehler, J. (2015). Can you see what i am talking about? Human speech triggers referential expectation in four-month-old infants. *Sci. Rep.* 5, 1–10. doi: 10.1038/srep13594

Martin, A., Onishi, K. H., and Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12months. *Cognition* 123, 50–60. doi: 10.1016/j.cognition.2011.12.003

Mazzarella, D., and Noveck, I. (2021). Pragmatics and mind reading: the puzzle of autism (response to kissine). *Language* 97, e198–e210. doi: 10.1353/LAN.2021.0037

McCleery, J. P., Surtees, A. D. R., Graham, K. A., Richards, J. E., and Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *J. Neurosci.* 31, 12849–12854. doi: 10.1523/JNEUROSCI.1392-11.2011

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science* 317:631. doi: 10.1126/science.1144073

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

Milligan, K., Astington, J. W., and Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* 78, 622–646. doi: 10.1111/j.1467-8624.2007.01018.x

Millikan, R. G. (2005). *Language: A biological model*. Oxford: Oxford University Press.

Moll, H., Richter, N., Carpenter, M., and Tomasello, M. (2008). Fourteen-month-olds know what "we" have shared in a special way. *Infancy* 13, 90–101. doi: 10.1080/15250000701779402

Musso, M., Moro, A., Glauchel, V., Rijntjes, M., Reichenbach, J., Büchel, C., et al. (2003). Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781. doi: 10.1038/nn1077

Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x

Nallet, C., Berent, I., Werker, J. F., and Gervain, J. (2023). The neonate brain's sensitivity to repetition-based structure: specific to speech? *Dev. Sci.*, 1–9. doi: 10.1111/desc.13408

Nelson, K. (1973). Structure and strategy in learning to talk. *Monogr. Soc. Res. Child Dev.* 38:1. doi: 10.2307/1165788

Nelson, K. (2009). Wittgenstein and contemporary theories of word learning. *New Ideas Psychol.* 27, 275–287. doi: 10.1016/j.newideapsych.2008.04.003

Nilsen, E. S., and Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cogn. Psychol.* 58, 220–249. doi: 10.1016/j.cogpsych.2008.07.002

Noveck, I. A. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge, MA: Cambridge University Press.

Noveck, I. A., and Chevaux, F. (2002). The pragmatic development of 'and.' Proceedings of the 26th Annual Boston University Conference on Language Development, 453–463.

Noveck, I. A., and Reboul, A. (2008). Experimental pragmatics: a Gricean turn in the study of language. *Trends Cogn. Sci.* 12, 425–431. doi: 10.1016/j.tics.2008.07.009

Noveck, I. A., and Sperber, D. (Eds.) (2004). *Experimental Pragmatics*. London: Palgrave Macmillan UK.

Olson, D. R. (1988). "On the origins of beliefs and other intentional states in children" in *Developing theories of mind*. eds. J. W. Astington, P. L. Harris and D. R. Olson (Cambridge: Cambridge University Press).

Papafragou, A. (2002). Mindreading and verbal communication. *Mind and Language* 17, 55–67. doi: 10.1111/1468-0017.00189_17_1-2

Papafragou, A., Cassidy, K., and Gleitman, L. (2007). When we think about thinking: the acquisition of belief verbs. *Cognition* 105, 125–165. doi: 10.1016/j.cognition.2006.09.008

Papafragou, A., and Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition* 86, 253–282. doi: 10.1016/S0010-0277(02)00179-8

Parise, E., and Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychol. Sci.* 23, 728–733. doi: 10.1177/0956797612438734

Parise, E., and Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-month-old infants. *PLoS One* 8:e72360. doi: 10.1371/journal.pone.0072360

Parise, E., Friederici, A. D., and Striano, T. (2010). "Did you call me?" 5-month-old infants own name guides their attention. *PLoS One* 5:e14208. doi: 10.1371/journal.pone.0014208

Parise, E., Reid, V. M., Stets, M., and Striano, T. (2008). Direct eye contact infuences the neural processing of objects in 5-month-old infants. *Soc. Neurosci.* 3, 141–150. doi: 10.1080/17470910701865458

Paunov, A. M., Blank, I. A., and Fedorenko, E. (2019). Functionally distinct language and theory of mind networks are synchronized at rest and during language comprehension. *J. Neurophysiol.* 121, 1244–1265. doi: 10.1152/jn.00619.2018

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumit, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. USA* 100, 11702–11705. doi: 10.1073/pnas.1934290100

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., et al. (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. USA* 108, 16056–16061. doi: 10.1073/pnas.1102991108

Perner, J., Leekam, S. R., and Wimmer, H. (1987). Three-year-olds' difficulty with false belief: the case for a conceptual deficit. *Br. J. Dev. Psychol.* 5, 125–137. doi: 10.1111/j.2044-835x.1987.tb01048.x

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

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pléh, C. (2000). Modularity and pragmatics. *Pragmatics* 10, 415–438. doi: 10.1075/prag.10.4.04ple

Pléh, C. (2024). *Laying the foundations of independent psychology*, vol. *1*. London: Routledge.

Pomiechowska, B., Bródy, G., Csibra, G., and Gliga, T. (2021). Twelve-month-olds disambiguate new words using mutual-exclusivity inferences. *Cognition* 213:104691. doi: 10.1016/j.cognition.2021.104691

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., et al. (2018). Do infants understand false beliefs? We don't know yet – a commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cogn. Dev.* 48, 302–315. doi: 10.1016/j.cogdev.2018.09.005

Pouscoulous, N., and Noveck, I. A. (2009). "Going beyond semantics: the development of pragmatic enrichment" in *Language Acquisition* (London: Palgrave Macmillan UK), 196–215.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Reboul, A., and Moeschler, J. (1998). *La pragmatique aujourd'hui. Une nouvelle science de la communication*. Paris: Le Seuil.

Rubio-Fernandez, P. (2021). Pragmatic markers: the missing link between language and theory of mind. *Synthese* 199, 1125–1158. doi: 10.1007/s11229-020-02768-z

Rueschemeyer, S.-A., Gardner, T., and Stoner, C. (2015). The social N400 effect: how the presence of other listeners affects language comprehension. *Psychon. Bull. Rev.* 22, 128–134. doi: 10.3758/s13423-014-0654-x

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Dev. Rev.* 34, 265–293. doi: 10.1016/j.dr.2014.04.001

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind.". *NeuroImage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., et al. (2020). Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* 147, 293–327. doi: 10.1037/bul0000303

Scott, R. M., and Baillargeon, R. (2017). Early false-belief understanding. *Trends Cogn. Sci.* 21, 237–249. doi: 10.1016/j.tics.2017.01.012

Scott-Phillips, T. C. (2015). Meaning in animal and human communication. *Anim. Cogn.* 18, 801–805. doi: 10.1007/s10071-015-0845-5

Searle, J. R. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Senju, A., and Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Shain, C., Paunov, A., Chen, X., Lipkin, B., and Fedorenko, E. (2023). No evidence of theory of mind reasoning in the human language network. *Cereb. Cortex* 33, 6299–6319. doi: 10.1093/cercor/bhac505

Shannon, C. E., and Weaver, W. (1949). *The mathematical theory of communication*, vol. *97*. Champaign, IL: The University of Illinois Press.

Sirri, L., Linnert, S., Reid, V., and Parise, E. (2020). Speech intonation induces enhanced face perception in infants. *Sci. Rep.* 10:3225. doi: 10.1038/s41598-020-60074-7

Sloutsky, V. M., Yim, H., Yao, X., and Dennis, S. (2017). An associative account of the development of word learning. *Cogn. Psychol.* 97, 1–30. doi: 10.1016/j.cogpsych.2017.06.001

Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010

Southgate, V. (2020). Are infants Altercentric? The other and the self in early social cognition. *Psychol. Rev.* 127, 505–523. doi: 10.1037/rev0000182

Southgate, V., and Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition* 130, 1–10. doi: 10.1016/j.cognition.2013.08.008

Sperber, D., and Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Steil, J. N., Friedrich, C. K., and Schild, U. (2021). No evidence of robust noun-referent associations in German-learning 6-to 14-month-olds. *Front. Psychol.* 12, 1–15. doi: 10.3389/fpsyg.2021.718742

Tauzin, T., and Gergely, G. (2018). Communicative mind-reading in preverbal infants. *Sci. Rep.* 8, 9534–9539. doi: 10.1038/s41598-018-27804-4

Tauzin, T., and Gergely, G. (2019). Variability of signal sequences in turn-taking exchanges induces agency attribution in 10.5-mo-olds. *Proc. Natl. Acad. Sci. USA* 116, 15441–15446. doi: 10.1073/pnas.1816709116

Tauzin, T., and Gergely, G. (2021). Co-dependency of exchanged behaviors is a cue for agency attribution in 10-month-olds. *Sci. Rep.* 11:18217. doi: 10.1038/s41598-021-97811-5

Thompson, R. J. (2014). Meaning and mindreading. *Mind Lang.* 29, 167–200. doi: 10.1111/mila.12046

Tincoff, R., and Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy* 17, 432–444. doi: 10.1111/j.1532-7078.2011.00084.x

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). "Origins of human communication" in *Origins of human communication* (Cambridge, MA: The MIT Press).

Tomasello, M. (2018). How children come to understand false beliefs: a shared intentionality account. *Proc. Natl. Acad. Sci.* 115, 8491–8498. doi: 10.1073/pnas.1804761115

Urbach, T. P., DeLong, K. A., Chan, W.-H., and Kutas, M. (2020). An exploratory data analysis of word form prediction during word-by-word reading. *Proc. Natl. Acad. Sci.* 117, 20483–20494. doi: 10.1073/pnas.1922028117

Van Berkum, J. J. A. (2009). "The neuropragmatics of "simple" utterance comprehension: an ERP review" in *Semantics and pragmatics: From experiment to theory*, eds. R. Breheny, U. Sauerland, and K. A. Loparo (London: Palgrave Macmillian), 276–316.

Vouloumanos, A. (2018). Voulez-vous jouer avec moi? Twelve-month-olds understand that foreign languages can communicate. *Cognition* 173, 87–92. doi: 10.1016/j.cognition.2018.01.002

Vouloumanos, A., Martin, A., and Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Dev. Sci.* 17, 872–879. doi: 10.1111/desc.12170

Vouloumanos, A., Onishi, K. H., and Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proc. Natl. Acad. Sci. USA* 109, 12933–12937. doi: 10.1073/pnas.1121057109

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Warren, E., and Call, J. (2022). Inferential communication: bridging the gap between intentional and ostensive communication in non-human Primates. *Front. Psychol.* 12:718251. doi: 10.3389/fpsyg.2021.718251

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

Waxman, S. R., and Lidz, J. L. (2007). "Early world learning" in *Handbook of child psychology*. eds. D. Kuhn and R. Siegler. *6th* ed (Hoboken, NJ: Wiley).

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., and Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Dev. Psychol.* 34, 1289–1309. doi: 10.1037/0012-1649.34.6.1289

Werker, J. F., and Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* 50, 509–535. doi: 10.1146/annurev.psych.50.1.509

Westley, A., Kohút, Z., and Rueschemeyer, S. A. (2017). "I know something you don't know": discourse and social context effects on the N400 in adolescents. *J. Exp. Child Psychol.* 164, 45–54. doi: 10.1016/j.jecp.2017.06.016

Wilson, D., and Sperber, D. (2004). "Relevance theory" in *Handbook of pragmatics*. eds. L. Horn and G. Ward (Oxford: Blackwell), 163–176.

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Wittgenstein, L. (1953). *Philosophical investigations*. (AnscombeG. E. M.). Oxford: Blackwell.

Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *J. Child Psychol. Psychiatry* 17, 89–100. doi: 10.1111/j.1469-7610.1976.tb00381.x

Yoon, J. M. D., Johnson, M. H., and Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proc. Natl. Acad. Sci. USA* 105, 13690–13695. doi: 10.1073/pnas.0804388105

frontiers | Frontiers in Human Neuroscience

TYPE Brief Research Report
PUBLISHED 13 June 2024
DOI 10.3389/fnhum.2024.1386207

OPEN ACCESS

EDITED BY
Abdellah Fourtassi,
Aix Marseille Université, France

REVIEWED BY
Alexandra P. Key,
Vanderbilt University Medical Center,
United States
Linda J. Larson-Prior,
University of Arkansas for Medical Sciences,
United States
Graciela Catalina Alatorre Cruz,
National Autonomous University of Mexico,
Mexico

*CORRESPONDENCE
Sari Ylinen
✉ sari.p.ylinen@tuni.fi

RECEIVED 14 February 2024
ACCEPTED 21 May 2024
PUBLISHED 13 June 2024

CITATION
Ylinen S, Suppanen E, Winkler I and Kujala T (2024) Establishing neural representations for new word forms in 12-month-old infants.
Front. Hum. Neurosci. 18:1386207.
doi: 10.3389/fnhum.2024.1386207

Establishing neural representations for new word forms in 12-month-old infants

Sari Ylinen[1,2]*, Emma Suppanen[2], István Winkler[3] and Teija Kujala[2]

[1]Logopedics, Welfare Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland, [2]Cognitive Brain Research Unit, Centre of Excellence in Music, Mind, Body and Brain, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [3]Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary

During the first year of life, infants start to learn the lexicon of their native language. Word learning includes the establishment of longer-term representations for the phonological form and the meaning of the word in the brain, as well as the link between them. However, it is not known how the brain processes word forms immediately after they have been learned. We familiarized 12-month-old infants (N = 52) with two pseudowords and studied their neural signatures. Specifically, we determined whether a newly learned word form elicits neural signatures similar to those observed when a known word is recognized (i.e., when a well-established word representation is activated, eliciting enhanced mismatch responses) or whether the processing of a newly learned word form shows the suppression of the neural response along with the principles of predictive coding of a learned rule (i.e., the order of the syllables of the new word form). The pattern of results obtained in the current study suggests that recognized word forms elicit a mismatch response of negative polarity, similar to newly learned and previously known words with an established representation in long-term memory. In contrast, prediction errors caused by acoustic novelty or deviation from the expected order in a sequence of (pseudo)words elicit responses of positive polarity. This suggests that electric brain activity is not fully explained by the predictive coding framework.

KEYWORDS

auditory processing, electroencephalography (EEG), event-related potential (ERP), language development, word learning

Introduction

Infants readily learn from their auditory environment, including features of their native language spoken by their family. For example, infants can extract different patterns or rules from speech they hear (statistical learning; Gómez and Gerken, 2000; Saffran and Kirkham, 2018) and make predictions based on them (Emberson et al., 2019; Suppanen et al., 2022). Our previous research has suggested that the ability of newborn infants to learn from speech exposure and to make predictions about future input is linked to their later language skills (Suppanen et al., 2022). Learning from speech also enables infants to start building their mental lexicon during the first year of life, which requires the establishment of word representations in the brain that link the phonological representation of the word form with the corresponding semantic representation (Gupta and Tisdale, 2009). These neural representations serve as top-down templates and enable infants to recognize words from bottom-up input and understand their meaning. These neural representations are also reflected in infant brain activity: previous studies have shown distinct brain responses for learned words and unknown words or pseudowords in infants at 12–16 months (Molfese, 1989, 1990; Molfese et al., 1993; Mills et al., 1997, 2004; St. George and Mills, 2001; Ylinen et al., 2017).

Because newborn infants do not yet have long-term representations of words, their learning from speech input may rather be dominated by learning regularities, patterns, or rules and using them in predictive processing. According to the predictive coding theory, during perception, feedback signals are generated in a hierarchically organized neural network to predict the perceptual input. The difference between the predicted and actual input drives changes in the predictions to minimize this difference, thereby reducing surprise (Rao and Ballard, 1999) or free energy (Friston, 2005). As a result, predicted items result in weak or no prediction error signals (i.e., weaker brain responses), whereas unpredicted items evoke strong prediction error signals (i.e., stronger brain responses). This pattern was observed in our previous study of newborn infants (Suppanen et al., 2022). However, it is not clear how the processes and neural signatures of prediction and recognition change when infants are able to establish word representations during the second half of the first year of life. Our previous study (Ylinen et al., 2017) utilized disyllabic words and a study paradigm in which generating predictions of word endings based on word beginnings resulted in either enhanced negative-polarity mismatch responses (MMRs) due to the activation of long-term representations for a familiar word, or prominent positive-polarity prediction error responses for an unfamiliar word form in 12-month-old infants. While these results concern the processing of previously learned words, they raise the question of how the infant brain processes newly learned word forms at the same age. To this end, here we studied whether, in 12-month-old infants, a newly learned word form elicits neural signatures that resemble those of the recognition of a word by activation of an established word representation (Ylinen et al., 2017), or, rather, the processing of a newly learned word form shows the suppression of the neural response along with the principles of predictive encoding of a learned rule.

To study the learning of novel word forms, we presented infants with pseudowords in an experiment comprising two phases: (1) a familiarization phase in which the infants were presented with two spoken native-language disyllabic pseudowords (designated as “AB” and “CD,” where A, B, C, and D denote different syllables), and (2) a test phase with an oddball sequence in which one of the familiarized pseudowords (AB) served as the frequent standard stimulus interspersed with three rare deviant word forms (pseudowords or actual words): CD, AD, and AX (where X represents a syllable that did not appear during familiarization). The AD deviant was an actual word that is often known by 12-month-old infants: ‘kukka’ (/kuk:a/; a flower). The first syllable of CD was expected to elicit a frontocentral positive-polarity MMR for the acoustic change from A to C [for reviews of the mismatch negativity (MMN) or MMRs in adults and infants, see Näätänen, 2001; Kujala et al., 2023]. In addition, since we were particularly interested in word-level processing, six hypotheses were tested regarding how infants process the second syllables of the (pseudo)words, for which processing cues commence at the onset of the second syllable (300 ms from word onset in the current study).

I) Because the auditory sequence with the frequent stimulus AB was likely to create a prediction for the repetition of AB, the syllables D of AD, X of AX, and C of CD could all elicit MMRs, reflecting the prediction error within the sound sequence (Ylinen et al., 2017). These MMRs are expected to have a frontocentral scalp distribution, and they could be either negative or positive in their polarity (see Näätänen et al., 2019, for a review). At 6–12 months, their latency has been reported to range from approximately 150 ms (Ylinen et al., 2017) to 450 ms from change onset (Cheng et al., 2013), depending on the characteristics of the stimuli and their context. Therefore, in the majority of cases, it is difficult to set specific hypotheses about MMR latency (see Näätänen et al., 2019), but the responses of interest were expected to occur between 150 and 450 ms from the onset of the second syllable.

II) The first syllables of the familiarized disyllabic word forms create predictions for their familiarized second syllable. Because AD and AX violate the familiarized continuation of the AB word form, they should cause within-word prediction errors, as shown in our previous studies (Ylinen et al., 2017; Suppanen et al., 2022). However, these prediction errors are expected to differ from each other (see additional hypotheses III and IV; Suppanen et al., 2022).

III) The response to the syllable X in AX should show the effects of novelty, as the X syllable has not appeared during the familiarization phase, and it was also rare within the test sequences. Novelty is typically associated with a frontocentral positive-polarity auditory ERP component in both infants and adults (see Kushnerenko et al., 2013, for a review). In our previous study with neonates (Suppanen et al., 2022), stimulus AX elicited a robust positive response that peaked at 300 ms from change onset. At 12 months of age, the latency may be slightly shorter due to maturation.

IV) Because AD is an actual word that could have been learned by the infants in their normal language environment, AD may elicit an enhanced response representing word or word-form recognition (Pulvermüller et al., 2001; for the long-term memory contribution to the MMN, see also Näätänen et al., 1997; Winkler et al., 1999; Ylinen et al., 2010). Based on our previous study in the same age group delivering the same stimulus as in the current study (but with a different kind of context in the sound sequence; Ylinen et al., 2017), we expected the word recognition response to be of negative polarity.

V) The response to D in the deviant CD might be suppressed if predicted based on the learned rule that C is followed by D.

VI) Alternatively, the response to D in the deviant CD might elicit an enhanced response of negative polarity, similar to what was hypothesized for a real word AD (hypothesis IV), if CD activates a word-form representation established during the learning phase. (Note that Hypotheses V and VI are mutually exclusive.)

Methods

Ethics statement

The study protocol was approved by the Ethics Committee for Gynecology and Obstetrics, Pediatrics, and Psychiatry of the Hospital

District of Helsinki and Uusimaa, Finland. Participants' parents gave their informed written consent.

Participants

This study was part of a larger project (Suppanen et al., 2022) in which 75 healthy, full-term newborn infants born into Finnish-speaking families were studied (see Suppanen et al., 2022, for details). Of these 75 infants, 68 participated in an EEG measurement at 12 months of age. Data from 16 participants were excluded due to participants missing or discontinuing the EEG recording, technical problems, or failure to meet the criterion of 50 accepted epochs per stimulus type. Thus, the data from 52 participants were included in the analyses (26 boys and 26 girls, average age 369 days, and SD 14 days).

Stimuli and study design

The auditory stimuli (see Figure 1) consisted of the phonotactically legal Finnish disyllabic pseudowords AB (/kut:o/), CD (/tek:a/), and AX (/kup:e/) and the word AD (/kuk:a/, which means flower). They were spoken in a sound-isolated studio by two native speakers of the Finnish language (one male and one female). For each syllable, the two most prototypical exemplars without clear co-articulatory cues revealing the original context were selected from each speaker and further processed with Praat (Boersma and Weenink, 2010). The intensities of the syllables in each position were matched as closely as possible. The duration of the stimuli was adjusted to 426 ms (the first syllable was 90 ms, a silent pause mimicking the occlusion phase of a stop consonant was 210 ms, and the second syllable was 126 ms). The onset of the second syllable was at 300 ms from the stimulus onset. Some variation in F0 was allowed within each speaker because we aimed for natural-sounding stimuli (for acoustic details, see Supplementary Table S1).

In the familiarization phase, the disyllabic pseudowords /kut:o/ and /tek:a/, denoted here as AB and CD, respectively, were presented to the participants with 50% probability (3 blocks, each having 250 stimuli; total duration 11.7 min). In the familiarization phase, half of the infants heard sequences in which 80% of the stimuli were spoken by a female speaker and the rest by a male speaker; the ratios were reversed in the other half of the infants.[1] In the test phase, participants were presented with four oddball blocks (540 stimuli in each block, total duration 33.7 min) with the familiarized pseudoword AB as the standard stimulus ($p = 0.79$) and three other word forms (CD, AD, and AX) as deviants ($p = 0.07$ for each). The test phase took place immediately after the familiarization phase. The interstimulus interval (offset to onset) was 510 ms in both phases. The total recording time was approximately 45 min.

The presentation order was randomized with the following constraints: Each stimulus block started with at least eight standard stimuli, and at least two standards followed each deviant. Stimuli in the test phase were spoken by the same speaker, with half of the infants receiving male-only stimuli and the other receiving female-only stimuli in a counterbalanced fashion. All the data were pooled together for the current data analysis.

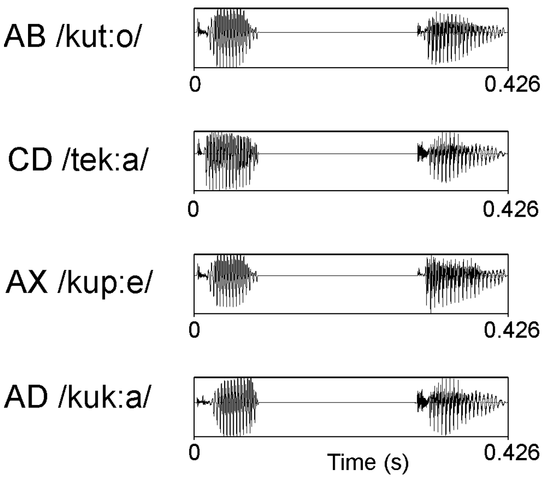


FIGURE 1
Waveforms of example stimuli spoken by a female native speaker of Finnish.

Data acquisition and procedure

EEG data were recorded with 16 active electrodes placed according to the international 10–20 system (Fp1/2, F3/4, Fz, C3/4, Cz, P3/4, Pz, O1/2, Oz), with additional electrodes on the left and right mastoids (LM and RM). The used amplifier was QuickAmp (version 10.08.14; Brain Products GmbH, Gilching, Germany), and the recording software was BrainVision Recorder (version 1.20.0801; Brain Products GmbH). The sampling rate was 500 Hz with a 100 Hz online lowpass filtering cutoff frequency. The recording reference was the average of all electrodes.

The participants were awake and sitting on their parents' laps during the measurement, and the parents entertained the participants during the measurement by silently showing them toys. The stimuli were presented in Presentation 17.2 Software (Neurobehavioral Systems Ltd., Berkeley, CA, United States) via two Genelec speakers: (Genelec Oy, Iisalmi, Finland) placed front left and front right approximately 160 cm from the participant. The approximate sound pressure level (SPL) was 65 dB.

Data analysis

Only the data collected in the test phase are reported here. The data were preprocessed using BESA Research 6.0 (Besa GmbH, Gräfelting, Germany), MATLAB Release 2018b (The MathWorks Inc., Natick, Massachusetts, United States), EEGlab 2019.0 (Delorme and Makeig, 2004), and in-house MATLAB scripts (CBRUPlugin2.1b, Tommi Makkonen, Cognitive Brain Research Unit, University of Helsinki). The data were first bandpass-filtered offline (0.5–30 Hz, 24 dB/octave), re-referenced to the average of the two mastoid signals (RM and LM), and segmented into −100 to 800 ms epochs with respect to stimulus

[1] The two voices in the learning phase were designed to address the infants' processing of the voice; they will be reported separately. In the current study, the ERPs of the test phase were analyzed.

onset, separately for each stimulus and participant. The epochs were baseline-corrected by the average voltage in the 100 ms pre-stimulus interval. Epochs with an absolute amplitude exceeding ±100 μV and the responses to the first two standard stimuli immediately after a deviant were rejected. The data from participants with less than 50 accepted epochs for any stimulus type were excluded from further analysis. The average number of remaining epochs per participant was 360 for the standard stimulus and 69 for the deviant stimulus.

The epochs were binned and averaged according to the stimulus type. ERP difference responses for each deviant type were calculated by subtracting the standard waveform from that of the deviant. Mean amplitudes of frontocentral channels (F3, Fz, F4, C3, Cz, and C4) were extracted for four 60 ms time windows based on the peak latencies observed in the grand-average deviant-minus-standard difference waveforms: 120–180 ms from stimulus onset (Time Window 1), 460–520 ms (Time Window 2), 520–580 ms (Time Window 3), and 620–680 ms (Time Window 4). Frontocentral channels were included in line with previous infant MMR studies (e.g., Choudhury and Benasich, 2011; Ylinen et al., 2017); this is also in line with the frontocentral dominance of the MMN in adults (Näätänen, 2001).

The presence of MMRs or prediction error responses in each condition and time window was tested using one-sample, two-tailed *t*-tests. This involved comparing the response amplitudes derived from deviant-minus-standard difference waveforms, averaged across frontocentral channels (F3, Fz, F4, C3, Cz, and C4), to zero (the baseline). Effect sizes were estimated using Cohen's *d.* In addition, to compare the response amplitudes for the three deviant types within each time window, the amplitudes derived from deviant-minus-standard difference waveforms were submitted to one-way analyses of variance (ANOVA) with the factor *Deviant* (CD vs. AD vs. AX). The effect sizes are reported using the η^2 measure. *Post-hoc* tests were conducted using Bonferroni-corrected t-tests (effect sizes: Cohen's *d*).

Results

All deviant types elicited a response that differed significantly from the baseline (see Table 1 for mean amplitudes derived from the deviant-minus-standard difference waveforms and *t*-test results for the significant responses in each time window; see Figure 2 for the original responses and Figure 3 for the group-averaged deviant-minus-standard waveforms).

The ANOVA for Time Window 1 (120–180 ms, the 1st syllable) yielded a significant effect of *Deviant* [$F(2,102) = 8.83$, $p < 0.001$, $\eta^2 = 0.15$]. Investigating this effect further with Bonferroni-corrected pairwise comparisons, the response to CD was significantly more positive than that to AD and AX [$p < 0.01$ for both, $d = 0.49$ and $d = 0.57$, respectively]. Similarly, the ANOVA for Time Window 2 (460–520 ms) showed a significant *Deviant* effect [$F(2, 102) = 5.4$, $p < 0.01$, $\eta^2 = 0.10$], and Bonferroni-corrected pairwise comparisons showed that the response to CD was significantly more negative than that to AX [$p < 0.01$, $d = 0.49$]. In Time Window 3 (520–580 ms), the ANOVA also showed a significant effect of *Deviant* [F(2, 102) = 22.3, $p < 0.001$, $\eta^2 = 0.31$]. Bonferroni-corrected pairwise comparisons revealed that the response elicited by AX was significantly more positive than those elicited by CD and AD [$p < 0.001$ for both, $d = 0.82$ and $d = 0.87$, respectively]. Furthermore, the *Deviant* effect was significant in the ANOVA for Time Window 4 (620–680 ms) [F(2, 102) = 4.33, $p < 0.05$, $\eta^2 = 0.08$]. According to Bonferroni-corrected pairwise comparisons, AD elicited a significantly more negative response than either CD or AX [$p < 0.05$ and $d = 0.37$ for both].

Discussion

The current study examined whether a newly learned word form elicits brain responses reflecting word-form recognition in 12-month-old infants or whether predictive processing is enabled by a learned rule. The answer to this question was assessed by measuring ERP responses to familiarized and unfamiliarized (pseudo)words. ERPs showed positive-polarity responses for the first and second syllable changes in pseudowords (CD and AX, respectively). However, the second syllables of a common word AD and a familiarized pseudoword CD elicited negative-polarity responses.

Confirming Hypothesis I (sequential deviation), all sequential deviants elicited responses that differed from the standard after their onset of deviation (Table 1 and Figure 3; Kushnerenko et al., 2013). The positive-polarity response to CD in Time Window 1 likely reflects MMR to the acoustic deviance of the first syllable of CD from the standard AB. In line with Hypotheses I (sequential deviation), II (within-word prediction error), and III (novelty response), the robust positive-polarity response in Time Window 3 elicited by the novel syllable X completing an unfamiliar word form likely reflects the sum of the word-level prediction error response (Ylinen et al., 2017; Suppanen et al., 2022), the novelty response (Kushnerenko et al., 2013), and the response to rare acoustic parameters (Kushnerenko et al., 2007). This interpretation of the

TABLE 1 Deviant-minus-standard difference amplitudes significantly differing from zero and the results of the one-sample *t*-tests (two-tailed), separately for each deviant in each Time Window.

Difference response	CD /tek:a/ (vs. AB /kut:o/)	AD /kuk:a/ (vs. AB /kut:o/)	AX /kup:e/ (vs. AB /kut:o/)
Time Window 1 (120–180 ms)	****1.3** (2.9) $t(51) = 3.2$, $p < 0.01$, $d = 0.44$		
Time Window 2 (460–520 ms)	***−1.06** (3.3) $t(51) = −2.3$, $p < 0.05$, $d = −0.32$		
Time Window 3 (520–580 ms)	***−0.86** (3.1) $t(51) = −2.01$, $p < 0.05$, $d = −0.28$	***−0.70** (2.4) $t(51) = −2.1$, $p < 0.05$, $d = −0.29$	****2.07** (2.5) $t(51) = 5.9$, $p < 0.001$, $d = 0.81$
Time Window 4 (620–680 ms)		****−0.85** (2.3) $t(51) = −2.7$, $p < 0.05$, $d = −0.37$	

Mean amplitudes (in bold) and standard deviations (in parentheses) are given in μV for the average of the frontocentral channels. The *t*-statistics (*t*, df—degrees of freedom, in parentheses, *p*—significance level, Cohen's *d*—effect size) are also shown. Statistical significance is marked with asterisks (**p* < 0.05, ***p* < 0.01).

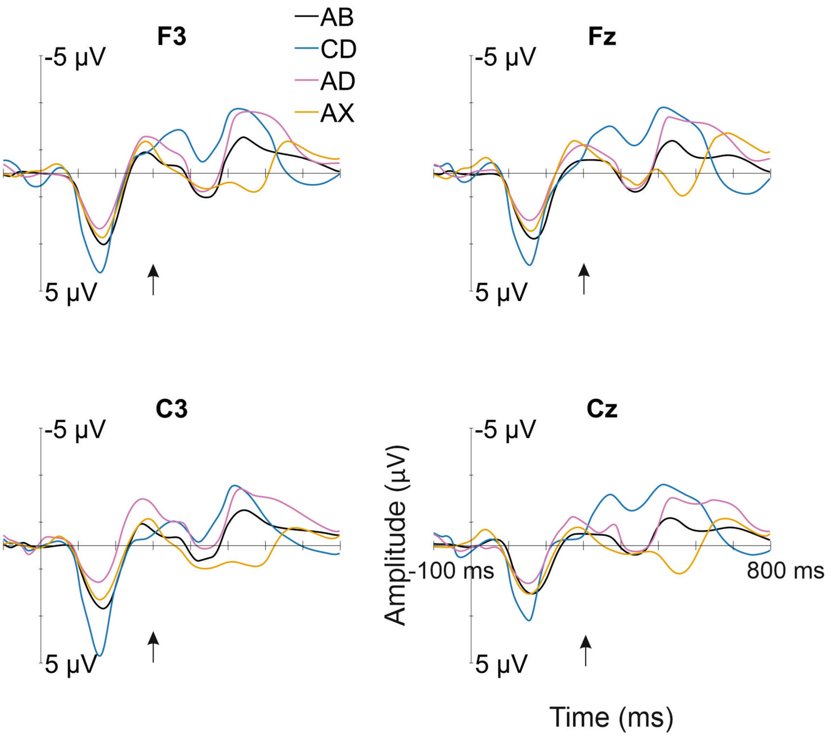


FIGURE 2
Group-averaged (N = 52) responses to the familiarized standard (AB—/kut:o/; black line) and the deviants (CD—/tek:a/, the other familiarized pseudoword; AD—/kuk:a/, the combination of the syllables of the two familiarized pseudowords forming a common word that infants might know; AX—/kup:e/, a novel pseudoword starting as the standard, but containing an unfamiliar syllable) from the electrode sites used in the statistical analyses. The Y-axis is at the stimulus onset, and the onset of the second syllable is marked with a black arrow.

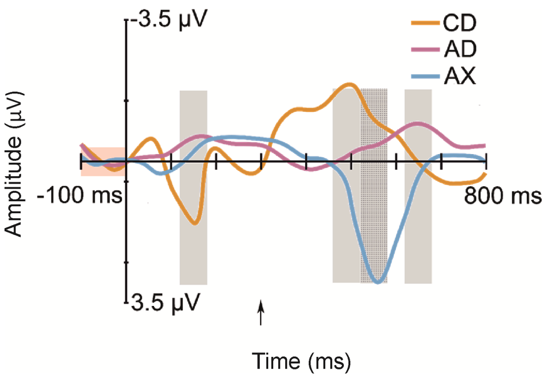


FIGURE 3
Group-averaged (N = 52) frontal (Fz) difference responses to the deviants: CD—/tek:a/, the other familiarized pseudoword besides the standard; AD—/kuk:a/, the combination of the syllables of the two familiarized pseudowords that forms a real word; AX—/kup:e/, a novel pseudoword starting as the standard, but containing an unfamiliar syllable. Measurement time windows are marked with light gray rectangles, and the baseline is marked with a light red rectangle. The Y-axis is at the stimulus onset, and the onset of the second syllable is marked with a black arrow.

functional distinction between the positive MMR to acoustic deviation for the first syllable of CD in Time Window 1 and the positive response to AX in Time Window 3 is also supported by different latencies from change onset (120–180 ms from the 1st syllable onset vs. 220–280 ms from the 2nd syllable onset).

Because sequential acoustic deviance resulted in positive-polarity responses in the current data (see also Kushnerenko et al., 2007), acoustic deviance cannot account for the negative response for the second syllable of AD, which differed significantly from the baseline in Time Window 4. In contrast, the observed response is in line with our earlier study (Ylinen et al., 2017), in which we found, in 12-month-old infants, a negative-polarity response for the syllable completing the word /kuk:a/, but not for the acoustically identical syllable /ka/ presented in isolation, and, thus, the observed negative-polarity response was explained by the activation of the word representation for /kuk:a/ ('flower'). Similarly, in line with Hypothesis IV, the negative-polarity response elicited by the same word as in our previous study (referred to as AD in the current description; the only actual word delivered in the sequences) can be interpreted as reflecting word recognition via the activation of a word representation in the infant brain (see also Garagnani and Pulvermüller, 2011).

If the processing of newly learned word forms is dominated by predictive processing resulting from rule learning (Hypothesis V), then the response to the D syllable in the CD pseudoword should be suppressed because the infants have learned during familiarization that C is followed by D and, thus, C predicts D. The alternative Hypothesis VI, in turn, states that if the processing of newly learned word forms in the infant's brain is dominated by recognition of the newly learned word form, then the D syllable in the CD pseudoword should elicit an enhanced response resembling that observed for AD (the actual word that could be known by the infants; see above). The pattern of current responses is compatible with the latter hypothesis: CD elicited a prominent negative-polarity response similar to AD rather than a suppressed response. Therefore, we interpret the negative-polarity response to the familiarized CD pseudoword as reflecting the activation of a newly established representation and the recognition of the word form for CD learned from speech exposure (during the familiarization phase). Thus, the present pattern of data supports Hypothesis VI, suggesting that successful word-level prediction (the first syllable predicting the second syllable) can be indexed by an enhanced ERP response of negative polarity at 12 months, even for newly learned word forms that had no long-term memory representation before.

Despite both AD and CD showing negative-polarity responses to D, there were also differences between these responses: the response to the common word AD peaked at a longer latency than that to CD, which may be explained by the recent familiarization (possibly higher activation state) of the pseudoword CD. In addition, the response to the common word AD was not as distinct as the one to the familiarized pseudoword, likely because, according to parental reports, not all infants did yet know the word /kuk:a/ (here, AD), which could cause variation in the individual responses and result in a less sharp or lower-amplitude response. (Note that in the study by Ylinen et al., 2017, the infants were familiarized with the word *kukka* beforehand, whereas in the current study, they were not.)

The current study has, however, some limitations. Using stimulus types that violate the infants' expectations in different ways would have allowed us to obtain a more detailed picture of infants' predictive inference. However, the experiment would probably have been too long for the infants. In addition, a control condition in which the same syllables of interest (here, the final syllables) were presented in isolation, without a word context, would have allowed us to tease apart factors that might contribute to infants' responses, including the acoustic properties of the speech stimuli and the effect of word context in creating expectations about the future input. Again, such a control condition was not possible due to time constraints. The latter limitation concerns mostly the novel deviant AX; however, since the other two deviants, namely a potentially familiar word AD and a novel (pseudo)word CD, shared the same critical syllable D, the observed differences in the responses to their second syllable could not be explained by the acoustic properties of the stimuli.

In conclusion, in 12-month-old infants, a newly learned word form appears to elicit an ERP response of negative polarity, potentially reflecting word-form recognition and resembling the responses elicited by familiar words established in long-term memory. In contrast, acoustic changes and other prediction errors in a sequence consisting of (pseudo)words elicit ERP responses of positive polarity. This suggests that although predictive processing takes place, successful learning, which enables correct prediction, does not result in suppressed responses (*cf.* Heilbron and Chait, 2018).

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Ethics Committee for Gynecology and Obstetrics, Pediatrics, and Psychiatry of the Hospital District of Helsinki and Uusimaa. The studies were conducted in accordance with local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

SY: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. ES: Data curation, Formal analysis, Funding acquisition, Investigation, Visualization, Writing – original draft, Writing – review & editing. IW: Conceptualization, Writing – original draft, Writing – review & editing. TK: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Supported by the Research Funds of the University of Helsinki (project Predictive Coding in Language Learning, awarded to SY), the Academy of Finland (Project Nos. 346211, 274058, and 316970), the National Research Fund of Hungary (Grant Number K147135 to IW), and the Emil Aaltonen Foundation (to ES).

Acknowledgments

The authors would like to thank the participants' families for taking part in the study. In addition, the authors acknowledge Anna-Riikka Smolander for her help with the EEG measurements in addition to Tommi Makkonen, Jaakko Kauramäki, and Jaakko Sulkko for technical assistance. This project was conducted in cooperation with the Hospital District of Helsinki and Uusimaa.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2024.1386207/full#supplementary-material

References

Boersma, P., and Weenink, D. (2010). Praat: doing phonetics by computer. Version 6.0.12 [software]. Downloaded March 7. Available at: www.praat.org

Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., and Lee, C. Y. (2013). The development of mismatch responses to mandarin lexical tones in early infancy. *Dev. Neuropsychol.* 38, 281–300. doi: 10.1080/87565641.2013.799672

Choudhury, N., and Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year language and cognitive abilities. *Clin. Neurophysiol.* 122, 320–338. doi: 10.1016/j.clinph.2010.05.035

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

Emberson, L. L., Boldin, A. M., Robertson, C. E., Cannon, G., and Aslin, R. N. (2019). Expectation affects neural repetition suppression in infancy. *Dev. Cogn. Neurosci.* 37:100597. doi: 10.1016/j.dcn.2018.11.001

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Garagnani, M., and Pulvermüller, F. (2011). From sounds to words: a neurocomputational model of adaptation, inhibition and memory processes in auditory change detection. *NeuroImage* 54, 170–181. doi: 10.1016/j.neuroimage.2010.08.031

Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186. doi: 10.1016/S1364-6613(00)01467-4

Gupta, P., and Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 364, 3755–3771. doi: 10.1098/rstb.2009.0132

Heilbron, M., and Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061

Kujala, T., Partanen, E., Virtala, P., and Winkler, I. (2023). Prerequisites of language acquisition in the newborn brain. *Trends Neurosci.* 46, 726–737. doi: 10.1016/j.tins.2023.05.011

Kushnerenko, E. V., Van den Bergh, B. R., and Winkler, I. (2013). Separating acoustic deviance from novelty during the first year of life: a review of event-related potential evidence. *Front. Psychol.* 4:595. doi: 10.3389/fpsyg.2013.00595

Kushnerenko, E., Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., et al. (2007). Processing acoustic change and novelty in newborn infants. *Eur. J. Neurosci.* 26, 265–274. doi: 10.1111/j.1460-9568.2007.05628.x

Mills, D. L., Coffey-Corina, S. A., and Neville, H. J. (1997). Language comprehension and cerebral specialization from 13–20 months. *Dev. Neuropsychol.* 13, 397–445. doi: 10.1080/87565649709540685

Mills, D. L., Prat, C., Zangl, R., Stager, C. L., Neville, H. J., and Werker, J. F. (2004). Language experience and the Organization of Brain Activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *J. Cogn. Neurosci.* 16, 1452–1464. doi: 10.1162/0898929042304697

Molfese, D. L. (1989). Electrophysiological correlates of word meanings in 14-month-old human infants. *Dev. Neuropsychol.* 5, 79–103. doi: 10.1080/87565648909540425

Molfese, D. L. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain Lang.* 38, 345–363. doi: 10.1016/0093-934X(90)90120-6

Molfese, D. L., Wetzel, W. F., and Gill, L. A. (1993). Known versus unknown word discriminations in 12-month-old human infants: electrophysiological correlates. *Dev. Neuropsychol.* 9, 241–258. doi: 10.1080/87565649309540555

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21. doi: 10.1111/1469-8986.3810001

Näätänen, R., Kujala, T., and Light, G. (2019). "The development of MMN" in *The mismatch negativity: A window to the brain*. eds. R. Näätänen, T. Kujala and G. Light (Oxford: Oxford Academic).

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., et al. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage* 14, 607–616. doi: 10.1006/nimg.2001.0864

Rao, R., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Saffran, J. R., and Kirkham, N. Z. (2018). Infant statistical learning. *Annu. Rev. Psychol.* 69, 181–203. doi: 10.1146/annurev-psych-122216-011805

St. George, M., and Mills, D. L. (2001). "Electrophysiological studies of language development" in *Language acquisition and language disorders*. eds. J. Weissenborn and B. Hoehle (Amsterdam: John Benjamins Publishing), 247–259.

Suppanen, E., Winkler, I., Kujala, T., and Ylinen, S. (2022). More efficient formation of longer-term representations for word forms at birth can be linked to better language skills at 2 years. *Dev. Cogn. Neurosci.* 55:101113. doi: 10.1016/j.dcn.2022.101113

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642. doi: 10.1017/S0048577299981908

Ylinen, S., Bosseler, A., Junttila, K., and Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Dev. Sci.* 20:e12472. doi: 10.1111/desc.12472

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., et al. (2010). Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *J. Cogn. Neurosci.* 22, 1319–1332. doi: 10.1162/jocn.2009.21272

TYPE Original Research
PUBLISHED 17 June 2024
DOI 10.3389/fpsyg.2024.1370007

Check for updates

OPEN ACCESS

EDITED BY
Marcela Pena,
Pontificia Universidad Católica de Chile, Chile

REVIEWED BY
Bahia Guellai,
Université Paris Nanterre, France
Paula Maarit Virtala,
University of Helsinki, Finland

*CORRESPONDENCE
Natalie Boll-Avetisyan
✉ nboll@uni-potsdam.de

RECEIVED 13 January 2024
ACCEPTED 29 May 2024
PUBLISHED 17 June 2024

CITATION
Boll-Avetisyan N, Shandala A and Langus A (2024) Infants show systematic rhythmic motor responses while listening to rhythmic speech.
Front. Psychol. 15:1370007.
doi: 10.3389/fpsyg.2024.1370007

Infants show systematic rhythmic motor responses while listening to rhythmic speech

Natalie Boll-Avetisyan[1,2]*, Arina Shandala[1,3] and Alan Langus[1]

[1]Department of Linguistics, University of Potsdam, Potsdam, Germany, [2]Research Focus Cognitive Sciences, University of Potsdam, Potsdam, Germany, [3]International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), University of Groningen, Netherlands/University of Newcastle, United Kingdom/University of Potsdam, Germany and Macquarie University, Sydney, NSW, Australia

Rhythm is known to play an important role in infant language acquisition, but few infant language development studies have considered that rhythm is multimodal and shows strong connections between speech and the body. Based on the observation that infants sometimes show rhythmic motor responses when listening to auditory rhythms, the present study asked whether specific rhythm cues (pitch, intensity, or duration) would systematically increase infants' spontaneous rhythmic body movement, and whether their rhythmic movements would be associated with their speech processing abilities. We used pre-existing experimental and video data of 148 German-learning 7.5- and 9.5-month-old infants tested on their use of rhythm as a cue for speech segmentation. The infants were familiarized with an artificial language featuring syllables alternating in pitch, intensity, duration, or none of these cues. Subsequently, they were tested on their recognition of bisyllables based on perceived rhythm. We annotated infants' rhythmic movements in the videos, analyzed whether the rhythmic moving durations depended on the perceived rhythmic cue, and correlated them with the speech segmentation performance. The result was that infants' motor engagement was highest when they heard a duration-based speech rhythm. Moreover, we found an association of the quantity of infants' rhythmic motor responses and speech segmentation. However, contrary to the predictions, infants who exhibited fewer rhythmic movements showed a more mature performance in speech segmentation. In sum, the present study provides initial exploratory evidence that infants' spontaneous rhythmic body movements while listening to rhythmic speech are systematic, and may be linked with their language processing. Moreover, the results highlight the need for considering infants' spontaneous rhythmic body movements as a source of individual differences in infant auditory and speech perception.

KEYWORDS

infants, rhythm perception, rhythmic body movements, rhythmic cues, speech segmentation, individual differences

1 Introduction

It is widely acknowledged that the perception of speech rhythm helps infants to tune into the language that surrounds them (Gleitman and Wanner, 1982; Langus et al., 2017). Newborns can already distinguish between languages that differ in their overall speech rhythm (Nazzi et al., 1998; for a meta-analysis, see Gasparini et al., 2021). At 4–6 months, infants have internalized their native languages' metrical structure (Friederici et al., 2007; Höhle et al., 2009), and from 7 months onwards they can use rhythmic cues for identifying words (e.g., Echols et al., 1997; Jusczyk et al., 1999; Abboub et al., 2016) and phrases (Hirsh-Pasek et al., 1987; Nazzi et al., 2000) in continuous speech. These abilities are an important step for acquiring words and syntax (Gleitman and Wanner, 1982; Christophe et al., 1994; Jusczyk, 1997; Weissenborn and Höhle, 2001), and predict later language skills (Newman et al., 2006; Junge et al., 2012; Cristia et al., 2014; Höhle et al., 2014; Marimon et al., 2022). Surprisingly, only a few studies investigating the role of rhythm in language acquisition have considered that rhythm is multimodal and has intrinsic connections between speech and the body, although it is well established that regular rhythm – such as in music – facilitates sensorimotor synchronization of bodily movements to perceived rhythmic regularities. Hence, if infants perceive rhythm in spoken language, they might express this sensitivity by showing rhythmic engagement with the speech rhythm. The goal of the present study was thus to explore whether infants spontaneously produce systematic rhythmic body movements while listening to the rhythm of spoken language and whether such rhythmic engagement supports the perception and acquisition of spoken language.

Our investigation of the potential link between infants' body and speech rhythm perception was inspired by a coincidental observation of infants' spontaneous body movements in experiments that investigate their language development. Language acquisition research often employs artificial language learning paradigms for studying what type of speech cues infants rely on for extracting word-like units from continuous speech. In this paradigm, infants are familiarized with artificial miniature languages that are highly reduced nonsense speech streams. After listening to these streams for a few minutes, infants are tested on their recognition of specific syllable combinations that were or were not part of the artificial speech stream (for example, the co-occurrence probabilities of syllables: Saffran et al., 1996; Aslin et al., 1998; rhythmic/prosodic cues: Thiessen and Saffran, 2003; Abboub et al., 2016; Marimon et al., 2022; phonotactic patterns: Mintz et al., 2018). These artificial languages are often designed such that they present syllables organized in a repetitive order resulting in a highly rhythmic auditory signal that seems to facilitate the perception of these artificial languages (Johnson and Tyler, 2010; Lew-Williams and Saffran, 2012; Mersad and Nazzi, 2012; Marimon et al., 2022). Interestingly, when running such experiments, we have also observed that many infants spontaneously move their bodies rhythmically, as if dancing to the rhythm of these artificial languages. Here we therefore investigate whether such rhythmic motor engagement with the rhythmic speech in artificial language learning experiments reflects infants' processing of the speech rhythms, and whether individual differences in motor engagement is associated with their speech processing performance.

There is some evidence that speech perception and bodily movements are intrinsically connected. Previous research into the multimodality of speech perception has drawn a link between sensorimotor information and speech perception by showing that phoneme perception is modulated by restricting lip movements in infants as early as 4.5 months of age (Yeung and Werker, 2013; Bruderer et al., 2015). Beyond articulatory gestures, when speaking, humans' body gestures move in synchrony with speech prosody, that is, the melody and rhythm of speech (e.g., Wilson and Wilson, 2005; Schmidt-Kassow and Kotz, 2008; Cummins, 2009; Guellai et al., 2014). In fact, hand movements are typically entrained with the prosodically strongest syllables of words and phrases (e.g., De Ruiter, 1998). This synchrony between speech prosody and hand movements starts to emerge when infants produce their earliest babbles (Esteve-Gibert and Prieto, 2014). Also, when listening to speech, infants have been found to follow the rhythm of speech with their body movements from birth (e.g., Condon and Sander, 1974; Mundy-Castle, 1980; Papoušek et al., 1991; Papoušek, 1992; Masataka, 1993; Kuhl et al., 1997). While spoken language is not perfectly rhythmic, it can assume a regular rhythm in many everyday activities including music, poetry, and nursery rhymes, where the regular metrical rhythmic patterns occur in an exaggerated form. In the context of these highly regularized rhythms, the link between body and perception is even more noteworthy: across cultures, both adults and children dance, bounce, or tap in synchrony with the beat when listening to music (e.g., Brown et al., 2004; Kirschner and Tomasello, 2009; Fujii et al., 2014). This raises the question of whether the association of body rhythm and speech rhythm has a function in language perception and acquisition.

Young infants occasionally show rhythmic body movements that have been described as vegetative reflexes produced by the limbs, torso, and head (Thelen, 1981). While these movements have been suggested to be precursors of more coordinated rhythmic movements in dancing (Thelen, 1981), theories have also highlighted their role as a transitional point to later communication abilities in the first year of life (Iverson and Thelen, 1999). For example, rhythmic movements of the hands and the arms have been linked with the maturation of oral articulators, with an observed decrease in produced rhythmic body movements occurring around the time when infants start producing more vocalizations (Iverson and Thelen, 1999). Whereas very young infants move rhythmically even in absolute silence, these movements are enhanced in social, interactive contexts, when the caregiver enters a room, or when infants are being presented with a toy (Thelen, 1981). This suggests that the development of rhythmic movements may be linked to communicative processes, of which language forms part.

While Thelen (1979, 1981) studies were purely observational, controlled studies have attested that the degree to which infants spontaneously produce rhythmic motor movements seems to depend on specific auditory conditions. For example, Zentner and Eerola (2010) explored the rhythmic motor engagement of 5- to 24-month-old infants under experimentally controlled conditions in a laboratory when listening to different types of music (i.e., isochronous drumbeats and naturalistic music) and speech (i.e., naturalistic adult- and infant-directed French and English). Results revealed that all age groups engaged less rhythmically when listening to speech than when listening to music. Furthermore, the duration of infants' movements in this study was faster to isochronous beats with faster tempi. These results from Finnish and Swiss infants were replicated with Brazilian infants in a study by Ilari (2015), which additionally showed that Brazilian infants tended to produce more movement to music stimuli than the European infants in Zentner and Eerola (2010). This may indicate that the development of spontaneous rhythmic motor responses to auditory signals is influenced by the culture that infants are surrounded by (Note that previous studies also more generally report cross-cultural differences regarding infants' gross motor activities; e.g., Bril and Sabatier, 1986; Victora et al., 1990; Venetsanou and Kambas, 2010, so the question remains whether the cross-cultural differences depend on the auditory context in Zentner and Eerola's and Ilari's studies).

The understanding of infants' spontaneous rhythmic movements to different auditory stimuli was further refined in a study by de l'Etoile et al. (2020), which focused on 6- to 10-month-old infants. Their results indicated that infants' movements were more regular when they were listening to regular beats than when the beat was irregular. Moreover, a longitudinal study by Mazokopaki and Kugiumutzakis (2009) followed infants' development of spontaneous rhythmic vocalizations and movements from 2 to 10 months of age. According to their results, infants produced more vocalizations, hand gestures and dance-like movements when hearing baby songs than when their behavior was observed in silence. However, Fujii and colleagues (Fujii et al., 2014), who explored 3- to 4-month-old infants, did not observe more limb movements when infants listened to pop music than when there was silence but found the presence of music to influence infants' vocal quality. Empirical studies thus suggest that infants manifest an increase in their rhythmic movements to auditory rhythms, and that the quantity of movements is modulated by the type and the register of the auditory rhythms. The present study will extend this work by asking whether the quantity of rhythmic movements also depends on the type of speech rhythms.

Prior research on adult (e.g., Bolton, 1894; Woodrow, 1909, 1911; Hay and Diehl, 2007) and infant listeners (e.g., Yoshida et al., 2010) has established that different acoustic types of speech rhythms lead to differences in rhythmic perception: Streams of sounds alternating in duration (short-long-short-long…) tend to be perceived as iambic (i.e., binary parsings with a weak-strong stress pattern), but streams of sounds alternating in intensity (loud-soft-loud-soft…) or pitch (high-low-high-low…) tend to be perceived as trochaic (i.e., binary parsings with a strong-weak stress pattern). The finding that this rhythmic grouping asymmetry is consistent across speakers of many different languages (Hay and Diehl, 2007; Bhatara et al., 2016; Boll-Avetisyan et al., 2020b) has given rise to proposals that the rhythmic grouping biases are based on an innate domain-general auditory mechanism (Hayes, 1995; Nespor et al., 2008). However, in adults, the magnitude of the effect seems to depend on experience (e.g., language experience: Iversen et al., 2008; Bhatara et al., 2013, 2016; Crowhurst and Olivares, 2014; Langus et al., 2016; music experience: Boll-Avetisyan et al., 2016) and individual skills (i.e., musical aptitude: Boll-Avetisyan et al., 2017, 2020a), indicating individual variability in the reliance of this perceptual mechanism that warrants further investigation, in particular in infants, for whom such data is yet lacking.

Research that addressed the potential functions of this auditory mechanism established that these rhythmic biases influence infants' segmentation of artificial languages into word-like units (Bion et al., 2011; Hay and Saffran, 2012; Abboub et al., 2016). Abboub et al. (2016) (which serves as the basis for the current one) focused on two questions: what acoustic cues infants aged 7.5 month-old used to rhythmically group speech, and how linguistic experience influenced this use. They employed a cross-linguistic comparison of German- and French-learning 7.5-month-old infants to assess whether such rhythmic biases influencing speech segmentation were language-general or language-specific. In their study, infants were first familiarized with artificial language streams composed of syllables that alternated either in pitch, duration, or intensity, while all other rhythmic cues were kept constant, or to a stream that showed no rhythmic alternation (control condition). Following this familiarization, infants were tested on their recognition of syllable pairs. As a result, irrespective of language background, infants showed recognition of syllable pairs if they were familiarized with pitch- or duration-varied streams, but not when they heard intensity-varied or unvaried streams. They concluded that this result may speak for language-general rhythmic biases in young infants, noting that specific cues (here: duration, pitch) might be more accessible than others (here: intensity) for them. The present study aimed to build on Abboub et al. (2016) by investigating infants' spontaneous rhythmic movements and whether the infants' speech segmentation performance in that task was linked to their rhythmic movements.

Past research has revealed that infants' speech segmentation skills are subject to individual variability. Factors that are associated with their speech segmentation performance are, for example, their babbling skills (Hoareau et al., 2019) and their later lexicon size (Newman et al., 2006; Junge et al., 2012), but also environmental factors such as the quantity of infant-directed speech input (Hoareau et al., 2019) and social interactions with their mother (Vanoncini et al., 2022, 2024). In segmentation experiments using the head-turn paradigm, individual differences are often revealed as follows. Infants first hear a speech stream, and in a subsequent test phase, it is probed whether they look longer to an unrelated visual stimulus (a lamp) while hearing familiar or unfamiliar test words. Generally, the direction of their looking preference, that is, whether they look longer when hearing the familiar or the unfamiliar/novel, is deemed irrelevant, as both directions indicate that infants at the group level perceived the difference between the stimuli. However, it has been noted (see Hunter and Ames, 1988 for a model) that infants who are more mature in their development (e.g., older infants) are more likely to express novelty preferences (i.e., they listen longer while hearing unfamiliar test words). Results of studies targeted at identifying predictors of infant speech segmentation performance are in line with this: novelty preferences are exhibited by infants who have more advanced babbling skills (Hoareau et al., 2019) and higher later word knowledge (Singh et al., 2012), reflecting that matureness in language development is associated with novelty preferences. Moreover, novelty preferences are more likely in infants whose mothers show less predictable social gaze behavior (Vanoncini et al., 2024) and whose mothers are more in emotional synchrony with their babies, suggesting that social aspects may also be a source of individual differences. The significant correlations between infants' listening preferences and the probed predictors in these studies also signalize that the strength of infants' preferences, expressed by the magnitude of their listening preferences, is increased in infants with a more mature language development. Following up on this background, we explored infants' rhythmic body movements as another potential source of individual variability in infants' speech segmentation performance.

With the present study, we addressed the question of whether rhythmic motor responses could have a function in language acquisition, following our assumption that infants' multisensory experience of producing (motor) rhythm while perceiving (auditory) rhythm should reinforce the perception of the rhythmic structure of language. We asked two specific research questions: (1) whether infants would show systematic rhythmic movements in the presence of auditory rhythmic cues that guide their speech segmentation and (2) whether infants showing more rhythmic motor engagement would show a more mature (i.e., stronger) speech segmentation performance. To address this, we used video data from Abboub et al. (2016), which included recordings of 7.5-month-olds, and their unpublished data from 9.5-month-olds. We predicted that infants would show rhythmic body movements to the regular occurrence of the rhythmic pattern in the auditory stimuli in these experiments.

We also predicted that infants would show more rhythmic movements when perceiving rhythms cued by pitch and duration (the two conditions prior studies found infants to be sensitive to, showing rhythmic grouping) than when cued by intensity or no prosodic property. Given that our data comprised two age groups (7.5 vs. 9.5 month), and that infants had been exposed to stimuli that were either pronounced as French or German, we additionally asked whether rhythmic movements would depend on these factors, but we did not have any specific predictions with regards to these. Regarding (2), we expected a positive relationship between infants' speech segmentation performance and their quantity of rhythmic body movements, with more rhythmic movements being linked to a more mature speech segmentation performance (matureness expressed by novelty listening preferences and the magnitude of their listening preferences).

2 Materials and methods

2.1 Participants

For the present study, we used Abboub et al. (2016) data of German-learning 7.5-month-olds ($n = 72$; 38 female, mean age: 7.43 months, range: 7.00–8.30; the set originally included $n = 80$, but no videos were available for 8 infants) and added unpublished data of 9.5-month-olds ($n = 76$; 44 female, mean age: 9.47 months, range: 9.03–10.00; no videos were available for 4 additional babies). Table 1 indicates the infants' distribution across conditions. Both of the samples did not include additional infants that were tested but removed due to Abboub et al.'s drop-out criteria. All infants were born full-term, had no reported family risk for language-related developmental disorders, and were growing up in monolingual households. Before the experiment, parents signed informed consent and filled out a demographic questionnaire. Ethics approval was obtained from the ethics committee at University of Potsdam.

TABLE 1 Overview of the number of participants per age and condition.

Condition	N of 7.5 months	N of 9.5 months
Pitch	20	18
Duration	20	18
Intensity	12	20
Control	20	20
Sum	**72**	76

TABLE 2 Acoustic values of the rhythmic cues per condition.

	Condition	Pitch (Hz)	Intensity (dB)	Duration (ms)
Familiarization	Duration	200	70	260–460
	Intensity	200	66–74	360
	Pitch	200–420	70	360
	Control	200	70	360
Test	(All Conditions)	200	70	360

2.2 Stimuli

In Abboub et al. (2016) original study, there were both familiarization and test stimuli. As familiarization stimuli, artificial language streams were used. These were rhythmic speech streams of six syllables consisting of six different vowels and consonants, which were concatenated into syllable streams with a 100 ms pause between each syllable. The speech stimuli were synthesized with the MBROLA software (Dutoit et al., 1996) with both a French (FR4) and a German pronunciation (DE5). The six syllables within a stream always co-occurred in the same fixed order (i.e., /na: zu: gi: pe: fy: ro: na: zu: …/), yielding a transitional probability of 1.0. This fixed order was repeated 66 times, to yield a stream that would last for 3 min. There were four conditions for marking prominence in the syllable streams: in the first three conditions, every second syllable in a stream was stressed (i.e., strong) by pitch, duration, or intensity cues. In the control condition, no syllable was stressed. Syllable durations, pitch and intensity values and ranges were similar to that of previous studies (e.g., Hay and Diehl, 2007; Bion et al., 2011; Bhatara et al., 2013), which were based on acoustic measurements of child-directed speech (for acoustic values, see Table 2). As test stimuli, six syllable pairs were used. Importantly, all test stimuli were flat in prosody, so that infants' recognition of a syllable pair as familiar or novel could only be based on their identification of the phonemic/syllabic information of the test stimulus, and not on its rhythm. Three of these syllable pairs would have occured as strong-weak in the familiarized artificial language, and three that occur as weak-strong; that is, if an infant was familiarized with a stream alternating as /NA zu GI pe FY ro…/, three "strong-weak" test stimuli were /na zu/, /gi pe/ and /fy ro/, while the "weak-strong" test stimuli were /zu gi/, /pe fy/ and /ro na/. For more information about the choice of phonemes for the generation of syllables and a more detailed description of the stimuli, consult Abboub et al. (2016).

2.3 Procedure

In the original study, infants were tested using the head-turn preference procedure (HPP, Kemler Nelson et al., 1995). During the experiment, infants were seated on their caregiver's lap in a soundproof booth. In front of the infant, there was a green light. On both sides of the room, there was a red light located on the same level as the green light. Loudspeakers were hidden under the red lights. Throughout the experiment, the caregiver was wearing headphones that played music to mask the stimuli the infant perceived. Video recordings of the experiment were made to verify the coding of infants' looking times. During the experiment, stimulus presentation was controlled by the experimenter via blinking lights, which, depending on the infant's head movement, were manipulated via button pressing. At the start of the experiment, infant's gaze was centered with the green blinking light. Then, the experimenter would make one of the red lights blink to attract infant's attention to it. When the infant looked at the blinking red light, the auditory stimulus would start.

During familiarization, infants listened to an artificial language stream with one of the four acoustic manipulations (pitch/duration/intensity/control). The sound came from both loudspeakers and lasted for 3 minutes. The blinking of the red light would stop if the infant turned away for more than 2 s, and in this case, the green light would begin blinking again (but the sound was never stopped during this phase). In the following test phase, which was the same for all infants regardless of familiarization conditions, segmentation was tested with two test trial types: "strong-weak" trials versus "weak-strong" trials. In sum, there were 12 test trials, half of which were "strong-weak" trials, and half of which were "weak-strong" trials. Test trials were constructed on the basis of the 6 syllable pair test stimuli (see 2.2), each containing 16 repetitions of one test stimulus (e.g., nazu nazu nazu). In order to arrive at 12 test trials, each of the 6 test trials was presented twice; once in trial position 1–6 (block 1), and once in trial position 7–12 (block 2). The order of test trials within a block was randomized across infants. Whether the stimuli were German- or French-sounding was counterbalanced (in our data: French-sounding: $n = 78$, German-sounding: $n = 70$). The test trials came randomly from either the right or the left side, in association with the blinking lamp on the same side. The blinking and the stimulus would stop if the infant turned away for more than 2 s, and in this case, the green light would begin blinking. All stimuli were played at comfortable volume (*ca.* 65 dB). In total, the experiment maximally took 6 min. Summed looking times per test trial (i.e., looking times per trial excluding the intervals during which the infant turned the head away from the side lamp) were taken to indicate infants' interest in a trial. Following the standard HPP procedure, at the group level, the difference in looking time between the average of the three "strong-weak" and the average of the three "weak-strong" test trials can be taken to indicate that infants have recognized the difference between the test trial types, namely that one test trial type was familiar to them (i.e., the syllable pairs following from their grouping of the continuous artificial language stream, e.g., /na zu/ when familiarized with a /NA zu GI pe …/ stream, if they perceived a strong-weak grouping), and half of them was novel to them (i.e., the syllable pairs that would not follow from their perceived grouping; in this case, e.g., /zu gi/).

2.4 Data preprocessing and analysis

Video recordings were annotated for infants' rhythmic body movements. Manual annotations were performed by three independent coders using the free annotation software ELAN (2016). For each video, coders identified and marked all intervals of rhythmic movements during the 3-min familiarization phase. Rhythmic movements were defined as comprising a minimum of three immediate repetitions of a bodily movement (Thelen, 1979) of any body part (e.g., limb, hand, legs). To test the reliability of rhythmic movement locations, an independent tester performed the inter-rater reliability check by calculating Cohen's Kappa for a random subset of approximately 7% of the cases (11 cases, including six 7.5- and five 9.5-month-olds). The kappa value ranged from 0.72 to 0.86 for two different rater pairs, suggesting moderate to strong agreement (McHugh, 2012).

To determine the predictors of rhythmic movements, an analysis of variance (ANOVA) was run. As a dependent variable, we used infants' total rhythmic moving times defined as the sum of all rhythmic intervals in milliseconds. Condition (pitch, intensity, duration or control), Age (7.5 or 9.5 months old) and Pronunciation (French or German-sounding) were between-participant fixed factors. Results are reported in section 3.1.

To investigate whether infants' rhythmic motor engagement correlated with their performance in the speech segmentation task, we analyzed a subset of infants that exhibited rhythmic movements during the experiment ($n = 62$). We performed two correlation analyses, both of which used infants' total rhythmic moving times (i.e., the sum of all intervals infants produced rhythmic movements while listening to the artificial language) as one variable, and both of which based the second variable on infants' summed looking time per trial, with the following difference: For the *first* correlation analysis, we generated a difference (Δ) score of the infants' looking times at test by subtracting the average of the looking times to "strong-weak" test trials from the average of the looking times to "weak-strong" test trials. Hence, positive Δ score values reflect longer looking times for "weak-strong" test trials, and negative Δ score values reflect longer looking times for "strong-weak" test trials. Similar difference scores were also used in Jusczyk and Aslin (1995), Singh et al. (2012), and Hoareau et al. (2019).

The *second* correlation analysis was motivated as follows: Since in the original study, the expected grouping depended on the acoustic cue (i.e., strong-weak grouping with pitch and intensity; weak-strong grouping with duration), pooled looking-time data may be difficult to interpret. However, because of the small sample size, it was impossible to further sub-set the data by the four conditions. Hence, we did the following: for the second correlational analysis, we changed the looking time Δ scores between trial types to *absolute* values by removing the positive/negative sign: all Δ scores were turned into positive values. Consequently, these absolute looking-time scores would reflect the magnitude of a preference, with higher values indicating more mature (i.e., stronger) listening preferences, independent of the direction of the preference. In both cases, this Δ score was based on looking times during the first six trials, following Abboub et al. (2016), who found that differences in listening preferences at test only occurred during the first of the two blocks in their study. The corresponding results are reported in section 3.2.

All statistical analyses and visualizations were done in R (version 4.3.2., R Core Team, 2021). The ANOVAs were performed using the package ez (version 4.4–0, Lawrence, 2016), the plots were based on the package ggplot2 (Wickham, 2016). The significance criterion was set to $p < 0.05$, but given the exploratory nature of the present study, we decided to report effects of $p < 0.1$ as marginally significant, as they may be insightful for future studies.

3 Results

3.1 Rhythmic body movements while listening to the speech stream

Overall, 42% of the babies in our sample occasionally produced rhythmic movements while listening to speech stimuli. Results of the statistical analysis (see Figure 1) revealed a significant main effect of condition ($F(3, 131) = 2.84$, $p < 0.05$, with a small effect size of $\eta^2 = 0.06$), and a significant Condition * Pronunciation interaction ($F(3, 131) = 3.02$, $p < 0.05$, with a moderate effect size of $\eta^2 = 0.07$). There were no main effects of Age ($F(1, 131) = 0.01$, $p = 0.92$) or Pronunciation ($F(1, 131) = 2.22$, $p = 0.14$), no Condition * Age ($F(3, 131) = 0.29$, $p = 0.83$), and no Condition * Age * Pronunciation interaction ($F(3, 131) = 0.68$, $p = 0.56$). The Age * Pronunciation

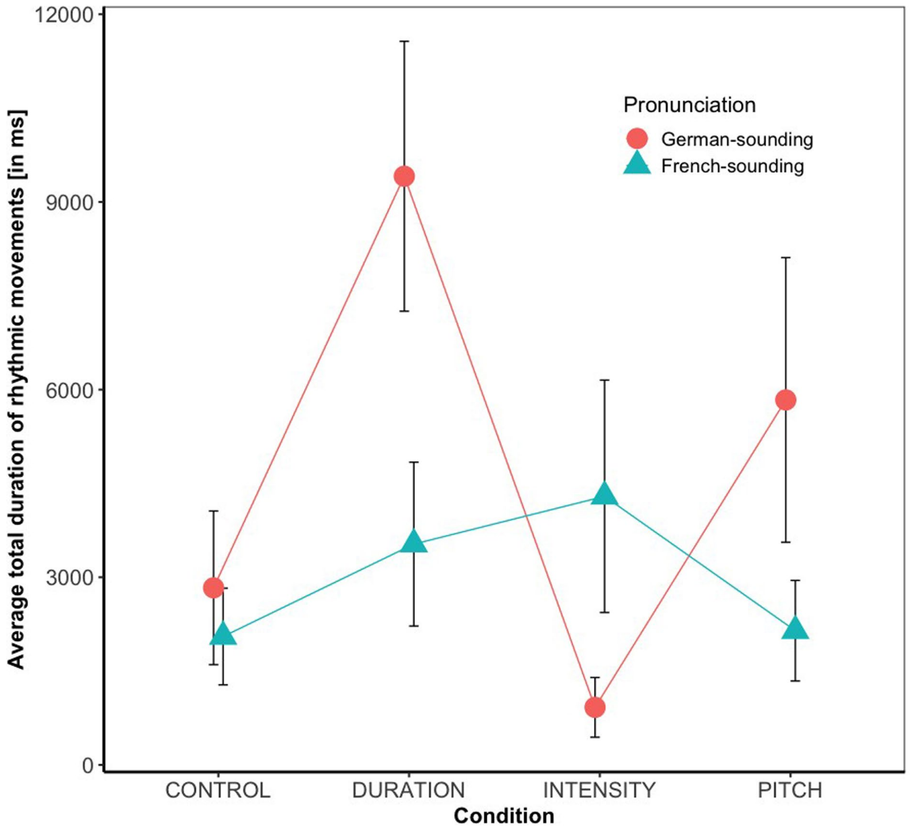


FIGURE 1
Average rhythmic body moving times (in ms) and their standard errors split by condition and pronunciation.

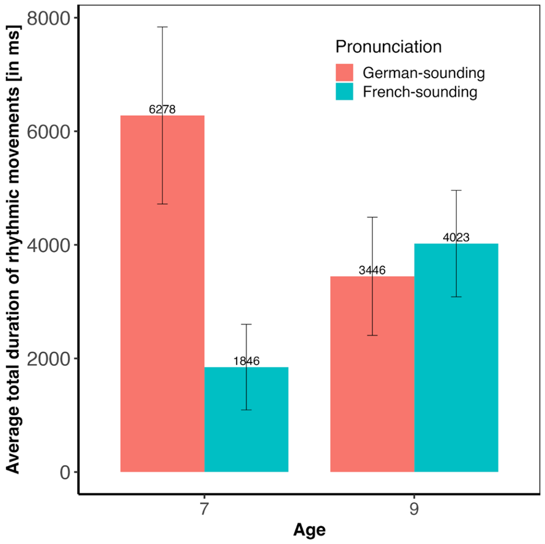


FIGURE 2
Average rhythmic body moving times (in ms) and their standard errors split by age and pronunciation.

interaction ($F(1, 131) = 2.91$, $p = 0.09$) was marginally significant. Planned post-hoc tests revealed that the significant main effect of condition was due to significant differences between Duration versus Intensity ($p < 0.05$) and Duration versus Control ($p < 0.05$). An exploration of the Condition * Pronunciation interaction showed that the effect of condition was only significant with the German- ($F(3, 65) = 4.43$, $p < 0.01$) but not with the French-sounding pronunciation ($F(3, 71) = 0.70$, $p = 0.55$). Post-hoc pairwise comparisons within the subset of infants that had listened to the German pronunciation revealed significantly longer total moving times in the Duration compared to the Intensity ($p < 0.01$) and to the Control condition ($p < 0.01$), and marginally significantly longer moving times in the Pitch compared to the Intensity condition ($p = 0.07$). No other comparison reached significance. As the Age * Pronunciation did not reach significance, no further post-hoc comparisons were conducted, but Figure 2 suggests that the trend was that 7-month-olds moved more when listening to German- than when listening to French-sounding streams, while neither of the two pronunciations elicited more or less movements in 9-month-olds.

3.2 Correlations of rhythmic body movements and speech segmentation performance

Both correlation analyses indicate an association between infants' rhythmic movements while listening to the artificial language and their speech segmentation performance. The first correlation analysis revealed significantly longer rhythmic moving times with more negative looking time Δ scores ($r = -0.26$, $p = 0.039$), that is, when hearing "strong-weak" test trials (see Figure 3A). The second

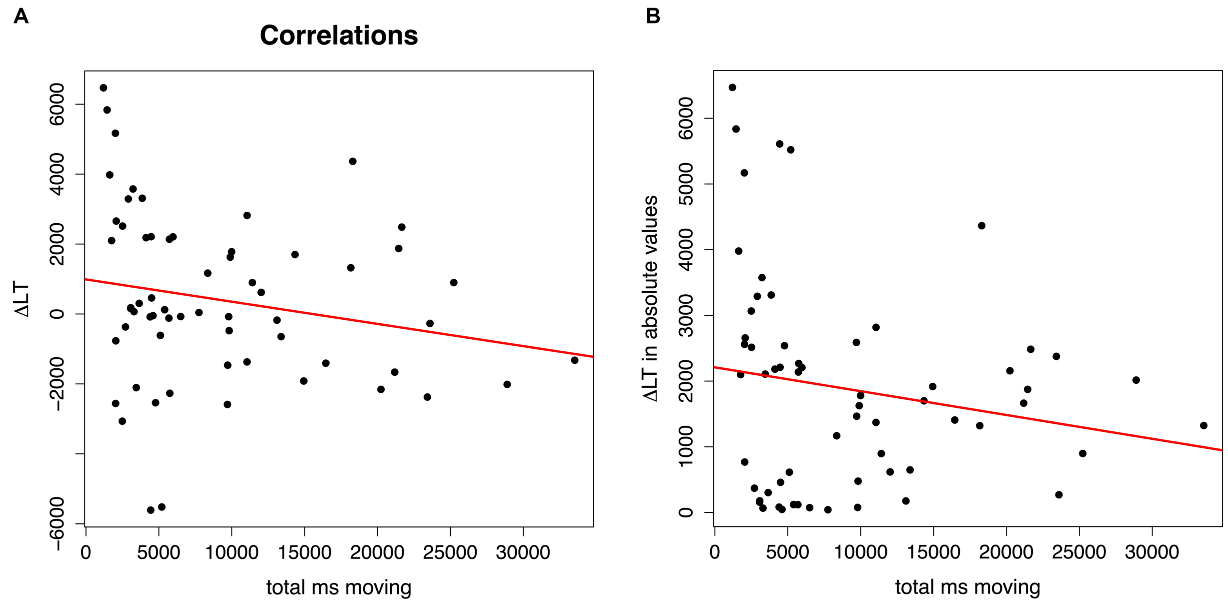


FIGURE 3
Plots of the correlations of infants' total moving times (in ms) while listening to the artificial speech streams (x-axis) and their speech segmentation performance at test. In **(A)**, y-axis = Δ in looking time (Δ LT) between "strong-weak" or "weak-strong" trials in real numbers, with positive Δ LT scores reflecting longer looking times for test trials with syllable pairs that had occurred as weak-strong, and negative Δ LT scores for test trials with syllable pairs that had occurred as strong-weak during familiarization; in **(B)**, y-axis = Δ in looking time (Δ LT) in absolute values between "strong-weak" and "weak-strong" trials.

correlation analysis revealed a marginally significant correlation: the longer were the infants' rhythmic moving times, the smaller were the absolute looking time Δ scores ($r = -0.24$, $p = 0.059$), that is, the weaker were their preferences for looking longer toward one of the two types of trials at test (see Figure 3B). Given the exploratory nature of the present study, and this p-value of 0.059 being very close to the significance criterion of $p < 0.05$, we will discuss this marginal effect as well as all significant effects below.

4 Discussion

The present study sought to explore infants' rhythmic motor engagement while listening to rhythmic speech, asking whether infants' rhythmic motor responses to rhythmic speech have a functional link with their language processing abilities. For this purpose, we analyzed 7.5- and 9.5-month-old German-learning infants' rhythmic body movements to German- and French-sounding artificial language streams. We reasoned that rhythmic motor engagement may enhance infants' perception of speech rhythm, which may be reflected in enhanced rhythmic motor responses to specific auditory rhythmic cues, namely pitch and duration. Moreover, we hypothesized that individual differences in infants' early rhythmic engagement may be associated with their word segmentation abilities. Our analysis of infants' reactions to the speech streams revealed that many infants moved rhythmically across conditions, but those who listened to duration-varied streams showed most rhythmic engagement. Moreover, the difference in rhythmic moving times between conditions was only present when infants listened to native, German-sounding language streams, but not when listening to non-native, French-sounding streams. Lastly, as expected, we found an association between infants' rhythmic movements and their speech segmentation performance; but unexpectedly, those who moved less demonstrated more mature segmentation skills. We discuss these results in detail below.

The comparison of rhythmic moving times between the different acoustic conditions revealed that infants showed most rhythmic engagement when perceiving duration-based rhythm in speech. More specifically, their rhythmic moving times were longer in the duration than in both the intensity and control condition, and no statistical difference in their moving times was found when comparing the duration and pitch condition. This result was in line with our expectation, as it complements the original findings by Abboub et al. (2016), in which 7.5-month-olds did not show rhythm-based speech segmentation in the intensity and control condition, while they did in the duration condition. However, since Abboub et al. also found infants to show rhythm-based speech segmentation in the pitch condition, it was unexpected to find no statistical evidence for (or against) enhanced rhythmic movements in this condition, too. (However, as can be seen in Figure 1, there was a marginal effect of longer moving times in the pitch than in the intensity condition when stimuli were German-sounding, so it is possible that the lack of a significant effect is due to low power). In any case, the results suggest that infants' body movements are indicative of their sensitivity to the auditory duration cues.

Notably, the above-described differences in rhythmic moving times between acoustic conditions were only attested by their reactions to speech synthesized from German but not French phonemes. This effect is in contrast with the behavioral results reported in Abboub et al. (2016) original study, where no effect of nativeness of pronunciation on infants' speech segmentation was observed, which led the authors to suggest that language-specific perception may emerge later in development. In turn, the present findings indicate that German-learning infants are sensitive to native versus non-native pronunciation differences: seemingly, they recognize the non-nativeness of French-sounding speech, and find the native German-sounding speech more engaging as manifested in their rhythmic body response. This exploratory result can also be taken as evidence that the infants must have processed the synthesized artificial speech streams as speech-like. Since the presented stimuli were synthesized with a text-to-speech software (Dutoit et al., 1996) that drew on recordings of phonemes from a German (DE5) and a French (FR4) speaker, phoneme-level information was the only language-specific cue, as all prosodic information was later superimposed onto the synthesized stimuli. Hence, infants' rhythmic motor engagement must have depended on their identification of the phoneme-level information as native-sounding.

Regarding the link between infants' rhythmic motor responses to speech and their segmentation abilities, the correlational analyses revealed some interesting patterns. The first correlation analysis, which used infants' Δ score in looking time in "strong-weak" versus "weak-strong" test trials attested that the longer infants had moved while listening to the language streams, the more they preferred listening to "strong-weak" trials at test, that is, to trials that followed from a trochaic grouping. It is possible that this result can be explained as a consequence of infants' knowledge of the German prosodic system. Linguistically, German is described as a trochaic language (Wiese, 1996), and infants predominantly receive words with a strong-weak (trochaic) stress pattern in their input (Stärk et al., 2022). Experimental evidence suggests that already young German-learning infants have knowledge of their language's predominant stress pattern, as indicated by their preferences to listen longer to trochaic than to iambic patterns in preferential looking paradigms between the ages of 4 and 6 months (Höhle et al., 2009, 2014; Marimon and Höhle, 2022). We may, hence, speculate that infants who produced more rhythmic movements to the rhythm of the speech streams found it more difficult to disengage from the perceived trochaic rhythm pattern and, hence, showed a trochaic listening preference at test, while infants with less motor engagement would more readily succeed at disengaging from the trochaic pattern and display instead a novelty preference for iambs at test. Since novelty preferences are generally interpreted to reflect more mature language processing skills (Hunter and Ames, 1988), this result may suggest that infants who showed less rhythmic engagement had more mature language segmentation abilities.

The second correlation analysis resulted in a marginally significant association between infants' rhythmic moving times and the Δ score in looking time turned into absolute values. This result, which we need to interpret with caution as it did not fully reach our significance criterion, draws a link between infants' rhythmic motor responses and the strength of a segmentation preference (without considering whether the preference was for trials reflecting a specific grouping, a familiarity or novelty preference). However, the direction of this association was unexpected: the longer infants had moved while listening to the language streams, the less strong was their preference. Like the results of the first correlation analysis, this second correlation also suggests that segmentation performance was lower when infants were more rhythmically engaged with the perceived rhythm. This raises the question of whether rhythmic motor engagement with speech rhythms has an inhibiting effect on language acquisition. Two explanations for the found association are possible: First, it may be that infants with more rhythmic movements were simply less attentive during the task. A second possibility is that the results reflect individual differences in whether the infants paid more attention to the rhythm of the auditory signal or to the specific syllable orders (i.e., phoneme-level information), with the former enhancing rhythmic motor responses and the latter enhancing segmentation performance. This is plausible, as studies with adults report that listeners can switch between focusing on musical or linguistic information when listening to music (e.g., Schön et al., 2005) or speech, and that whether one focuses more on the musical or linguistic properties is subject to individual variation (e.g., Rathcke et al., 2021). This interpretation is also in line with results of a recent study (Marimon et al., 2022) that probed whether German-learning 9-month-old infants rely on prosodic or syllable co-occurrence cues for segmenting an artificial language stream if the two cues are in conflict. Infants who relied more on prosody had stronger grammar skills at 3 years, whereas infants who relied more on syllable co-occurrences had larger vocabulary outcomes. This may suggest that, at least for infants between 7 and 9 months of age, focusing attention on rhythmic information rather than phoneme-level information may be a sign of a less mature response that is related to both a reduced ability to segment words from speech (the present finding) and a reduced ability to build up a large vocabulary (Marimon et al., 2022), which is interesting in light of proposals that vocabulary acquisition depends on speech segmentation (e.g., Jusczyk, 1997; Junge et al., 2012). Overall, the results suggest that infants' general auditory preferences as well as differences in novelty and familiarity preferences typically observed in looking-time paradigms (e.g., Bergmann and Cristia, 2016; Gasparini et al., 2021) may not only be driven by stimulus familiarity or novelty, task complexity, or the infant age (Hunter and Ames, 1988), but may also result from bodily engagement with the rhythm of auditory stimuli. It would be ideal to replicate this study and to clarify whether infants who produce less rhythmic movements are indeed more mature in their language development, for example, by gathering information on their acquisition milestones (e.g., on their babbling, see Hoareau et al., 2019).

This finding of the individual differences raises the question of whether infants in our study focused either *only* on rhythm or only on phoneme-level information. However, this is highly unlikely: first, a strong preference for one test trial type over another at test (i.e., in infants that may have focused on phoneme-level information) can only occur if infants have perceived the rhythm cue that would bias them toward perceiving a grouping of syllables into pairs. Second, for the infants who moved more (i.e., infants that may have focused on rhythm), we found that their quantity of rhythmic movements was higher when they listened to native rather than non-native language streams, and since nativeness was related to the acoustic properties of phoneme-level information, it is not possible to conclude that these infants ignored one of the levels (phonemic or rhythmic) either. Hence, it is more likely that at the ages between 7 and 9 months, infants perceive both rhythm and phoneme-level information, with individual differences leading some infants to focus more on rhythm and others on syllable co-occurrence characteristics in the language streams.

The present study shows for the first time that infants learning the same language demonstrate individual differences in rhythmic grouping. Prior research by Yoshida and colleagues (Yoshida et al., 2010) had already established that infants' rhythmic grouping is influenced by language experience: when exposed to streams of sounds alternating in duration, English-learning 7-8-month-olds favored a weak-strong (iambic) grouping, while Japanese-learning infants favored a strong-weak (trochaic) grouping at the same age. The authors related this cross-linguistic difference to infants' experience with different word orders, with English showing many phrases with head-complement order (i.e., short function words before long content words), and Japanese showing many phrases with complement-head order (i.e., long content words before short function words). Differences within a group of individuals with the same native language had, so far, only been attested for German-speaking adults, who showed more consistent rhythmic grouping if they had higher musical rhythm perception acuity (Boll-Avetisyan et al., 2017, 2020a). Hence, the present research adds further evidence to the body of research indicating individual variability in infants' speech perception and processing, which may systematically relate to internal or external factors beyond their language experience.

Our study is one of the first behavioral studies to tie early motor engagement with speech rhythm and language acquisition. Whereas spontaneous rhythmic movements in infants cannot be used as a reliable measure of tracking the auditory signal (Mandke and Rocha, 2023), our results suggest that they may still be indicative of the underlying speech perception processes. For example, while previous research has shown that infants' passive sensorimotor experience (i.e., being moved to certain beats) reinforces their rhythm perception (Phillips-Silver and Trainor, 2005), our evidence contributes to it by demonstrating that infants' active sensorimotor engagement may also play a role in rhythm processing. Moreover, according to our findings, rhythmic engagement in infancy may indeed modulate speech processing, e.g., in speech segmentation tasks. However, since the present results indicate reduced rather than enhanced speech processing in rhythmically moving infants, our findings add a new perspective to the body of research on infants' improved sensitivity to auditory rhythms in multimodal contexts (Bahrick and Lickliter, 2000; Lewkowicz, 2000).

Previous accounts have suggested that infants' enhanced rhythmic movements may be related to the positive affect that infants experience when listening to music (Zentner and Eerola, 2010), or that it could be a precursor of later entrainment abilities (Provasi et al., 2014; de l'Etoile et al., 2020), broadening the role of such engagement beyond mere manifestations of infants' arousal. It has also been suggested that rhythmic entrainment may play a functional role in language acquisition. For example, educators have suggested that rhythmic movement with nursery rhymes and songs can help children to understand rhymes in language (e.g., Berger Cardany, 2013). The present results, however, do not provide support for this possibility. Instead, they suggest that for 7 to 9 month old infants motor engagement with auditory rhythms may actually hinder them from perceiving rhythmic grouping in spoken language – an important aspect in speech processing. We need to emphasize that the present results are correlational and based on exploratory analyses; hence this effect clearly requires replication before any conclusions can be drawn regarding the functional role of infants' rhythmic body movements to auditory rhythms. Future studies will need to further explore the conditions that might enhance infants' rhythmic movements, and how these associate with their language processing and language abilities. In order to get a better understanding of the causal link between rhythmic motor movements and rhythmic speech perception in language development, it might also be interesting to investigate the effects of hindering infants from moving their bodies on their speech processing (similar to Bruderer et al. (2015) and Yeung and Werker (2013), who showed that infants' phoneme perception ability was affected, if they sucked on toys or pacifiers that constrained their lip movements).

A limitation of the present study is that it was not specifically designed to investigate infants' body movements to speech rhythm. Consequently, the original design with numerous conditions and age groups was not ideal for the purpose. For example, we were not able to run correlation analyses that would probe the association of rhythmic motor responses and speech segmentation performance for each age group and each condition separately due to the limited observations after exclusion of the non-moving infants. Future studies should build on this exploration with study designs that are well-suited for establishing such correlations. Moreover, with the pre-existing data we had at hand, it was not possible to establish a baseline for probing whether infants who move more in silence also show more rhythmic movements while listening to speech. While not all prior studies used such baselines (Zentner and Eerola, 2010; Ilari, 2015; Nguyen et al., 2023), some indeed used them (Fujii et al., 2014; de l'Etoile et al., 2020), and future studies should consider collecting such data. Furthermore, the scope of this research could also be extended to other age groups in future studies. For example, it could be interesting to study whether rhythmic engagement with speech rhythm is more relevant for auditory processing in younger, pre-lexical age groups and whether rhythmic engagement with speech changes in maturation. Another direction for future research could be to examine whether and how rhythmic engagement to speech is modulated by infants' experience with music and musical rhythms. Music experience in infancy has been previously shown to influence language acquisition (Langus et al., 2023; for a review, see Nayak et al., 2022), and music and speech rhythm have been suggested to share a number of properties aiding language processing (Fiveash et al., 2021). Lastly, it would be insightful to obtain more sensitive measures of infants' rhythmic movements in the future by employing motion capture systems or even electromyography, to measure subtle muscle responses.

In sum, the present study provides the first exploratory evidence that infants' spontaneous rhythmic body movements while listening to rhythmic speech are systematic. We observe that infants' motor engagement is highest when they hear duration cues to rhythm. Moreover, they produce more rhythmic movements to native than to non-native speech. Lastly, individual differences in the quantity of infants' rhythmic motor responses is associated with their speech segmentation ability, but other than expected, infants who showed less rhythmic movements showed more mature and stronger performance in speech segmentation. Future studies should further investigate how rhythmic body movements to speech rhythm may interact with language acquisition.

Data availability statement

The pre-processed data supporting the conclusions of this article will be made available by the authors upon request.

Ethics statement

The studies involving humans were approved by the ethics committee of the University of Potsdam. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

NB-A: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. AS: Writing – review & editing, Writing – original draft, Validation, Methodology. AL: Writing – review & editing, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The publication of this article was funded by the Publication Fund of the University of Potsdam.

Acknowledgments

We thank Olivia Malotka, Stefanie Meister and Anne Triebs for assistance with coding the video data.

Conflict of interest

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

Publisher's note

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

References

Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., and Nazzi, T. (2016). Rhythmic grouping according to the iambic-trochaic law in French- and German-learning infants. *Front. Hum. Neurosci.* 10:292. doi: 10.3389/fnhum.2016.00292

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.* 9, 321–324. doi: 10.1111/1467-9280.00063

Bahrick, L. E., and Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev. Psychol.* 36, 190–201. doi: 10.1037/0012-1649.36.2.190

Berger Cardany, A. (2013). Nursery rhymes in music and language literacy. *Gen. Music Today* 26, 30–36. doi: 10.1177/1048371312462869

Bergmann, C., and Cristia, A. (2016). Development of infants' segmentation of words from native speech: a meta-analytic approach. *Dev. Sci.* 19, 901–917. doi: 10.1111/desc.12341

Bhatara, A., Boll-Avetisyan, N., Agus, T., Höhle, B., and Nazzi, T. (2016). Language experience affects grouping of musical instrument sounds. *Cogn. Sci.* 40, 1816–1830. doi: 10.1111/cogs.12300

Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., and Höhle, B. (2013). Native language affects rhythmic grouping of speech. *J. Acoust. Soc. Am.* 134, 3828–3843. doi: 10.1121/1.4823848

Bion, R. A., Benavides-Varela, S., and Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Lang. Speech* 54, 123–140. doi: 10.1177/0023830910388018

Boll-Avetisyan, N., Bhatara, A., and Höhle, B. (2017). Effects of musicality on the perception of rhythmic structure in speech. *Lab. Phonol.* 8:9. doi: 10.5334/labphon.91

Boll-Avetisyan, N., Bhatara, A., and Höhle, B. (2020a). Processing of rhythm in speech and music in adult dyslexia. *Brain Sci.* 10:261. doi: 10.3390/brainsci10050261

Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., and Höhle, B. (2016). Effects of experience with L2 and music on rhythmic grouping by French listeners. *Biling. Lang. Congn.* 19, 971–986. doi: 10.1017/S1366728915000425

Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., and Höhle, B. (2020b). Rhythmic grouping biases in simultaneous bilinguals. *Biling. Lang. Congn.* 23, 1070–1081. doi: 10.1017/S1366728920000140

Bolton, T. L. (1894). Rhythm. *Am. J. Psychol.* 6, 145–238. doi: 10.2307/1410948

Bril, B., and Sabatier, C. (1986). The cultural context of motor development: postural manipulations in the daily life of Bambara babies (Mali). *Int. J. Behav. Dev.* 9, 439–453. doi: 10.1177/016502548600900403

Brown, S., Martinez, M. J., and Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *Neuroreport* 15, 2033–2037. doi: 10.1097/00001756-200409150-00008

Bruderer, A. G., Danielson, D. K., Kandhadai, P., and Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proc. Natl. Acad. Sci.* 112, 13531–13536. doi: 10.1073/pnas.1508631112

Christophe, A., Dupoux, E., Bertoncini, J., and Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *J. Acoust. Soc. Am.* 95, 1570–1580. doi: 10.1121/1.408544

Condon, W. S., and Sander, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science* 183, 99–101. doi: 10.1126/science.183.4120.99

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., and Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Dev.* 85, 1330–1345. doi: 10.1111/cdev.12193

Crowhurst, M. J., and Olivares, A. T. (2014). Beyond the iambic-trochaic law: the joint influence of duration and intensity on the perception of rhythmic speech. *Phonology* 31, 51–94. doi: 10.1017/S0952675714000037

Cummins, F. (2009). Rhythm as entrainment: the case of synchronous speech. *J. Phon.* 37, 16–28. doi: 10.1016/j.wocn.2008.08.003

de l'Etoile, S. K., Bennett, C., and Zopluoglu, C. (2020). Infant movement response to auditory rhythm. *Percept. Mot. Skills* 127, 651–670. doi: 10.1177/0031512520922642

De Ruiter, J. P. (1998). *Gesture and speech production*. Nijmegen: Katholieke Universiteit.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., and Van der Vrecken, O. (1996). The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes. In Proceeding of fourth international conference on spoken language processing. Philadelphia, PA: IEEE.

Echols, C. H., Crowhurst, M. J., and Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *J. Mem. Lang.* 36, 202–225. doi: 10.1006/jmla.1996.2483

ELAN. (Version 4.9.4) [Computer software] (2016). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Available at: https://archive.mpi.nl/tla/elan

Esteve-Gibert, N., and Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Comm.* 57, 301–316. doi: 10.1016/j.specom.2013.06.006

Fiveash, A., Bedoin, N., Gordon, R. L., and Tillmann, B. (2021). Processing rhythm in speech and music: shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology* 35, 771–791. doi: 10.1037/neu0000766

Friederici, A. D., Friedrich, M., and Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Curr. Biol.* 17, 1208–1211. doi: 10.1016/j.cub.2007.06.011

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

Gasparini, L., Langus, A., Tsuji, S., and Boll-Avetisyan, N. (2021). Quantifying the role of rhythm in infants' language discrimination abilities: a meta-analysis. *Cognition* 213:104757. doi: 10.1016/j.cognition.2021.104757

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

Guellai, B., Langus, A., and Nespor, M. (2014). Prosody in the hands of the speaker. *Front. Psychol.* 5:700. doi: 10.3389/fpsyg.2014.00700

Hay, J. S., and Diehl, R. L. (2007). Perception of rhythmic grouping: testing the iambic/trochaic law. *Percept. Psychophys.* 69, 113–122. doi: 10.3758/BF03194458

Hay, J. F., and Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy* 17, 610–641. doi: 10.1111/j.1532-7078.2011.00110.x

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., and Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition* 26, 269–286. doi: 10.1016/S0010-0277(87)80002-1

Hoareau, M., Yeung, H. H., and Nazzi, T. (2019). Infants' statistical word segmentation in an artificial language is linked to both parental speech input and reported production abilities. *Dev. Sci.* 22:e12803. doi: 10.1111/desc.12803

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

Höhle, B., Pauen, S., Hesse, V., and Weissenborn, J. (2014). Discrimination of rhythmic pattern at 4 months and language performance at 5 years: a longitudinal analysis of data from German-learning children. *Lang. Learn.* 64, 141–164. doi: 10.1111/lang.12075

Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res.* 5, 69–95.

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

Iversen, J. R., Patel, A. D., and Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* 124, 2263–2271. doi: 10.1121/1.2973189

Iverson, J. M., and Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *J. Conscious. Stud.* 6, 19–40.

Johnson, E. K., and Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Dev. Sci.* 13, 339–345. doi: 10.1111/j.1467-7687.2009.00886.x

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* 15, 463–473. doi: 10.1111/j.1467-7687.2012.1144.x

Jusczyk, P. W. (1997). Finding and remembering words: some beginnings by English-learning infants. *Curr. Dir. Psychol. Sci.* 6, 170–174. doi: 10.1111/1467-8721.ep10772947

Jusczyk, P. W., and Aslin, R. N. (1995). Infants′ detection of the sound patterns of words in fluent speech. *Cogn. Psychol.* 29, 1–23. doi: 10.1006/cogp.1995.1010

Jusczyk, P. W., Houston, D. M., and Newsome, M. (1999). The beginnings of word-segmentation in English-learning infants. *Cogn. Psychol.* 39, 159–207. doi: 10.1006/cogp.1999.0716

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behav. Dev.* 18, 111–116. doi: 10.1016/0163-6383(95)90012-8

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

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Langus, A., Boll-Avetisyan, N., van Ommen, S., and Nazzi, T. (2023). Music and language in the crib: early cross-domain effects of experience on categorical perception of prominence in spoken language. *Dev. Sci.* 26:13383. doi: 10.1111/desc.13383

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

Langus, A., Seyed-Allaei, S., Uysal, E., Pirmoradian, S., Marino, C., Asaadi, S., et al. (2016). Listening natively across perceptual domains. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 1127–1139. doi: 10.1037/xlm0000226

Lawrence, M. A. (2016). Ez: easy analysis and visualization of factorial experiments. Available at: https://CRAN.R-project.org/package=ez

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol. Bull.* 126, 281–308. doi: 10.1037/0033-2909.126.2.281

Lew-Williams, C., and Saffran, J. R. (2012). All words are not created equal: expectations about word length guide infant statistical learning. *Cognition* 122, 241–246. doi: 10.1016/j.cognition.2011.10.007

Mandke, K., and Rocha, S. (2023). "Neural and behavioural rhythmic tracking during language acquisition: the story so far" in *Rhythms of speech and language*. eds. L. Meyer and A. Strauss (Cambridge: Cambridge University Press).

Marimon, M., and Höhle, B. (2022). Testing prosodic development with the Headturn preference procedure: a test-retest reliability study. *Infant Child Dev.* 31:e2362. doi: 10.1002/icd.2362

Marimon, M., Höhle, B., and Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition* 224:105054. doi: 10.1016/j.cognition.2022.105054

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three-to four-month-old Japanese infants. *J. Child Lang.* 20, 303–312. doi: 10.1017/S0305000900008291

Mazokopaki, K., and Kugiumutzakis, G. (2009). "Infant rhythms: expressions of musical companionship" in *Communicative musicality: Exploring the basis of human companionship*. eds. S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 185–208.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282. doi: 10.11613/BM.2012.031

Mersad, K., and Nazzi, T. (2012). When mommy comes to the rescue of statistics: infants combine top-down and bottom-up cues to segment speech. *Lang. Learn. Dev.* 8, 303–315. doi: 10.1080/15475441.2011.609106

Mintz, T. H., Walker, R. L., Welday, A., and Kidd, C. (2018). Infants' sensitivity to vowel harmony and its role in segmenting speech. *Cognition* 171, 95–107. doi: 10.1016/j.cognition.2017.10.020

Mundy-Castle, A. (1980). "Perception and communication in infancy: a cross-cultural study" in *The social foundations of language and thought: essays in honor of J.S. Bruner*. ed. D. Olson (New York: Norton), 231–253.

Nayak, S., Coleman, P. L., Ladányi, E., Nitin, R., Gustavson, D. E., Fisher, S. E., et al. (2022). The musical abilities, pleiotropy, language, and environment (MAPLE) framework for understanding musicality-language links across the lifespan. *Neurobiol. Lang.* 3, 615–664. doi: 10.1162/nol_a_00079

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 756–766. doi: 10.1037//0096-1523.24.3.756

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

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., and Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue Linguaggio* 7, 139–168. doi: 10.1418/28093

Newman, R., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., and Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Dev. Psychol.* 42, 643–655. doi: 10.1037/0012-1649.42.4.643

Nguyen, T., Reisner, S., Lueger, A., Wass, S. V., Hoehl, S., and Markova, G. (2023). Sing to me, baby: infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Dev. Cogn. Neurosci.* 64:101313. doi: 10.1016/j.dcn.2023.101313

Papoušek, M. (1992). "Early ontogeny of vocal communication in parent–infant interactions" in *Nonverbal vocal communication: comparative and developmental approaches*. eds. H. Papoušek, U. Jürgens and M. Papoušek (Cambridge: Cambridge University Press), 230–261.

Papoušek, M., Papoušek, H., and Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behav. Dev.* 14, 415–440. doi: 10.1016/0163-6383(91)90031-M

Phillips-Silver, J., and Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science* 308:1430. doi: 10.1126/science.1110922

Provasi, J., Anderson, D. I., and Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Front. Psychol.* 5:1048. doi: 10.3389/fpsyg.2014.01048

R Core Team (2021). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rathcke, T., Falk, S., and Dalla Bella, S. (2021). Music to your ears: sentence sonority and listener background modulate the "speech-to-song illusion". *Music Percept.* 38, 499–508. doi: 10.1525/mp.2021.38.5.499

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Schmidt-Kassow, M., and Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Res.* 1226, 144–155. doi: 10.1016/j.brainres.2008.06.017

Schön, D., Gordon, R. L., and Besson, M. (2005). Musical and linguistic processing in song perception. *Ann. N. Y. Acad. Sci.* 1060, 71–81. doi: 10.1196/annals.1360.006

Singh, L., Steven Reznick, J., and Xuehua, L. (2012). Infant word segmentation and childhood vocabulary development: a longitudinal analysis. *Dev. Sci.* 15, 482–495. doi: 10.1111/j.1467-7687.2012.01141.x

Stärk, K., Kidd, E., and Frost, R. L. (2022). Word segmentation cues in German child-directed speech: a corpus analysis. *Lang. Speech* 65, 3–27. doi: 10.1177/0023830920979016

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

Thelen, E. (1981). Kicking, rocking, and waving: contextual analysis of rhythmical stereotypies in normal human infants. *Anim. Behav.* 29, 3–11. doi: 10.1016/S0003-3472(81)80146-7

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Dev. Psychol.* 39, 706–716. doi: 10.1037/0012-1649.39.4.706

Vanoncini, M., Boll-Avetisyan, N., Elsner, B., Hoehl, S., and Kayhan, E. (2022). The role of mother-infant emotional synchrony in speech processing in 9-month-old infants. *Infant Behav. Dev.* 69:101772. doi: 10.1016/j.infbeh.2022.101772

Vanoncini, M., Hoehl, S., Elsner, B., Wallot, S., Boll-Avetisyan, N., and Kayhan, E. (2024). Mother-infant social gaze dynamics relate to infant brain activity and word segmentation. *Dev. Cogn. Neurosci.* 65:101331. doi: 10.1016/j.dcn.2023.101331

Venetsanou, F., and Kambas, A. (2010). Environmental factors affecting preschoolers' motor development. *Early Childhood Educ. J.* 37, 319–327. doi: 10.1007/s10643-009-0350-z

Victora, M. D., Victora, C. G., and Barros, F. C. (1990). Cross-cultural differences in developmental rates: a comparison between British and Brazilian children. *Child Care Health Dev.* 16, 151–164. doi: 10.1111/j.1365-2214.1990.tb00647.x

Weissenborn, J., and Höhle, B. (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Philadelphia, PA: John Benjamins.

Wickham, H. (2016). *Ggplot2: elegant graphics for data analysis*. New York: Springer International Publishing.

Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon.

Wilson, M., and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968. doi: 10.3758/BF03206432

Woodrow, H. (1909). *A quantitative study of rhythm: the effect of variations in intensity, rate and duration*. Beijing, China: Science Press.

Woodrow, H. (1911). The role of pitch in rhythm. *Psychol. Rev.* 18, 54–77. doi: 10.1037/h0075201

Yeung, H. H., and Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychol. Sci.* 24, 603–612. doi: 10.1177/0956797612458802

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., et al. (2010). The development of perceptual grouping biases in infancy: a Japanese-English cross-linguistic study. *Cognition* 115, 356–361. doi: 10.1016/j.cognition.2010.01.005

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

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 02 July 2024
DOI 10.3389/fpsyg.2024.1393295

Check for updates

OPEN ACCESS

EDITED BY
Judit Gervain,
Centre National de la Recherche Scientifique (CNRS), France

REVIEWED BY
Chiara Cantiani,
Eugenio Medea (IRCCS), Italy
Gábor Péter Háden,
Hungarian Academy of Sciences (MTA), Hungary

*CORRESPONDENCE
Laura K. Cirelli
✉ laura.cirelli@utoronto.ca

RECEIVED 28 February 2024
ACCEPTED 06 June 2024
PUBLISHED 02 July 2024

CITATION
Cirelli LK, Talukder LS and Kragness HE (2024) Infant attention to rhythmic audiovisual synchrony is modulated by stimulus properties.
Front. Psychol. 15:1393295.
doi: 10.3389/fpsyg.2024.1393295

Infant attention to rhythmic audiovisual synchrony is modulated by stimulus properties

Laura K. Cirelli[1]*, Labeeb S. Talukder[1] and Haley E. Kragness[1,2]

[1]Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada, [2]Psychology Department, Bucknell University, Lewisburg, PA, United States

Musical interactions are a common and multimodal part of an infant's daily experiences. Infants hear their parents sing while watching their lips move and see their older siblings dance along to music playing over the radio. Here, we explore whether 8- to 12-month-old infants associate musical rhythms they hear with synchronous visual displays by tracking their dynamic visual attention to matched and mismatched displays. Visual attention was measured using eye-tracking while they attended to a screen displaying two videos of a finger tapping at different speeds. These videos were presented side by side while infants listened to an auditory rhythm (high or low pitch) synchronized with one of the two videos. Infants attended more to the low-pitch trials than to the high-pitch trials but did not display a preference for attending to the synchronous hand over the asynchronous hand within trials. Exploratory evidence, however, suggests that tempo, pitch, and rhythmic complexity interactively engage infants' visual attention to a tapping hand, especially when that hand is aligned with the auditory stimulus. For example, when the rhythm was complex and the auditory stimulus was low in pitch, infants attended to the fast hand more when it aligned with the auditory stream than to misaligned trials. These results suggest that the audiovisual integration in rhythmic non-speech contexts is influenced by stimulus properties.

KEYWORDS

infant perception, audiovisual synchrony, rhythm, music development, eye-tracking

1 Introduction

Music and song are frequently encountered in infants' everyday soundscapes (Mendoza and Fausey, 2021). While these experiences are sometimes unimodal, such as when infants listen to music from their car seat during a drive, they are often multimodal events. Caregivers gently rock their infants while making eye contact and singing, a melody plays from a rotating mobile above the crib, or a song accompanied by a video plays from a nearby television. A growing body of research suggests that even newborn infants can track an unfolding auditory rhythm (for a review, see Provasi et al., 2014), but many questions remain about how infants integrate auditory rhythms with corresponding visual rhythms and how this integration guides attention over time.

When adults listen to music, synchronous visual displays (e.g., an expressive singer's face and the performer playing their instrument) have an impact on emotional, perceptual, and esthetic judgments (Schutz and Lipscomb, 2007; Thompson et al., 2008; Platz and Kopiez, 2012; Pan et al., 2019). Adults are also quite capable of detecting audiovisual asynchrony in musical displays, although musical expertise and stimulus features interact to affect task difficulty (Petrini et al., 2009).

Less is known about how and when infants begin to link rhythmic sounds that they hear with synchronous visual displays. The limited research suggests that infants can at least discriminate between synchronous and asynchronous audiovisual rhythmic displays by 6 months of age (Gerson et al., 2015; Hannon et al., 2017). However, beyond discrimination, little is known about how infants deploy attention to competing synchronous or asynchronous audiovisual rhythmic displays when both are present. Hypotheses informed by the auditory scene analysis framework (Bregman, 1994) would predict that infants will deploy visual attention to the object most likely to be creating the auditory stream—for example, a mouth moving in synchrony with the speech stream. This aligns with the intersensory redundancy hypothesis (Lickliter et al., 2017), which stipulates that redundancy in multimodal stimuli effectively recruits attention, facilitating the perception of amodal properties, such as rhythm. Conversely, if detecting audiovisual rhythmic synchrony is easily achieved by infants, they might quickly shift their attention to the asynchronous visual display. This would support information-seeking models of infant attention: for example, the Hunter and Ames (1988) model of infant attention, which predicts that infants will attend to stimuli worthy of continued exploration, as well as the discrepancy hypothesis, which predicts that infants will attend most to events that are moderately complex (Kinney and Kagan, 1976; Kidd et al., 2014). Taken together, these models suggest that infant attention toward rhythmic audiovisual synchrony is likely to be modulated by stimulus properties, such as complexity, and may shift as a scene unfolds over time.

Previous studies reveal substantial variability in infant attention to audiovisual synchrony, potentially stemming from cross-study differences in methodologies, variations in stimulus features (speech vs. non-speech, rhythmic or non-rhythmic), and stimulus complexity (Shaw and Bortfeld, 2015). Existing evidence suggests that very young infants demonstrate an early-emerging preference for synchronous displays. For example, newborns hearing either a vocal or non-vocal sound preferentially look at one of two videos of a vocalizing monkey with a matching temporal structure (Lewkowicz et al., 2010). By around 3 months of age, infants presented with alternating synchronous and asynchronous displays of a face reciting nursery rhymes focused longer on the synchronous displays (Dodd, 1979). Similar synchrony preferences were found in 4-month-old infants watching simultaneously presented synchronous and tempo-shifted displays of two puppets bouncing isochronously and generating impact sounds (Spelke, 1979).

However, studies with older infants and more complex stimuli suggest that audiovisual synchrony does not consistently guide attention across all contexts. For example, when infants listen to an unfolding speech stream alongside two talking faces, infants below 12 months look equally at both displays, whereas those 12 to 14 months look longer at the synchronous display (Lewkowicz et al., 2015). This might suggest that ongoing speech streams are more difficult for infants to associate with competing visual displays compared to the stimuli used in the experiments cited previously. Corroborating this interpretation, the observed preference for synchronous talking faces documented after the first birthday appears to be stimulus-dependent and is eliminated when adult-directed (as opposed to infant-directed) speech or non-native languages are presented (Kubicek et al., 2014). This may be surprising given the early-emerging audiovisual synchrony detection documented even by newborns using vocal stimuli. However, it could be linked to the timing of perceptual narrowing and native-language speech specialization, processes that unfold after 6 months of age (Danielson et al., 2017). Overall, it is unclear whether these results, which seem incongruent with the early emergence of synchrony preference, stem from increased task complexity, developmental changes in cognitive ability (i.e., information-seeking behavior), or properties specific to the stimulus being used.

One stimulus feature of potential importance is pitch. Previous research with adults and infants suggests that listeners focus on high-frequency sounds when identifying the melody of music (Fujioka et al., 2005; Marie and Trainor, 2014; Trainor et al., 2014) and on low-frequency sounds when tracking the rhythm of music (Hove et al., 2014; Lenc et al., 2018, 2023). This suggests that if low-tone rhythms are easier to track (i.e., the low-tone superiority effect), they may also be easier to integrate with synchronous visual displays.

Another potentially important dimension to investigate is how attention is distributed over time. Recent research exploring infant multimodal perception of song, for example, suggests that infants dynamically shift their attention between a singer's eyes and mouth (Lense et al., 2022). Specifically, infants increase their attention to a singer's eyes around the musical beat window. These shifting attentional processes, which align with the dynamic attending models of rhythm perception (Large and Jones, 1999), highlight that exploring overall looking patterns collapsed over time may mask indicators of audiovisual integration. Instead, a real-time analysis of infant attention as it unfolds over time, such as with eye-tracking technology, may uncover subtler indications of audiovisual synchrony.

In the present study, we investigated how 8- to 12-month-old infants deploy attention over time while synchronous and asynchronous videos are presented side by side, concurrent with an auditory stimulus. Using eye-tracking, we examined how infants allocated attention to audiovisual synchrony at the trial level. Additionally, we investigated the impact of pitch (high vs. low) on audiovisual integration, given previous observations of a low-tone superiority effect for auditory rhythm processing in infants and adults. Infants were presented with two side-by-side videos depicting a hand tapping with one finger, each playing at distinct rates. Meanwhile, infants listened to either a high- or low-frequency rhythmic pattern synchronized with one of the two videos. We measured infants' relative looking time to the synchronous and asynchronous videos, as well as the time course of looking as trials unfolded. The auditory scene analysis framework suggests that infants would spend more time looking at the probable source of the sound—the synchronous video. If infants instead spend more time looking at the asynchronous video, this would support the models of infant attention that highlight information-seeking and preferences for moderate complexity levels. Furthermore, we explored how the pitch of the rhythmic sequence might impact infant attention and preference for synchrony. If infants demonstrate low-tone superiority for rhythmic processing, they may detect synchrony more readily in low-frequency conditions.

2 Methods

2.1 Participants

Full-term infants (>36 weeks gestation) between 8 and 12 months were recruited from the University of Toronto Scarborough Infant and Child Database. Target sample sizes were determined based on laboratory resources and samples used in prior research, documenting infant auditory–visual integration from various research groups (Lewkowicz et al., 2010; Kubicek et al., 2014; Gerson et al., 2015). Data were collected from 44 infants before testing was paused in March 2020 due to COVID-19 laboratory shutdowns. Seven infants were tested but excluded from analysis due to fussiness (4), calibration errors (2), or equipment failure (1). This left data from 37 infants in the analyses (M age = 10.46 months, SD = 1.27; 21 girls, 16 boys). The first 21 participants were assigned to the isochronous rhythm condition. The following 16 participants were assigned to the syncopated rhythm condition.

Infants came from diverse language backgrounds, with 57% exposed to more than one language, and mean English exposure at 77% (1 of the 37 participants did not report language background). Household incomes exceeded medians ($84,000 CAD; Statistics Canada, 2023) reported in this geographic region, with 19% reporting <$60,000/year, 35% reporting between $60,000 and $120,000/year, and 46% reporting > $120,000/year. Two caregivers did not provide income information. Additionally, 46% of caregivers reported that their infants participated in organized music lessons (for example, paid weekly programs such as Kindermusik or Music Together or free community weekly drop-in classes; 5 did not respond).

The University of Toronto Research Ethics Board approved all experimental procedures (Protocol 36642). Informed written consent was obtained from all parents. Infants received a junior scientist t-shirt and certificate for participating.

2.2 Stimuli

Auditory stimuli were generated in Audacity (2.2.2) on a Windows computer. These stimuli consisted of 200 ms pure tones with inter-beat intervals (IOI) of 430 ms (100 beats per minute, or bpm) or 600 ms (140 bpm). Pure tones had a 10-ms rise time and a 50-ms fall time. High- and low-frequency patterns were created using pure tone sine waves with 1236.8 Hz and 130 Hz, respectively, consistent with frequencies utilized by Lenc et al. (2018). Isochronous (x-x-x-x-x-x-x-x-) and syncopated (x--x--x---x-x---) rhythm patterns were used.

In each trial, visual stimuli consisted of two side-by-side finger-tapping videos: one synchronous with the tempo of the auditory stimulus and one asynchronous. Both videos were oriented such that the fingers were pointed toward the middle against a black background (see Figure 1). Pointing the fingers inward ensured that the points of impact were equidistant from the fixation point in the center of the screen, which infants fixated on before the trial began.

These videos were recorded at 60 frames per second using a Google Pixel 2. The model was a white adult woman tapping with her dominant (right) hand and pointer finger. Two types of tapping videos were recorded: isochronous and syncopated, each initially recorded at 515 ms IOI (116.5 beats per minute). These videos were subsequently sped up and slowed down by 16.5% using iMovie (10.1.9) to generate the 430 ms IOI (fast) and 600 ms IOI (slow) versions of each video. Mirror images were created by duplicating and flipping the videos to create a version with the finger pointing to the opposite side. The video commenced with both fingers starting their ascent from the surface (a wooden table) at the same time. Audio files were aligned so that the first pure tone occurred in synchrony with the first impact point for the synchronous video. There were 8 unique stimulus combinations that counterbalanced synchronous video location (left/right), tempo of the auditory rhythm (fast/slow), and pitch of the auditory rhythm (high/low). These 8 trial types were randomized within each trial block. An attention-getter, presented during calibration and between trials, was obtained from the Open Science Framework website,[1] consisting of colorful concentric circles and auditory chimes.

2.3 Apparatus

Infants were tested sitting on their parent's laps in a small dark room surrounded by heavy white curtains (see Figure 1). Each parent was provided with blacked-out glasses obscuring their vision as well as noise-isolating headphones playing music. Infants sat 55 cm in front of a 1280 × 1024 computer monitor. The audio was presented at 78.8 dBC SPL from a KRK Rokit 5 speaker centered below the monitor. Stimuli were presented using Experiment Builder (SR Research).

Eye movements were recorded using an EyeLink 1000 Plus system (SR Research Ltd.). The eye tracker camera recorded reflections of infrared light on the infant's cornea in relation to their pupil at a sampling rate of 500 Hz. A head-free setup was utilized with a target sticker placed on the infant's forehead between their eyebrows. The right eye was tracked across all infants. A three-point calibration procedure with manual experimenter confirmation was used to map gaze position to screen position, using the attention-getter (colorful spinning circles accompanied with a chime) at each target point.

2.4 Procedure

Following calibration, the experiment began. The attention-getter was presented in the center of the screen before each trial. The experimenter manually triggered the trial presentation after confirming that the infant gaze was within 10 degrees of the attention-getter and correcting for drift. Following the attention-getter, trials were presented for 8 s. Blocks of the 8 trial types (counterbalanced for synchrony left/right, fast/slow tempo, and high/low pitch) were repeated six times (48 trials total). The trial order was randomized within each block. Once calibration was complete, the procedure took approximately 10 to 15 min.

Upon completion of the experiment, the caregiver completed a general demographics questionnaire and the "Music@Home-Infant" questionnaire (Politimou et al., 2018), which gathered information about infants' musical home environments.

2.5 Data processing

For the analyses below, trials were retained if infants looked at least once at the left and at least once at the right display. These criteria led to the exclusion of 235 out of the 1527 trials (15%). This criterion was selected *a priori* to prioritize trial inclusion. The remaining trials had looking times that ranged from 172 ms to 7903 ms (M = 4432 ms,

[1] https://osf.io/wh7md/

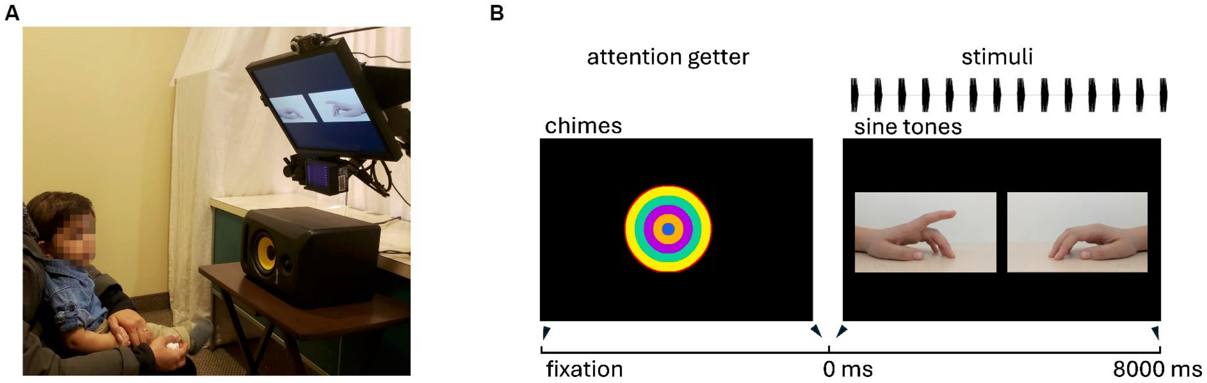


FIGURE 1
(A) Example of experimental setup. Infants sat on their parent's lap. Calibration stickers were placed on the children's foreheads. The loudspeaker was directly centered under the display screen and in front of the infant. Note that lights were dimmed during data collection. **(B)** An example of one trial. First, an attention-getter (shifting concentric circles accompanied by a chime sound) is displayed until the infant fixates. Then, the trial begins—two tapping hands are simultaneously presented, angled inward so that the point of contact is equidistant from the prior central fixation. The auditory rhythm is presented via the loudspeaker below the screen. Trials lasted 8000 ms.

SD = 1881 ms). Only 11 (<1%) of the included trials had looking times that were less than 2 SDs below the mean (670 ms). To liberally capture infant looking, our interest areas focused on the right vs. left half of the screen rather than specific interest areas in each video.

2.6 Analyses

Our primary dependent measures were (1) the proportion of time spent looking at the side of the screen displaying synchronous over the asynchronous display and (2) overall dwell times to either (synchronous/asynchronous) display. Exploratory dependent measures are described in more detail below. The proportion of looking at the synchronous and asynchronous displays was compared to chance levels (0.50) using one-sample t-tests. Linear mixed-effects models (LMEM; glmmTMB package, Brooks et al., 2017) in R (version 4.2.2, R Core Team, 2023) were used to evaluate the effects of pitch, tempo, and rhythmic complexity on infant-looking measures. We contrast-coded the repeated-measures variables pitch (low = −1, high = 1) and tempo (slow = −1, fast = 1) and the between-participants variable rhythmic complexity (isochronous = −1, complex = 1), such that a main effect of a factor represents the average effect across levels of the other factors.

Age, trial, and Music@Home scores were included as continuous predictors. For proportion-looking data, we assumed a beta distribution. For overall looking time, Gaussian distributions were assumed. Random intercepts for participants were included in the models to account for repeated measures.

3 Results

3.1 Preferential looking to synchronous or asynchronous displays

The infant proportion of time spent looking at the synchronous compared to the asynchronous side of the screen was calculated per trial. Overall, relative to the time infants spent looking at either half of the screen, they spent 49.9% of the time dwelling on the synchronous side. This did not differ significantly from chance levels (50%), $t(36) = -0.11$, $p = 0.913$ (one-sample test). To explore whether this null finding was driven by trials where infant looking may not have been long enough to notice synchrony, we ran the same test using a strict trial inclusion criterion requiring at least 1200 ms (at least two tap cycles) of looking to both the synchronous and asynchronous displays and found the same pattern, $t\ (36) = -0.61$, $p = 0.548$ (one-sample test). This pattern of distributed attention was consistent across conditions. A linear mixed-effects model demonstrated no significant effects of pitch, tempo, rhythmic complexity, or interaction between these terms on proportion time looking to the synchronous side (p's > 0.467). We also found no significant relationship between Music@Home general factor score, infant age, or trial number and proportion of synchronous looking (p's > 0.479). A simplified model exploring only the interaction between pitch and tempo while accounting for trial number revealed similar findings (p's > 0.282).

3.2 Overall attention across trials

The infant's total looking duration for either display was calculated per trial. A linear mixed-effects model was used to explore whether total looking changed across conditions (pitch, speed, rhythmic complexity, and trial number) and infant characteristics (age and Music@Home scores). While no interactions emerged, we found a simple effect of the trial ($B = -44.41$, $SE = 3.22$, $z = -13.77$, $p < 0.001$) and pitch ($B = -312.16$, $SE = 140.63$, $z = -2.22$, $p = 0.026$). As expected, overall attention to the displays reduced as trials progressed. Interestingly, infants spent more time looking at the screen in the low-pitch audio conditions ($M = 4435$ ms) than in the high-pitch audio conditions ($M = 4271$ ms).

3.3 Exploratory analyses around beat windows

Our initial hypothesis—that infants would prefer synchronous or asynchronous displays—was not supported. After completing our planned analyses, we further explored whether infants' attention to the synchronous and asynchronous hand shifted dynamically around the beat windows. This exploratory analysis was inspired by recent infant eye-tracking work showing that infants selectively attend a singer's eyes (compared to the mouth) at rhythmically important moments (Lense et al., 2022). For this analysis, 35 ms bins were identified across the window 210 ms before and after each beat for both the fast and slow taps within each trial. Then, for each trial, we determined if each infant fixated on the side of the screen displaying the tapping hand—looking at the fast hand around fast beat windows and the slow hand around slow beat windows—at least once within each of these 35ms bins. We then aggregated looks at the tapping hand around each beat window, considering whether the audio aligned with that beat window. This approach allowed us to calculate the proportion of bins containing looks at the same tapping hand when that hand was either congruent with the audio or incongruent with the audio. From these values, we calculated a difference score reflecting congruent–incongruent looking across bins surrounding the fast and slow beat windows. Positive values reflect more looking to the tapping hand on synchronous compared to asynchronous audio trials. For example, this would mean more looking to the fast hand around the fast beat window when fast audio is presented than when slow audio is presented. If infants did not integrate audiovisual information, they should distribute their attention similarly to a given hand regardless of audio congruence, resulting in a difference score close to 0. However, we hypothesized that if auditory stimuli guide visual attention to the tapping hand, infants should display a greater tendency to look at the tapping hand when it aligns with the audio.

A linear mixed-effects model was used to explore whether an infant looking at the congruent tapping hand was guided by features of the auditory stimuli. Our model explored the simple effects and interactions between pitch (high and low), speed of the tapping hand (fast and slow), and rhythmic complexity (isochronous and complex). A three-way interaction emerged, $B = -0.07$, $SE = 0.01$, $z = -4.89$, $p < 0.001$. Simple effects were explored within each rhythmic complexity condition (see Figure 2).

Within the isochronous rhythm condition, the main effects of pitch ($B = -0.02$, $SE = 0.005$, $z = -3.12$, $p = 0.002$) and speed of the tapping hand ($B = -0.01$, $SE = 0.005$, $z = -2.18$, $p = 0.029$) were qualified by an interaction between these factors ($B = 0.04$, $SE = 0.007$, $z = 5.09$, $p < 0.001$). Above-baseline congruent looks (more looking when audio is congruent) to the fast hand were greater in the high-pitch condition than in the low-pitch condition, $p < 0.001$. Conversely, above-baseline congruent looks at the slow hand were greater in the low-pitch condition than in the high-pitch condition, $p <. 001$.

In the syncopated rhythm condition, the main effects of pitch ($B = -0.03$, $SE = 0.009$, $z = -3.50$, $p < 0.001$) and speed of the tapping hand ($B = 0.03$, $SE = 0.009$, $z = 3.09$, $p = 0.002$) were again qualified by an interaction between these factors ($B = -0.03$, $SE = 0.01$, $z = -2.47$, $p = 0.014$). The difference scores for looks at the fast-tapping hand and the slow-tapping hand were both greater in the low-pitched conditions than in the high-pitched conditions. However, this pitch effect was more dramatic for congruent looking at the fast hand. Visual inspection suggests that variability across participants was higher in this condition than in the isochronous rhythm condition. While this increased variability may be reflective of the increased complexity of the stimulus, it may also be a by-product of the smaller sample (n = 16 compared to 21 infants).

4 Discussion

When presented with two side-by-side videos of fingers tapping rhythmically, 8- to 12-month-old infants did not show overall within-trial preferences for the video that aligned with the auditory rhythm. Furthermore, our analyses found no effect of rhythmic complexity (isochronous/syncopated), auditory pitch (high/low), tempo (fast/slow), or infant musical background on their interest in the synchronous vs. asynchronous display. This finding may be surprising, given that much younger infants prefer to attend to visual displays that align with presented audio (Spelke, 1979; Lewkowicz et al., 2010), but converges with other research within this age group, suggesting that this synchrony preference is inconsistent, if present at all (Kubicek et al., 2014; Lewkowicz et al., 2015). These findings are unlikely to reflect low interest in the stimuli, which are arguably less interesting than speech streams—trial-level dwell times exceeded 50% of the trial lengths.

A synchrony preference would have provided support for the auditory scene analysis framework (Bregman, 1994) and would have suggested that infants use auditory–visual synchrony to guide attention to likely sound sources. Overall preferences for attending asynchronous displays, on the other hand, would have suggested that infants in this age range find synchrony detection to be trivial and shift to the display, warranting more exploration (Hunter and Ames, 1988).

Not finding support for either model and inspired by recent research investigating infant attention to a singing face (Lense et al., 2022), we explored infant attention around the beat window. Specifically, we asked if cross-trial interest in the fast and slow displays was facilitated by hearing a congruent rhythm. Here, our analysis revealed preliminary evidence for integration and evidence that stimulus features mattered. Across most conditions, infants displayed a greater tendency for congruent compared to incongruent fixations to the tapping hand in the low-pitch condition compared to the high-pitch condition. This pattern was particularly pronounced in the syncopated rhythm/fast hand condition. These initial findings provide preliminary evidence that infants are integrating the rhythms that they hear with the rhythms that they see—if these streams were being processed independently, we would not expect to see above-baseline congruent looks. Above-baseline looking suggests that infants are especially likely to look at a particular rhythmic visual display when it aligns with the rhythms being heard. While this analysis is exploratory, it highlights the value of exploring fine-grained infant attention to synchronous displays instead of only looking at averaged interest collapsed across trial lengths.

We did not find any evidence for individual differences in synchrony preferences or congruent looks around beat windows relating to infant age or home music background (from the Music@Home scale). Due to the interruption of in-person data collection by the COVID-19 lockdowns, future research with larger samples may be able to address this question more directly. For example, previous research with 6-month-old infants shows that infants provided an opportunity to interact with a toy drum are subsequently more interested in videos showing the same toy drum being struck synchronously rather than asynchronously with auditory rhythms (Gerson et al., 2015). This example of short-term experience raises questions about whether long-term experience also impacts early attentional biases for audiovisual synchrony.

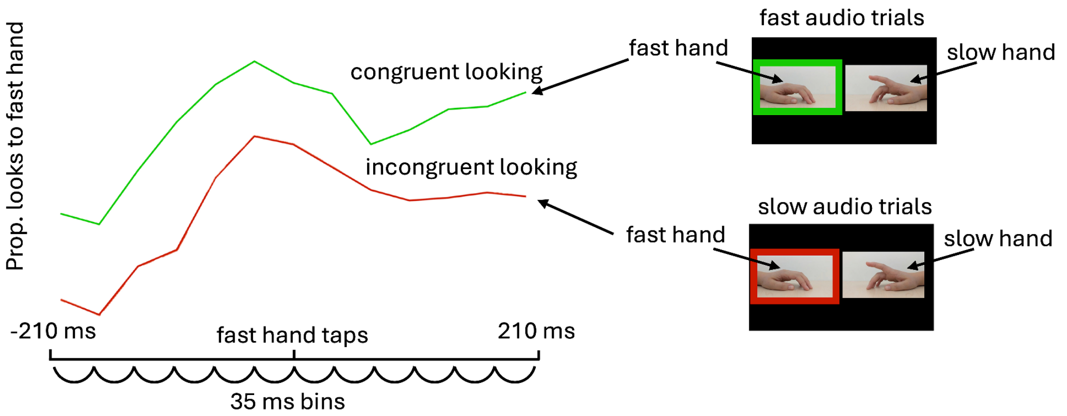


FIGURE 2
The exploratory analysis investigated whether attention to the stimuli was enhanced by audiovisual congruence around the tap window. Specifically, we asked whether attention around taps for a given video (e.g., the fast hand) was enhanced when the audio was congruent compared to incongruent. First, the window around each finger tap was divided into 35-ms bins (6 before and 6 after). Each bin was assigned 1 or 0 (1 = a fixation to the tapping hand occurred). These values were then aggregated across taps within trials and across trials within each pitch condition. Finally, the proportion of bins containing looks at the tapping hand on congruent trials (in this example, the fast hand in fast audio trials) was calculated and compared to the proportion of bins containing looks at the tapping hand on incongruent trials (here, the fast hand in slow audio trials). The difference scores in Figure 3 reflect incongruent looking subtracted from congruent looking.

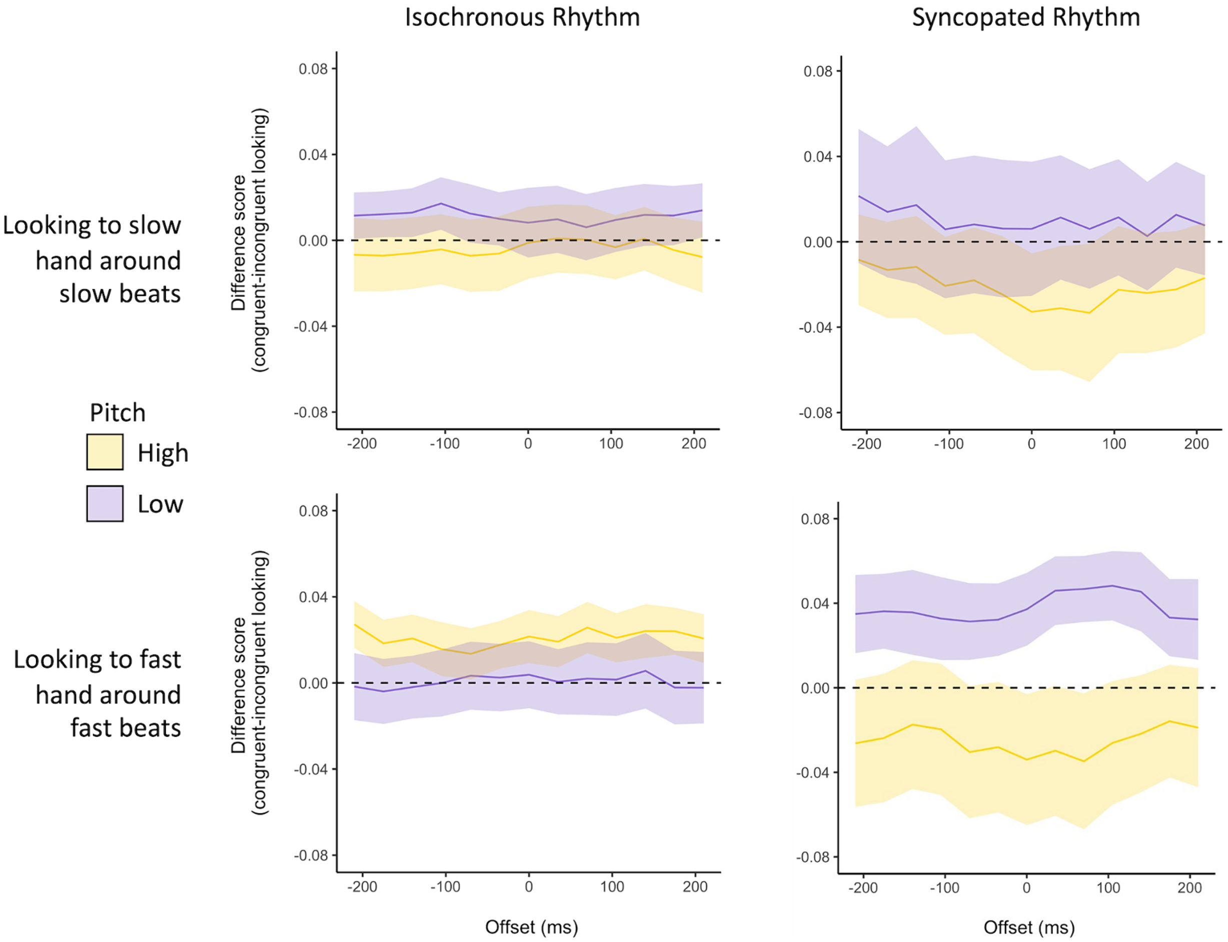


FIGURE 3
Here, we plot infant attention to the tapping fingers around the fast (top row) and slow (bottom row) beat windows for infants in the isochronous (left) and complex (right) rhythm conditions. The y-axis shows the difference score in looking at these hands when the audio is congruent vs. incongruent with that hand's tapping tempo (i.e., looking above baseline represents more looking when audio aligns than when audio does not align). The error bars represent the standard error of the mean.

Irrespective of whether infants engaged in synchronous or asynchronous looking, they demonstrated more time attending to the visual displays in the low-pitch condition compared to the high-pitch condition. This observation may be interpreted in light of the low-tone superiority effect, demonstrating that rhythmic information is better extracted from low-pitch signals (Hove et al., 2014; Lenc et al., 2023). Perhaps infants were more interested in exploring the two visual rhythms when the auditory stream provided a more salient rhythmic context. Future research could explore the effect of pitch on rhythm processing by asking whether infants are better able to detect rhythmic violations in low- compared to high-pitch streams. It is also worth noting that infant preferences for pitch in musical signals are context-dependent—for example, infants prefer to listen to low- over high-pitched lullabies but prefer to listen to high- over low-pitched playsongs (Volkova et al., 2006; Tsang and Conrad, 2010). Lullabies also tend to have slower and steadier rhythms (Trainor et al., 1997) and are more effective at downregulating infant arousal (Cirelli et al., 2020). Questions remain about how pitch interacts with rhythm and functional goals in shaping infants' perceptions and emotional reactions to everyday musical exchanges.

Future studies are needed to harmonize the existing research on the developmental trajectory of auditory–visual integration in infancy. Here, we opted to utilize musically relevant rhythmic patterns (isochronous and syncopated), which were selected to match those used in prior work exploring the low-tone superiority effect (Lenc et al., 2018). One potential consideration, however, is that the audiovisual pairings we selected—namely, sine tones and tapping fingers—do not occur naturally. Previous research has shown that infants as young as 6 months are sensitive to some aspects of audiovisual congruence in impact events. When presented with side-by-side videos that are *both* temporally synchronized with an auditory stimulus, infants preferentially watch the display that matches the acoustic properties of the heard material (Bahrick, 1987). In contrast, however, infants are likely to integrate natural speech and sine wave speech when presented synchronously with a talking face (Baart et al., 2014) and experience audiovisual illusions—such as the sound-bounce illusion—even when the "sound" paired with the bounce is an artificial beep (Sekuler et al., 1997; Scheier et al., 2003). Therefore, many unanswered questions remain about the potential facilitatory effects of naturalistic vs. artificial audiovisual pairing and the role of experience in informing infants' expectations about naturalistic audiovisual pairings. The present research highlights that considering stimulus properties and tracking dynamic attention is an important step toward building predictions about how audiovisual synchrony guides attention in early life.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the University of Toronto Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

LC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. LT: Conceptualization, Investigation, Methodology, Writing – review & editing. HK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) held by LKC (RGPIN-2019-04689).

Acknowledgments

We would like to thank Mark Schmuckler for the use of the eye-tracking equipment, laboratory space, and for technological support. We thank Julia Santiago and Deepika Elango for their assistance with data collection and Adam Frost for technological support.

Conflict of interest

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

Publisher's note

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

References

Baart, M., Vroomen, J., Shaw, K., and Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition* 130, 31–43. doi: 10.1016/j.cognition.2013.09.006

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behav. Dev.* 10, 387–416. doi: 10.1016/0163-6383(87)90039-7

Bregman, A. S. (1994). Auditory scene analysis. Cambridge, MA: MIT press.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. , Nielsen, A., et al (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

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

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., and Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cogn. Dev.* 42, 37–48. doi: 10.1016/j.cogdev.2017.02.004

Dodd, B. (1979). Lip reading in infants: attention to speech presented in-and out-of-synchrony. *Cogn. Psychol.* 11, 478–484. doi: 10.1016/0010-0285(79)90021-5

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592. doi: 10.1162/089892905774597263

Gerson, S. A., Schiavio, A., Timmers, R., and Hunnius, S. (2015). Active drumming experience increases infants' sensitivity to audiovisual synchrony during observed drumming actions. *PLoS One* 10:e0130960. doi: 10.1371/journal.pone.0130960

Hannon, E. E., Schachner, A., and Nave-Blodgett, J. E. (2017). Babies know bad dancing when they see it: older but not younger infants discriminate between synchronous and asynchronous audiovisual musical displays. *J. Exp. Child Psychol.* 159, 159–174. doi: 10.1016/j.jecp.2017.01.006

Hove, M. J., Marie, C., Bruce, I. C., and Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc. Natl. Acad. Sci.* 111, 10383–10388. doi: 10.1073/pnas.1402039111

Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*. 5, 69–95.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2014). The goldilocks effect in infant auditory attention. *Child Dev.* 85, 1795–1804. doi: 10.1111/cdev.12263

Kinney, D. K., and Kagan, J. (1976). Infant attention to auditory discrepancy. *Child Dev.* 47, 155–164. doi: 10.2307/1128294

Kubicek, C., Gervain, J., De Boisferon, A. H., Pascalis, O., Lœvenbruck, H., and Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behav. Dev.* 37, 644–651. doi: 10.1016/j.infbeh.2014.08.010

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295X.106.1.119

Lenc, T., Keller, P. E., Varlet, M., and Nozaradan, S. (2018). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proc. Natl. Acad. Sci.* 115, 8221–8226. doi: 10.1073/pnas.1801421115

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

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

Lewkowicz, D. J., Leo, I., and Simion, F. (2010). Intersensory perception at birth: newborns match nonhuman primate faces and voices. *Infancy* 15, 46–60. doi: 10.1111/j.1532-7078.2009.00005.x

Lewkowicz, D. J., Minar, N. J., Tift, A. H., and Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: its emergence and the role of experience. *J. Exp. Child Psychol.* 130, 147–162. doi: 10.1016/j.jecp.2014.10.006

Lickliter, R., Bahrick, L. E., and Vaillant-Mekras, J. (2017). The intersensory redundancy hypothesis: extending the principle of unimodal facilitation to prenatal development. *Dev. Psychobiol.* 59, 910–915. doi: 10.1002/dev.21551

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

Mendoza, J. K., and Fausey, C. M. (2021). Everyday music in infancy. *Dev. Sci.* 24:e13122. doi: 10.1111/desc.13122

Pan, F., Zhang, L., Ou, Y., and Zhang, X. (2019). The audio-visual integration effect on music emotion: behavioral and physiological evidence. *PLoS One* 14:e0217040. doi: 10.1371/journal.pone.0217040

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., et al. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Exp. Brain Res.* 198, 339–352. doi: 10.1007/s00221-009-1817-2

Platz, F., and Kopiez, R. (2012). When the eye listens: a meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Percept. Interdiscip. J.* 30, 71–83. doi: 10.1525/mp.2012.30.1.71

Politimou, N., Stewart, L., Müllensiefen, D., and Franco, F. (2018). Music@home: a novel instrument to assess the home musical environment in the early years. *PLoS One* 13:e0193819. doi: 10.1371/journal.pone.0193819

Provasi, J., Anderson, D. I., and Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Front. Psychol.* 5:1048. doi: 10.3389/fpsyg.2014.01048

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Scheier, C., Lewkowicz, D. J., and Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Dev. Sci.* 6, 233–241. doi: 10.1111/1467-7687.00276

Schutz, M., and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception* 36, 888–897. doi: 10.1068/p5635

Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* 385:308. doi: 10.1038/385308a0

Shaw, K. E., and Bortfeld, H. (2015). Sources of confusion in infant audiovisual speech perception research. *Front. Psychol.* 6:1844. doi: 10.3389/fpsyg.2015.01844

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Dev. Psychol.* 15, 626–636. doi: 10.1037/0012-1649.15.6.626

Statistics Canada. (2023). Census Profile. 2021 Census of Population. Statistics Canada Catalogue no. 98-316-X2021001. Ottawa. Released November 15, 2023. https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

Thompson, W. F., Russo, F. A., and Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognit. Emot.* 22, 1457–1470. doi: 10.1080/02699930701813974

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

Trainor, L. J., Marie, C., Bruce, I. C., and Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: evidence from cortical evoked potentials and peripheral auditory models. *Hear. Res.* 308, 60–70. doi: 10.1016/j.heares.2013.07.014

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

Volkova, A., Trehub, S. E., and Schellenberg, E. G. (2006). Infants' memory for musical performances. *Dev. Sci.* 9, 583–589. doi: 10.1111/j.1467-7687.2006.00536.x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 July 2024
DOI 10.3389/fpsyg.2024.1375353

Check for updates

OPEN ACCESS

EDITED BY
Marcela Pena,
Pontificia Universidad Católica de Chile, Chile

REVIEWED BY
Caroline Junge,
Utrecht University, Netherlands
Joana L. Gonçalves,
Lusíada University, Portugal

*CORRESPONDENCE
Krisztina Lakatos
✉ lakatos.krisztina@ttk.hu

†These authors share first authorship

RECEIVED 23 January 2024
ACCEPTED 17 June 2024
PUBLISHED 04 July 2024

CITATION
Balázs A, Lakatos K, Harmati-Pap V, Tóth I and Kas B (2024) The influence of temperament and perinatal factors on language development: a longitudinal study. *Front. Psychol.* 15:1375353.
doi: 10.3389/fpsyg.2024.1375353

The influence of temperament and perinatal factors on language development: a longitudinal study

Andrea Balázs[1,2†], Krisztina Lakatos[2*†], Veronika Harmati-Pap[1], Ildikó Tóth[2] and Bence Kas[1,3]

[1]Institute for General and Hungarian Linguistics, HUN-REN Hungarian Research Centre for Linguistics, Budapest, Hungary, [2]Sound and Speech Perception Research Group, Institute of Cognitive Neuroscience and Psychology, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary, [3]MTA-ELTE Language-Learning Disorders Research Group, Eötvös Loránd University, Bárczi Gusztáv Faculty of Special Needs Education, Budapest, Hungary

Early language development is characterized by large individual variation. Several factors were proposed to contribute to individual pathways of language acquisition in infancy and childhood. One of the biologically based explaining factors is temperament, however, the exact contributions and the timing of the effects merits further research. Pre-term status, infant sex, and environmental factors such as maternal education and maternal language are also involved. Our study aimed to investigate the longitudinal relationship between infant temperament and early language development, also considering infant gender, gestational age, and birthweight. Early temperament was assessed at 6, 9, 18, 24, and 30 months with the Very Short Form of Infant Behavior Questionnaire (IBQ-R) and the Very Short Form of Early Childhood Behavior Questionnaire (ECBQ). Early nonverbal communication skills, receptive and expressive vocabulary were evaluated with the Hungarian version of The MacArthur Communicative Development Inventory (HCDI). Our study adds further evidence to the contribution of infant temperament to early language development. Temperament, infant gender, and gestational age were associated with language development in infancy. Infants and toddlers with higher Surgency might enter communicative situations more readily and show more engagement with adult social partners, which is favorable for communication development. Gestational age was previously identified as a predictor for language in preterm infants. Our results extend this association to the later and narrower gestational age time window of term deliveries. Infants born after longer gestation develop better expressive vocabulary in toddlerhood. Gestational age may mark prenatal developmental processes that may exert influence on the development of verbal communication at later ages.

KEYWORDS

language development, temperament, gestational age, longitudinal study, sex differences

1 Introduction

Early language development shows great individual variation both in the extension and the expansion of receptive and expressive vocabulary. The range of a 12-month-old, typically developing child's receptive vocabulary may span from 25 to more than 200 words, and similar variation can be observed in expressive vocabulary (Fenson et al., 2007). Some children utter their first words by the age of 12 months while others only after 18 months. The rate of development begins to even out by the third year of life (Fenson et al., 2007). Previous studies indicated several biological factors influencing the dynamics of linguistic development in early childhood. Gestational age (Barre et al., 2011), birthweight (Stolt et al., 2009), infant sex (Law et al., 2019) and infant temperament (Ishikawa-Omori et al., 2022) were among these factors, however, social class, family history, certain environmental characteristics (AlHammadi, 2017) seem to play a role.

Concerning gestational age, studies suggest that children born very preterm and extremely preterm exhibit delayed language skills compared to full-term children (Foster-Cohen et al., 2007, 2010). In their earlier work, Foster-Cohen et al. (2007) studied 90 preterm children ($N = 36$ extremely preterm gestational age < 28 weeks, and $N = 54$ very preterm, gestational age 28–33) and 102 full-term children (gestational age 38–41). The MacArthur-Bates Communicative Development Inventory: Words and Sentences (CDI-WS) at 2 years of corrected age was associated with gestational age at birth. Vocabulary size, word use quality, morphological and syntactic complexity were related to longer gestation before birth. An association between gestational age and language outcomes persisted after the authors controlled for child and family factors otherwise related to gestational age. At 4 years (Foster-Cohen et al., 2010), the association between language development and very preterm birth was replicated. These children had significantly poorer linguistic outcomes even after excluding children with neurosensory impairment and statistical control for the effect of social risk. By contrast, Pérez-Pereira et al. (2016) studied language performance at 30 months with the Galician version of the CDI. Comparing low-risk preterm (mean gestational age, GA: 32.60 weeks) and full-term children (GA: 39.84 weeks), they found no significant differences in the language outcomes: word production, MLU and sentence complexity between groups.

However, the third trimester is characterized by important developmental changes in the brain. Shortened gestation, even within the normal term delivery range (greater than 37 weeks), had long-lasting effects on neural development in a healthy, low-risk population (Davis et al., 2011) with lower gray matter density detected by magnetic resonance imaging. These structural differences may lead to variation in later cognitive development as well. Can et al. (2013) identified several brain regions with early white matter and gray-matter concentrations in association with infants' receptive language ability and expressive language at 12 months. The indicated cerebellum, PLIC/cerebral peduncle, and hippocampus are suggested to be associated with early language development. These brain developmental processes may contribute to the underlying mechanism connecting higher gestational age with better receptive language at 24 months of age in a sample of toddlers born after 32 weeks of gestation (Snijders et al., 2020). The Norwegian Mother and Child Cohort Study found that children born both early-term and late-preterm had an increased risk for communication impairment at 18 months and for expressive language impairments at 36 months (Stene-Larsen et al., 2014). Thus, we hypothesize that the linguistic performance of a full-term child may also be related to gestational age in a middle-class, term infant sample.

The contribution of birthweight to variations in language development tends to be confounded by preterm status (Stolt et al., 2009; Barre et al., 2011). No effect of birthweight on language outcomes was detected in a sample of Hungarian children on the Hungarian version of CDI-III (Dale et al., 2001; Fenson et al., 2007; Kas et al., 2022) at 2–4 years of age. The sample of 1,424 term children included 9.3% low-birthweight (<2,500 g) children. CDI scores were predicted by children's age, gender, and parents' education level, whereas other factors including birthweight, birth problems, number of siblings, birth order, multilingualism, familial net income, and children's chronic illness did not have significant effects. Individual differences within normal birthweight (>2,500 g) have not yet been linked to language development, however, Full-scale IQ performance was positively associated with birthweight within the normal range (Matte et al., 2001). Marinopoulou et al. (2021) found that the number of words used by children at age 2.5 years was associated with deficits in intellectual functioning at age 7 years. Children who used 50 words or fewer at age 2.5 years had lower scores of Full-scale IQ, verbal comprehension, working memory, and perceptual reasoning at age 7 years. Given the contradictory results and the potential association via IQ, further investigation of the role of birthweight is needed.

Although there is a growing body of research on the role of infant temperament (Ishikawa-Omori et al., 2022), the results are inconclusive. Studies differ in the definitions of temperament, the stage of language development investigated, the age range of the children, the length of data collection, and the set of other variables included in the analyses. The diversity of these parameters makes it difficult to compare the results. Major theories agree that temperament is inherently present at an early age and influences the expression of behaviors related to activity, affectivity, and self-regulation (Goldsmith et al., 1987; Shiner et al., 2012). However, different approaches to temperament use divergent operational definitions and thus operate within somewhat different frameworks. According to Rothbart and Derryberry (1981) and Rothbart (2007), whose approach was applied in the present study, temperament is constitutionally based, can be measured from infancy, and shows a relatively stable pattern extending over the lifetime (Hampson and Goldberg, 2006; Putnam et al., 2008; Kopala-Sibley et al., 2018; Tang et al., 2020). It can be defined as individual differences in reactivity and self-regulation that manifest in emotions, activity, and attention. Temperament is described by 3 major, distinct factors: Positive Emotionality/Surgency, Negative Affectivity and Regulatory Capacity/Effortful Control (see Table 1 for example items assessing the three factors). Buss and Plomin's (1984) approach shares some of the concepts and behaviors observed, Thomas and Chess (1977) defined rather different temperament types based on nine dimensions of temperament that captured patterns relevant to clinical practice. While these theories consider emotions and affectivity as components of temperament, Goldsmith (1996) sees temperament as the expression and regulation of emotions. Thus, instruments based on one theory or the other may capture different aspects of temperament.

Based on Rothbart's concept, longitudinal positive associations were found between temperament and expressive language skills. Children's expressive vocabulary and length of utterance at 24 months were associated with Approach and Perceptual Sensitivity measured at 8 and 12 months of age (Davison et al., 2019). The scales of Approach and Perceptual Sensitivity, along with others, contribute to the Surgency factor (Gartstein and Rothbart, 2003). Laake and Bridgett (2014) also reported that a higher Surgency score measured at 10 months was predictive of improved expressive but not receptive language at 14 months. This relationship might be related to higher infant Surgency predicting higher levels of toddler Effortful Control

TABLE 1 Example items assessing the 3 factors of temperament (effortful control, Surgency, negative affectivity).

	Very Short Form of Infant Behavior Questionnaire (IBQ-R)	Very Short Form of Early Childhood Behavior Questionnaire (ECBQ)
Surgency	During a peekaboo game, how often did the baby laugh?	When offered a choice of activities, how often did your child decide what to do very quickly and go after it?
	When hair was washed, how often did the baby vocalize?	When encountering a new activity, how often did your child get involved immediately?
	How often during the week did your baby move quickly toward new objects?	While participating in daily activities, how often did your child seem full of energy, even in the evening?
Effortful control	How often during the last week did the baby enjoy being read to?	When told "no," how often did your child stop the forbidden activity?
	How often during the last week did the baby play with one toy or object for 5–10 min?	When asked to wait for a desirable item (such as ice cream), how often did your child wait patiently?
	How often during the last week did the baby stare at a mobile, crib bumper or picture for 5 min or longer?	When asked to do so, how often was your child able to be careful with something breakable?
Negative affectivity	When tired, how often did your baby show distress?	When visiting a new place, how often did your child not want to enter?
	When introduced to an unfamiliar adult, how often did the baby cling to a parent?	When told "no," how often did your child become sadly tearful?
	When introduced to an unfamiliar adult, how often did the baby refuse to go to the unfamiliar person?	Following an exciting activity or event, how often did your child seem to feel down or blue?

(Putnam et al., 2008), and in turn, Effortful Control was reported to be associated with expressive language (Bruce et al., 2022). Also, as Positive Anticipation contributes to the Surgency factor, the general learning enhancing aspect or/and the social aspect of positive affect might be considered here as well. Kort et al. (2001) reported that positive affect enhanced students' learning behavior. Yang et al. (2013) found that positive affect was related to better working memory and had a weaker relationship with short-term memory. They suggest that positive affect facilitates controlled cognitive processing, leading to improved learning ability. We may assume that improved learning ability may support language learning as well. Language learning is greatly facilitated by interactions with social partners. Dixon and Smith (2000) claim that individual differences in positive or negative emotionality might moderate the willingness of social partners to enter social dialogs in the first place, thus influencing exposure to language. Ishikawa-Omori et al. (2022) studied receptive and expressive vocabulary at 40 months. They found that two scales contributing to the Negative Affectivity factor, Motor Activation and Perceptual Sensitivity at 18 months predicted language skills at 40 months, however, the associations pointed in opposite directions. Higher scores on Perceptual Sensitivity were related to larger expressive and receptive vocabulary at 40 months, while higher scores on Motor Activation were related to poorer receptive and expressive vocabulary. Garello et al. (2012) also found concurrent negative correlations between Motor Activation and language development in 24- and 30-month-old children.

Early attentional control and the capacity for self-regulation, which consistently loaded on the Effortful Control factor, were associated positively with language development in infancy and early childhood as well. Dixon and Shore (1997) and Dixon and Smith (2000) reported that attentional control, positive affect and emotional stability measured at 13 months predicted the efficiency of language acquisition, including the time of appearance of first words and the time and speed of vocabulary expansion at 20–21 months. Dixon and Smith (2000) explained this pattern of longitudinal association by Rothbart and Bates's (2007) theory of an early attentional control system, which corresponds to the maturation of the anterior attentional system at the end of the first year. This early attentional control system allows children to voluntarily direct and maintain attention and allows flexibility in awareness. In fact, emergent control of attention indicated by increases in the duration of orientation from 7 to 10 months was found to be associated with advanced language production at 20 months.

In summary, higher Positive Affect and Effortful Control at the end of the first year and the beginning of the second year are associated with better language performance between 1 and 2 years of age. Conversely, a higher score on the Negative Affectivity between 18 and 30 months is associated with poorer language performance between 24 and 40 months. Additionally, Negative Affectivity may influence the rate of expressive language development around the age of 2 and beyond due to a lower likelihood of engaging in social interactions.

The present study focused on examining the role of perinatal variables in addition to temperament in language development and assessing concurrent and longitudinal relationships in a longitudinal design. Both language development and temperament were evaluated repeatedly, allowing for capturing the potentially changing patterns of associations between temperament and language skills. In addition to expressive language and receptive vocabulary, gestural communication was measured. According to the design of CDI, the latter two were assessed up to 18 months (Frank et al., 2021). Regression models were used to determine the effect of temperament, infant gender and perinatal factors.

2 Materials and methods

2.1 Participants

A longitudinal project on early language development was carried out by recruiting 186 families. The inclusion criteria for the present study were that the child was born on time (gestational age > 37 and birthweight >2,500) and was taken to the baby lab at least once. All infants included in the present investigation were of low social risk and were the first-born children of their mothers. As is common in developmental studies with voluntary participants, mothers with higher education were overrepresented, with 75% having college or university degrees. All participants came from metropolitan (Budapest) or agglomeration areas, all children were monolingual. No hearing problems were reported. The sample was ethnically homogeneous Caucasians of Hungarian origin. Families were recruited at the infant's birth, 4, 9, and 18 months of age (see Table 2). All families received detailed information on the study, and informed consent was obtained. The first wave included 74 middle-class mothers recruited in the HONVED PMC hospital's maternity ward. Data were collected up to 18 months in this phase. An additional recruitment at 18 months was planned to increase the sample size continuing to the second phase, however, as the dropout rate was higher than previously expected due to the COVID-19 epidemic, additional recruitment of 4- and 9-month-old infants was carried out (see Table 2). The sex ratio and infant characteristics in the participating and the dropout families did not differ significantly. The present data set includes varying numbers of infants at different ages due to the disruption caused by the pandemic breaking out during data collection and preventing families from visiting the child laboratory. The exact numbers of available data at each age are presented in Table 2.

2.2 Procedures and instruments

According to the original protocol, mothers were to fill in the questionnaires in the baby lab while lab assistants played with the children and administered tests in the passive presence of the mother. However, the Covid-19 pandemic resulted in some mothers completing the questionnaires from their homes online. There were no significant differences in temperament or language between the questionnaires administered before and during the pandemic.

Ethical approval for the study was granted by ETT-TUKEB (1942-12/2016) and EPKEB (77/2015).

2.2.1 Measurements of child temperament

Infant temperament was assessed using the Very Short Forms of the Infant Behavior Questionnaire (IBQ–R) (Putnam et al., 2014) and the Early Childhood Behavior Questionnaire (ECBQ) (Putnam et al., 2006). Mothers completed the 37-item IBQ–R and 36-item ECBQ (Hungarian versions: Lakatos et al., 2010) either in the baby lab or online at home. The IBQ–R was administered at 6 and 9 months of infant's age, whereas the ECBQ was at 18, 24 and 30 months. Mothers rated the frequency of their infants' behaviors over the past two weeks using seven-point Likert scales. Three main factors were computed: Surgency, Effortful Control, and Negative Affectivity. Cronbach alpha coefficients of internal consistency for these factors in this sample were between 0.607 and 0.805 (see Table 3). Missing data were not substituted.

2.2.2 Measurements of language and communication skills

For the assessment of early language development, the Hungarian adaptation of the MacArthur-Bates Communicative Development Inventory (CDI) Words & Gestures and Words & Sentences parent report forms (Fenson et al., 2007) has been used (Kas et al., 2010, 2022). This questionnaire relies on maternal (caregiver) reports to explore children's receptive and expressive vocabulary and assess their level of speech comprehension, gesture use, morpheme acquisition, and syntactic complexity through systematic questions. CDI forms are suitable for assessing language development in typically developing children aged 8–30 months or older with developmental disorders. The present study considers the following CDI variables: (1) receptive vocabulary total score, (2) expressive vocabulary total score, and (3) gestures total score including sub-scores of object manipulation, imitation of adults, symbolic activity, and non-verbal gesture use. The CDI was first administered at 9 months of age, followed by a second administration at 12 months. Thereafter, the course of language development was monitored at two-month intervals until the age of 30 months (Figure 1). For the present report, language data from 18, 24, and 30 months was included in the analyses. Eighteen months of age represents a major turning point in language development, as this is the last age when all 3 dimensions of the CDI (receptivity, expression and gesture) are assessed. Twenty-four-month language data was included because it showed the highest variability. Thirty-month expressive language as measured by CDI was also characterized by good variability. Temperament was also assessed at these ages, thus concurrent associations can be examined.

2.2.3 Analyses

Data were analyzed with IBM SPSS (Version 26). Descriptive measures of linguistic variables obtained from H-CDI (Receptive and Expressive Vocabulary and Communicative Gestures) and of the temperament variables obtained from IBQ-R and ECBQ (Surgency, Effortful Control, Negative Affectivity), perinatal variables, and infant sex were calculated (Table 4). According to the results of Shapiro–Wilk tests, parametric and non-parametric tests were carried out in the analyses. Sex differences were investigated for all variables. To examine the contribution of temperament, perinatal factors to the individual variation in language development, we first analyzed correlations of

TABLE 2 Number of participants at each data collection point in cohorts recruited at different ages.

Data collection points	Cohorts				Total number of participants
	Newborn	4-month	9-month	18-month	
6 months	68	39	-	-	107
9 months	56	39	26	-	121
18 months	53	36	25	37	151
24 months	42	35	25	37	139
30 months	40	35	25	31	131

these variables with language development at various ages (see Table 5). Variables with significant associations were entered as predictors in stepwise linear regression analysis to determine their predictive value on the dependent linguistic variables, such as receptive and expressive vocabulary and communicative gestures at 18 months of age, and expressive vocabulary at 24 and 30 months of age.

3 Results

3.1 Descriptive statistics

Means and standard deviations for the whole sample, and for boys and girls separately are presented in Table 4. The age-related growth of expressive vocabulary between 9 and 30 months is depicted in Figure 1.

One-way ANOVA or Kruskal-Wallis tests were applied to assess sex differences. No gender differences were observed in gestational age, but boys were significantly heavier at birth [$H(1, n = 179) = 20.394$, $p < 0.001$). Likewise, no significant differences between boys and girls appeared on temperament scales at any age, apart from a statistical trend towards boys scoring higher on Surgency at 9 months [$F(1,119) = 2.891$, $p = 0.092$]. However, significant sex differences were found in CDI language scores (Table 4). Girls scored higher on all CDI sub-scales at most time points, except for receptive vocabulary at 18 months [$H(1, n = 147] = 2.939$, $p = 0.086$), yet with a tendency in favor of girls.

TABLE 3 Cronbach's alpha coefficients of internal consistency of IBQ-R-SF and ECBQ-SF factors.

Age Scale name	Cronbach's Alpha
6 months	
Surgency	0.686
Effortful control	0.677
Negative affectivity	0.805
9 months	
Surgency	0.607
Effortful control	0.705
Negative affectivity	0.756
18 months	
Surgency	0.738
Effortful control	0.703
Negative affectivity	0.688
24 months	
Surgency	0.680
Effortful control	0.723
Negative affectivity	0.626

3.2 Correlations with language scores

Bivariate relationships between language development and temperament, demographic and perinatal variables were explored by correlation analyses (Table 5) to select variables for regression analyses predicting language outcomes.

Concurrent correlations were investigated at 18, 24 and 30 months. 18-month Surgency showed a consistent relationship with all measures of language development: higher Surgency was related to better language skills. Better Effortful Control was significantly related to more developed use of gestures and there was a tendency toward better receptive vocabulary. At 24 months, higher Surgency was related to better expressive vocabulary. At 30 months, the association between these two measures only showed a trend-level correlation, however, in the same direction as at earlier ages.

Longitudinal correlations were weak and sparse, however, Surgency at various ages tended to be related to measures of language and communicative development at later ages. Higher Surgency at 9 months was related to higher receptive vocabulary and gesture use at 18 months. Similarly, positive associations appeared between 18-month Surgency and expressive vocabulary at 24 months, and 24-month Surgency and expressive vocabulary at 30 months, with higher Surgency being related to a more extensive expressive

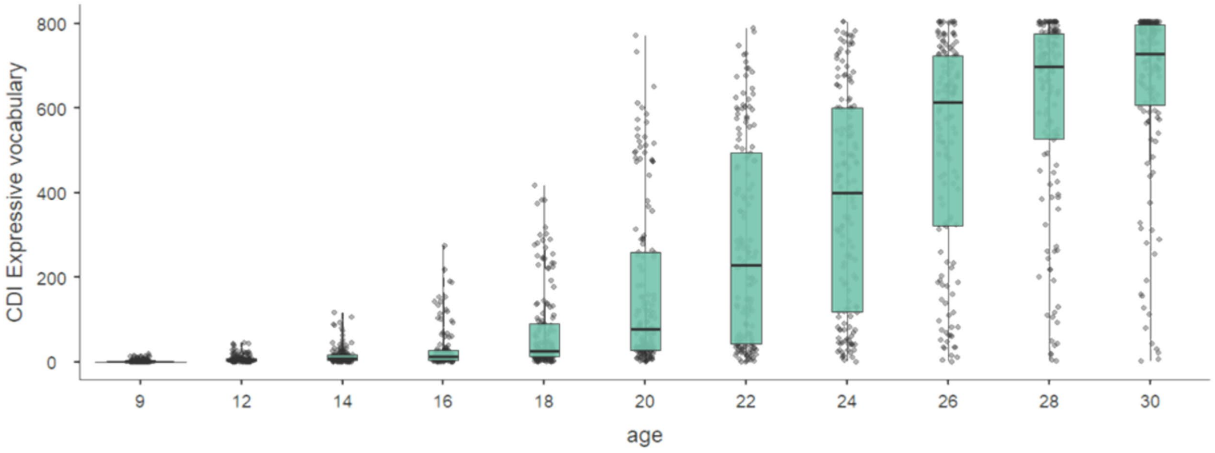


FIGURE 1
CDI expressive vocabulary scores by age.

TABLE 4 Descriptive statistics of and differences by infant sex in perinatal factors, temperament scales, and language skills.

	Variables at different ages	*N*	Mean (SD)	Range Min-Max	Girls *N*	Girls Mean (SD)	Boys Mean (SD)	*H or F* statistic
	Birthweight	171	3416.46 (405.50)	2,410–5,350	82	3278.51 (358.18)	3,543 (406.80)	**20.327***
	Gestational age	171	39.56 (1.128)	37–43	81	39.54 (1.15)	39.56 (1.11)	0.242
Surgency	6 months	105	5.26 (0.69)	3.75–6.72	53	5.18 (0.76)	5.34 (0.59)	1.427
	9 months	121	5.31 (0.65)	3.77–6.75	61	5.21 (0.65)	5.41 (0.64)	**2.891**$^{+}$
	18 months	140	5.26 (0.76)	2.83–6.58	69	5.15 (0.79)	5.37 (0.72)	2.644
	24 months	130	5.32 (0.73)	3.00–6.67	63	5.28 (0.68)	5.35 (0.77)	0.466
Effortful control	6 months	105	5.43 (3.54)	3.5–6.98	53	5.47 (0.648)	5.39 (0.66)	0.452
	9 months	121	5.28 (3.98)	2.75–6.73	61	5.30 (0.63)	5.23 (0.77)	0.010
	18 months	140	4.65 (2.77)	2.455–6.42	69	4.73 (0.75)	4.57 (0.71)	1.673
	24 months	130	4.91 (3.00)	2.75–6.83	63	4.95 (0.67)	4.88 (0.75)	0.322
Negative affectivity	6 months	105	3.54 (0.96)	1.25–6.20	53	3.64 (1.03)	3.44 (0.88)	1.104
	9 months	121	3.98 (0.90)	1.25–6.00	61	4.03 (0.97)	3.94 (0.83)	0.289
	18 months	140	2.77 (0.71)	1.29–4.91	69	2.70 (0.69)	2.84 (0.72)	1.349
	24 months	130	3 (0.73)	1.50–5.91	63	3.02 (0.80)	2.98 (0.66)	0.011
Language and communication	Receptive vocabulary 18 months	147	311.45 (100.25)	27–455	73	326.15 (95.17)	296.95 (103.62)	**2.939**$^{+}$
	Gestures 18 months	147	57.99 (11.41)	25–81	73	61 (11.26)	55.03 (10.83)	**10.751****
	Expressive vocabulary 18 months	147	72.66 (95.35)	0–383	73	84.60 (105.34)	60.88 (83.40)	**5.043***
	Expressive vocabulary 24 months	144	380.60 (256.98)	1–804	71	430.23 (237.59)	332.34 (267.38)	**5.097***
	Expressive vocabulary 30 months	138	655.32 (205.14)	2–804	69	691.16 (183.36)	619.58 (220.39)	**5.652***

+$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Bold highlights significant sex differences.

vocabulary. In addition, 18-month receptive vocabulary and gesture use were associated with 9-month Effortful Control. Higher Negative Affectivity at 6 months was significantly correlated with more developed expressive vocabulary at 24 months. However, later Negative Affectivity (at 18 months) had the opposite relationship with expressive vocabulary at 30 months: more negative affect was associated with lower expressive vocabulary a year later.

Of the temperament factors in infancy and early childhood, Surgency seems to be indicated in language acquisition both concurrently and longitudinally, spanning from receptive to expressive language.

3.3 Longitudinal predictors of language development

To better understand how temperament and perinatal factors affect each language and communication skill measured by the H-CDI, we conducted linear regression analyses with stepwise selection separately for each CDI variable at 18, 24 and 30 months. Temperament variables of preceding ages, perinatal variables, and infant sex showing significant correlations with the predicted variable were included in the regression.

First, we examined receptive vocabulary at 18 months (Table 6). Birthweight, Surgency and Effortful Control at 9 months were entered in the model. In the final model [$R^{2=}0.089$, $F(1,104) = 10.161$, $p = 0.002$], the single significant predictor was Surgency measured at 9 months ($\beta = 0.298$, $p = 0.002$).[1]

In the model predicting gestures at 18 months, sex, Surgency, and Effortful Control at 9 months were entered (see Table 7). In the final model [$R^{2=}0.111$, $F(2,103) = 7.559$, $p = 0.001$], predictive variables were Surgency measured at 9 months ($\beta = 0.272$, $p = 0.004$) and sex ($\beta = -0.275$, $p = 0.004$).[2]

To predict expressive vocabulary at 18 months, only sex and gestational age were entered, as no temperament variable showed a significant correlation with this language outcome (see Table 8). Here, only one model was generated [$R^{2=}0.076$, $F(1,143) = 11.778$, $p < 0.001$], with gestational age reaching significance ($\beta = 0.276$, $p < 0.001$).[3]

Table 9 presents the regression model predicting expressive vocabulary at 24 months, in which sex, gestational age, Surgency and Negative Affectivity at 18 months were entered. In the final model

[1] Multicollinearity was not detected (birthweight, Tolerance=0.99, VIF=1.00; Surgency at 9months, Tolerance=1.00 VIF=1.00; Effortful Control at 9months, Tolerance=0.84, VIF=1.19).

[2] Multicollinearity was not detected (sex, Tolerance=0.98, VIF=1.02; Surgency at 9months, Tolerance=0.98, VIF=1.02; Effortful Control at 9months, Tolerance=0.83, VIF=1.21).

[3] Multicollinearity was not detected (gestational age, Tolerance=1.00, VIF=1.00; sex, Tolerance=1.00, VIF=1.00).

TABLE 5 Correlations between perinatal factors, temperament, and language.

	Receptive vocabulary	Gestures	Expressive vocabulary		
	18 months	18 months	18 months	24 months	30 months
Birthweight	**0.204* (146)**	0.044 (146)	**0.153^{+} (146)**	0.130 (144)	**0.179* (138)**
Gestational age	0.153 (145)	**0.161^{+} (145)**	**0.265** (145)**	**0.212* (143)**	**0.167^{+} (137)**
Temperament					
Surgency					
6 months	0.110 (82)	**0.200^{+} (82)**	−0.034 (82)	−0.009 (78)	−0.022 (75)
9 months	**. 285** (106)**	**0.233* (106)**	**0.165^{+} (106)**	0.149 (102)	**0.183^{+} (98)**
18 months	**0.285*** (137)**	**0.319*** (137)**	**0.191* (137)**	**0.183* (133)**	−0.137 (128)
24 months				**0.213* (130)**	**0.200* (126)**
30 month				**0.064^{+} (114)**	
Effortful control					
6 months	0.089 (82)	**0.190^{+} (82)**	0.089 (82)	0.022 (78)	0.016 (75)
9 months	**0.221* (106)**	**0.221* (106)**	0.122 (106)	0.026 (102)	0.088 (98)
18 months	**0.160^{+} (137)**	**0.313*** (137)**	0.137 (137)	0.122 (133)	**0.152^{+} (128)**
24 months				0.096 (130)	0.133 (126)
30 month				0.183 (114)	
Negative affectivity					
6 months	0.169 (82)	0.0147 (82)	0.077 (82)	0.187 (78)	**0.233* (75)**
9 months	0.062 (106)	0.032 (106)	0.070 (106)	0.105 (102)	0.095 (98)
18 months	0.031 (137)	−0.138 (137)	−0.061 (137)	**−0.202* (133)**	**−0.159^{+} (128)**
24 months				−0.021 (130)	−0.024 (126)
30 month					−0.005 (114)

Pearson's r or Spearman's rho, +$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Bold: Correlation surviving Bonferroni correction ($p < 0.05/5$).

[R^{2}=0.112, F(2,129) = 8.163, $p < 0.001$], gestational age ($\beta = 0.248$, $p = 0.004$) and Negative Affectivity measured at 18 months ($\beta = -0.199$, $p = 0.019$) were the significant predictors.[4]

Infant sex, birthweight, Negative Affectivity at 6 months, and Surgency at 24 months were entered into the regression to predict expressive vocabulary at 30 months (see Table 10). In the final model [R^{2}=0.088, F(1,64) = 6.152, $p = 0.016$], the only significant contributor was Surgency at 24 months[5] ($\beta = 0.296$, $p = 0.016$).

4 Discussion

Our study aimed to investigate the longitudinal relationship between infant temperament and early language development, also considering infant sex and gestational age. Several data collection points for temperament (6–30 months) and language development (18–30 months) were included. Our findings support the role of both infant temperament and perinatal factors in early language development. Nine-month Surgency forecasted receptive vocabulary at 18 months and also contributed to gestural communication at 18 months in addition to infant sex. Gestational age predicted expressive vocabulary at 18 and 24 months. In addition, Negative Affectivity at 18 months also contributed to 24-month expressive vocabulary. Thirty-month expressive vocabulary was predicted by Surgency measured at 24 months.

While Surgency appears to have a significant influence on receptive language and gestures at 18 months, and expressive vocabulary at 30 months, there was a lack of association with expressive vocabulary at 18 and 24 months. Instead, expressive vocabulary at these ages was related to gestational age. Thus, there seems to be a discontinuity in the effect of Surgency, with the emergence of gestational age. Bates et al. (1992) describe increases in vocabulary and grammar along with increases in synaptic density and brain metabolism between the ages of 16–30 months. These brain developmental processes might not be independent of prenatal brain development potentially marked by gestational age. This may be reflected in gestational age predicting 18- and 24-month expressive vocabulary. Surgency, however, may play a role in the expansion of gestural and verbal communication via potentially increased exposure to communicative signals and engagement in social interaction (Laake and Bridgett, 2014). This may be reflected in the association with a more extensive receptive vocabulary and gestures at 18 months, and expressive language at a later age (30 months), when verbal communication is established in most of the children.

[4] Multicollinearity was not detected (gestational age, Tolerance=0.98, VIF=1.01; Negative Affectivity at 18months, Tolerance=0.98, VIF=1.01; sex, Tolerance=0.98, VIF=1.01; Surgency at 18months, Tolerance=0.97, VIF=1.03).

[5] Multicollinearity was not detected (Surgency at 24months, Tolerance=1.00, VIF=1.00; sex, Tolerance=0.98, VIF=1.02; birthweight, Tolerance=0.99, VIF=1.00; Negative Affectivity at 6months, Tolerance=0.98 VIF=1.02).

TABLE 6 Receptive vocabulary at 18 months was predicted by 9-month Surgency.

Predictors	Beta	Sig.	R2	Change in R^2	Change in F	Sig. change in F	F	df2	Sig.
Model 1			0.089	0.089	10.161	0.002	10.161	104	0.002
Surgency (9 months)	0.298	0.002							
				0.044	5.223	0.024	7.899	103	0.001

Stepwise linear regression analyses, *N* = 106, excluded variables birthweight, 9-month Effortful Control.

TABLE 7 Gestures at 18 months were predicted by infant sex and Surgency (9 months).

Predictors	Beta	Sig.	R^2	Change in R^2	Change in F	Sig. change in F	F	df2	Sig.
Model 1			0.056	0.056	6.125	0.001	6.125	104	0.015
Sex	−0.236	0.015							
Model 2			0.128	0.072	8.548	0.013	7.559	103	0.001
Sex	−0.275	0.004							
Surgency (9 months)	0.272	0.004							

Stepwise linear regression analyses, *N* = 106, excluded variables: effortful control (9 months).

TABLE 8 Expressive vocabulary at 18 months was predicted by gestational age.

Predicators	Beta	Sig.	R2	Change in R^2	Change in F	Sig. change in F	F	df2	Sig.
Model 1			0.076	0.076	11.778	0.001	11.778	143	0.001
Gestational age	0.276	0.001							

Stepwise linear regression analyses, *N* = 145, excluded variables: sex.

TABLE 9 Expressive vocabulary at 24 months was predicted by gestational age and negative affectivity (18 months).

Predictors	Beta	Sig.	R2	Change in R^2	Change in F	Sig. change in F	F	df2	Sig
Model 1			0.073	0.73	10.298	0.002	10.298	130	0.002
Gestational age	0.271	0.002							
Model 2			112	0.39	5.660	0.019	8.163	129	0.000
Gestational age	248	0.004							
Negative Affectivity (18 months)	−0.199	0.019							

Stepwise linear regression analyses, *N* = 132, excluded variables: sex, Surgency (18 months).

TABLE 10 Expressive vocabulary at 30 months was predicted by Surgency at 24 month.

Predictors	Beta	Sig.	R2	Change in R^2	Change in F	Sig. change in F	F	df2	Sig.
Model 1			0.088	0.088	6.152	0.003	6.15	64	0.016
Surgency (24 months)	0.296	0.016							

Stepwise linear regression analyses, *N* = 66, excluded variables: sex, birthweight, negative affectivity (6 months).

4.1 Surgency

Several studies have linked positive affectivity with language development in infancy and early childhood (Laake and Bridgett, 2014; Pérez-Pereira et al., 2016; Davison et al., 2019). Positive affectivity contributes to the Surgency factor in Rothbart's temperament model (Gartstein and Rothbart, 2003). Laake and Bridgett found that 10-month-old infants with higher Positive Affectivity/Surgency, as measured by IBQ-R, showed improved expressive language at 14 months. Davidson's study also supported these findings, as infant Positive Affectivity/Surgency measured at 8 and 12 months predicted expressive language skills at 24 months.

Consistent with the literature, we also found Surgency to be related to early language skills. Surgency at 9 months predicted receptive vocabulary and gesture use at 18 months, while Surgency measured at 24 months was a significant contributor to expressive vocabulary at 30 months. Of concurrent associations between Surgency and language measures, only correlations with 18-month receptive vocabulary and gesture use remained significant after Bonferroni correction. However, at least a trend-level association with concurrent Surgency pointing in the same direction can be observed for expressive vocabulary at all ages. Thus, infants with higher Surgency scores demonstrated better language abilities, both in terms of receptive and expressive language. These results suggest that Surgency may be related to language development over an extended period. Since there is some stability in Surgency over time (correlations among Surgency values measured between 9–30 months ranged between 0.358–0.694), temperament can be expected to show a weak longitudinal correlation with expressive communication.

As children with high Positive Affectivity/Surgency are more likely to engage in and elicit social interactions, they have more opportunities to practice and improve their expressive language skills (Laake and Bridgett, 2014). This assumption could also apply to gesture use and receptive language, as both are related to expressive language use. Extensive social interactions provide more opportunities not only for the use of expressive vocabulary but also for gestural communication. More social interactions may result in varied, and increased amounts of language stimuli, fostering the development of language skills.

4.2 Effortful control

In our study, Effortful Control was not a significant predictor of language development in the regression models. Only weak correlations were observed between Effortful Control at 9 months and gesture use and receptive vocabulary at 18 months. Medium concurrent correlation with gesture use was also observed at 18 months.

The link between effortful control and language development remains unclear, despite some studies (Salley and Dixon, 2007; Keller et al., 2016) suggesting a positive relationship that could potentially be attributed to varying attentional capacities, which are thought to support language acquisition (Snijders et al., 2020). Effortful Control, as measured by Rothbart's temperament questionnaires, is related to the functioning of the executive network (Posner et al., 2016). In turn, a link was demonstrated between the executive network and language development, production, and comprehension (Ye and Zhou, 2009; Shokrkon and Nicoladis, 2022). Furthermore, language development may also contribute to executive function development and self-regulation (Roben et al., 2013; Bruce et al., 2023).

However, Bruce et al. (2022) found that Effortful Control was only related to concurrent language, and 10-month Orienting/Effortful Control did not predict 24-month expressive language. Similarly, Ishikawa-Omori et al. (2022) did not find a predictive link between Effortful Control at 18 months and language development at 40 months. Keller et al. (2016) only demonstrated a significant relationship in the second language competence of dual language learners in childhood. The lack of predictive power of Effortful Control preceding the age of language assessment in the regression models and the separately observed concurrent correlation are in line with these results.

4.3 Negative affectivity

Negative Affectivity was entered in regressions at 24 and 30 months, however, only 18-month Negative Affectivity proved to be a significant predictor for lower expressive vocabulary at 24 months. This result supports earlier findings that Negative Affectivity may be associated with worse language skills (Dixon and Smith, 2000; Garello et al., 2012; Ishikawa-Omori et al., 2022). For instance, Garello et al. found that at the ages of 24–30 months, increased Negative Emotionality and Motor Activity correlated with poorer language production and comprehension. Similarly, Ishikawa-Omori et al. (2022) reported that Motor Activity, a scale of the Negative Affectivity factor, measured at 18 months, predicted lower expressive and receptive language skills at 40 months. They suggested that fidgeting behavior may reduce the availability of attentional resources, and as a result, it could hinder language learning. Excessive negative emotions could limit the resources children can allocate for information processing and language learning. They may also influence the way the children and their social partners interact. Children displaying more negative affect indeed performed worse on a joint attention task at 21 months (Salley and Dixon, 2007).

4.4 Gestational age

Gestational age proved to be a significant predictor for expressive vocabulary at 18 and 24 months. It's been well-documented that both preterm birth and low birthweight can negatively impact language development into school age and beyond (Husby et al., 2023). Emerging findings, however, suggest variation in the development of term-born children, indicating differing developmental trajectories for early-term, full-term, late-term and post-term children (MacKay et al., 2010; Espel et al., 2014; Bentley et al., 2016; Snijders et al., 2020; Dhamrait et al., 2021). Our results suggest that longer *in-utero* development may support the development of expressive language. The final weeks of intrauterine development are characterized by rapid brain development. Children born early-term will not benefit from the effect of uterine neurosteroids (Hüppi et al., 1998; Limperopoulos et al., 2005; Shaw et al., 2019) as long as children born at later gestational ages. Increasing evidence shows long-lasting brain structure differences in preterm infants (Inder et al., 2005; Rogers et al., 2018). For instance, variations in functional connectivity were present even in adolescence after preterm birth, suggesting distinctive neurodevelopment potentially underlying behavioral differences (Lubsen et al., 2011). Term-born infants' brain development also seems to benefit from longer gestation, within the time window of 37–41st weeks. Gestational age was related to differences in brain development in school-age children (Davis et al., 2011; Nivins et al., 2023). Such a variation may contribute to the observed differences in cognitive functioning and language skills (Ma et al., 2022).

4.5 Infant sex

Other than temperament and gestational age, infant sex also seems to contribute to variations in language skills. Previous studies have shown sex differences in language acquisition (Eriksson et al., 2012; Law et al., 2019), which aligns with our findings. Except for receptive vocabulary at 18 months, girls performed significantly better on all language measures and infant sex predicted the use of gestures at 18 months. Although we have only assessed gestures at 18 months, previous studies found girls using more gestures and starting earlier than boys (Özçalişkan and Goldin-Meadow, 2010; Germain et al., 2022).

4.6 Conclusion

Our aim was to investigate the role of temperament and some perinatal and maternal characteristics on early language development in a sample of low-social-risk, first-born term infants. Our sample was rather homogeneous as all participants were Caucasian of Hungarian origin, and the maternal education level was generally high across the sample. Results indicate the contribution of Surgency both concurrently and longitudinally on various measures of language development and the influence of gestational age on expressive vocabulary at 18 and 24 months. Negative Affectivity only predicted expressive vocabulary at 24 months. Despite Effortful Control being correlated with 18-month language, it was not a significant predictor in the regression models.

However, a major limitation of the study was the sample size and the missing data due to the pandemic. The COVID-19 pandemic hit during data collection and caused unexpected loss of data (data collection could not be conducted due to closures) and thus a higher dropout rate. Although the pandemic had no direct influence on the data presented (no differences were observed on any measures between data collected pre-pandemic and pandemic, post-pandemic periods), there might be hidden underlying effects of the quarantine period. Only 3 families reported contracting COVID-19 during the data collection. Thus, we may assume that results were not influenced by the neurological effects of the viral infection. Another limitation was the relatively low reliability of some temperament factors (Surgency at 9 months: 0.61, Negative Affectivity at 24 months: 0.63). Since correlations of the Surgency factor were consistent with those of other ages (albeit weak across the board), we have decided to include it in the regression analyses.

Our results extend previous findings as we have demonstrated associations with Surgency at the early stages of language acquisition for both receptive and expressive vocabulary, and showed the additional significant contribution of gestational age and Negative Affectivity. Gestational age was identified as a predictor for language in preterm infants previously. Our results extend this association to the narrower time window of gestational age of full-term infants. The latter finding may have relevance for medical practice and child educational support agencies. In line with other studies highlighting difficulties in later academic performance, this calls for increased attention to the early development of early-term and term infants.

4.7 Future directions

Our results highlight the importance of longitudinal studies using tools to measure temperament based on the same theoretical concept over time. Also, investigating the small differences in gestational age in term infants in a larger sample may reveal important effects on language acquisition. With more evidence on how early-term status may influence later cognitive and language development, research on how certain environmental factors, such as socioeconomic status, maternal education, and quality of mother–child interaction might interact with gestational age can yield important results that can be translated into practices supporting early childhood development.

Extending the study beyond 30 months is crucial to identify early characteristics of developmental pathways leading to language impairment. The role of Surgency and the relative lack of power for Effortful Control in this sample calls for experimental investigation of the development of very early executive functions and attentional functioning.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Medical Research Council, Scientific and Research Ethics Committee (ETT-TUKEB) United Ethical Review Committee for Research in Psychology (EPKEB). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

AB: Formal analysis, Investigation, Project administration, Writing – original draft, Writing – review & editing, Data curation. KL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. VH-P: Investigation, Writing – review & editing. IT: Data curation, Investigation, Project administration, Writing – review & editing. BK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Hungarian National Research Development and Innovation Office (Grant: K124477 to BK).

Acknowledgments

The first phase of the study providing infant temperament data was funded by the grant K115385 awarded to I. Winkler. We are grateful for making these data available. The authors thank Judit Rochéné Farkas, Lívia Őzéné Kende, Zsófia Haraszti, Tamara Lakatos, Ágnes Kata Szerafin, Kinga Kerner, Anna Radányi, Blanka Horváth, Beatrix Uhrin, Rebeka Szekeres, and Eszter Rozgonyiné Lányi for their contribution in data collection.

Conflict of interest

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

Publisher's note

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

References

AlHammadi, F. S. (2017). Prediction of child language development: A review of literature in early childhood communication disorders. *Lingua* 199, 27–35. doi: 10.1016/j.lingua.2017.07.007

Barre, N., Morgan, A., Doyle, L. W., and Anderson, P. J. (2011). Language abilities in children who were very preterm and/or very low birthweight: a meta-analysis. *J. Pediatr.* 158, 766–774.e1. doi: 10.1016/j.jpeds.2010.10.032

Bates, E., Thal, D., and Janowsky, J. S. (1992). Early language development and its neural correlates. In *Handbook of neuropsychology*. (eds.) S. J. Segalowitz, and I. Rapin Vol. 7. Child neuropsychology. Amsterdam: Elsevier Science Publishers.

Bentley, J. P., Roberts, C. L., Bowen, J. R., Martin, A. J., Morris, J. M., and Nassar, N. (2016). Planned birth before 39 weeks and child development: a population-based study. *Pediatrics* 138:e20162002. doi: 10.1542/peds.2016-2002

Bruce, M., Ermanni, B., and Bell, M. A. (2023). The longitudinal contributions of child language, negative emotionality, and maternal positive affect on toddler executive functioning development. *Infant Behav. Dev.* 72:101847. doi: 10.1016/j.infbeh.2023.101847

Bruce, M., McFayden, T. C., Ollendick, T. H., and Bell, M. A. (2022). Expressive language in infancyand toddlerhood: the roles of child temperament and maternal parenting behaviors. *Dev. Psychobiol.* 64:e22287. doi: 10.1002/dev.22287

Buss, A. H., and Plomin, R. (1984). Temperament: Early developing personality traits. Hillsdale, NJ: Erlbaum.

Can, D. D., Richards, T., and Kuhl, P. K. (2013). Early gray-matter and white-matter concentration in infancy predict later language skills: a whole brain voxel-based morphometry study. *Brain Lang.* 124, 34–44. doi: 10.1016/j.bandl.2012.10.007

Dale, P. S., Reznick, J. S., Thal, D. J., and Marchman, V. A. (2001). A parent report measure of language development for three-year-olds. Columbia, MO: University of Missouri-Columbia.

Davis, E. P., Buss, C., Muftuler, L. T., Head, K., Hasso, A., Wing, D. A., et al. (2011). Children's brain development benefits from longer gestation. *Front. Psychol.* 2:1. doi: 10.3389/fpsyg.2011.00001

Davison, L., Warwick, H., Campbell, K., and Gartstein, M. A. (2019). Infant temperament affects toddler language development. *J. Educ. e-Learn. Res.* 6, 122–128. doi: 10.20448/journal.509.2019.63.122.128

Dhamrait, G. K., Christian, H., O'Donnell, M., and Pereira, G. (2021). Gestational age and child development at school entry. *Sci. Rep.* 11:14522. doi: 10.1038/s41598-021-93701-y

Dixon, W. E. Jr., and Shore, C. (1997). Temperamental predictors of linguistic style during multiword acquisition. *Infant Behav. Dev.* 20, 99–103. doi: 10.1016/S0163-6383(97)90065-5

Dixon, W. E., and Smith, P. H. (2000). Links between early temperament and language acquisition. *Merrill-Palmer Q.* 46, 417–440.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x

Espel, E. V., Glynn, L. M., Sandman, C. A., and Davis, E. P. (2014). Longer gestation among children born full term influences cognitive and motor development. *PLoS One* 9:e113758. doi: 10.1371/journal.pone.0113758

Fenson, L., Marchman, V. A., Thal, D., Dale, P., Reznick, J. S., and Bates, E. (2007). MacArthur-Bates communicative development inventories: User's guide and technical manual. *2nd* Edn. Baltimore, MD: Brookes Publishing Co.

Foster-Cohen, S., Edgin, J., Champion, P., and Woodward, L. (2007). Early delayed language development in very preterm infants: evidence from the MacArthur-Bates CDI. *J. Child Lang.* 34, 655–675. doi: 10.1017/S0305000907008070

Foster-Cohen, S. H., Friesen, M. D., Champion, P. R., and Woodward, L. J. (2010). High prevalence/low severity language delay in preschool children born very preterm. *J. Dev. Behav. Pediatr.* 31, 658–667. doi: 10.1097/DBP.0b013e3181e5ab7e

Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. Cambridge, MA: MIT Press.

Garello, V., Viterbori, P., and Usai, M. C. (2012). Temperamental profiles and language development: A replication and an extension. *Infant Behav. Dev.* 35, 71–82. doi: 10.1016/j.infbeh.2011.09.003

Gartstein, M. A., and Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behav. Dev.* 26, 64–86. doi: 10.1016/S0163-6383(02)00169-8

Germain, N., Gonzalez-Barrero, A. M., and Byers-Heinlein, K. (2022). Gesture development in infancy: effects of gender but not bilingualism. *Infancy* 27, 663–681. doi: 10.1111/infa.12469

Goldsmith, H. H. (1996). Studying temperament via construction of the toddler behavior assessment questionnaire. *Child Dev.* 67, 218–235. doi: 10.2307/1131697

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., et al. (1987). Roundtable: what is temperament? Four approaches. *Child Dev.* 58, 505–529. doi: 10.2307/1130527

Hampson, S. E., and Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *J. Pers. Soc. Psychol.* 91, 763–779. doi: 10.1037/0022-3514.91.4.763

Hüppi, P. S., Warfield, S., Kikinis, R., Barnes, P. D., Zientara, G. P., Jolesz, F. A., et al. (1998). Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Ann. Neurol.* 43, 224–235. doi: 10.1002/ana.410430213

Husby, A., Wohlfahrt, J., and Melbye, M. (2023). Gestational age at birth and cognitive outcomes in adolescence: population based full sibling cohort study. *BMJ* 380:e072779. doi: 10.1136/bmj-2022-072779

Inder, T. E., Warfield, S. K., Wang, H., Hüppi, P. S., and Volpe, J. J. (2005). Abnormal cerebral structure is present at term in premature infants. *Pediatrics* 115, 286–294. doi: 10.1542/peds.2004-0326

Ishikawa-Omori, Y., Nishimura, T., Nakagawa, A., Okumura, A., Harada, T., Nakayasu, C., et al. (2022). Early temperament as a predictor of language skills at 40 months. *BMC Pediatr.* 22, 1–10. doi: 10.1186/s12887-022-03116-5

Kas, B., Jakab, Z., and Lőrik, J. (2022). Development and norming of the Hungarian CDI-III: A screening tool for language delay. *Int. J. Lang. Commun. Disord.* 57, 252–273. doi: 10.1111/1460-6984.12686

Kas, B., Lőrik, J., Andrea, S. V., and Henrietta, K. K. (2010). A korai nyelvi fejlődés új vizsgálóeszköze, a MacArthur-Bates Kommunikatív Fejlődési Adattár (KOFA) bemutatása és validitási vizsgálata. *Gyógypedagógiai Szemle* 38, 114–125.

Keller, K., Troesch, L. M., Loher, S., and Grob, A. (2016). The relation between effortful control and language competence—A small but mighty difference between first and second language learners. *Front. Psychol.* 7:1015. doi: 10.3389/fpsyg.2016.01015

Kopala-Sibley, D. C., Olino, T., Durbin, E., Dyson, M. W., and Klein, D. N. (2018). The stability of temperament from early childhood to early adolescence: A multi-method, multi-informant examination. *Eur. J. Personal.* 32, 128–145. doi: 10.1002/per.2151

Kort, B., Reilly, R., and Picard, R. W. (2001). "An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion" in Proceedings IEEE international conference on advanced learning technologies, 43–46.

Laake, L. M., and Bridgett, D. J. (2014). Happy babies, chatty toddlers: infant positive affect facilitates early expressive, but not receptive language. *Infant Behav. Dev.* 37, 29–32. doi: 10.1016/j.infbeh.2013.12.006

Lakatos, K., Tóth, I., and Gervai, J. (2010). Csecsemő Viselkedési Kérdőív és Kora Gyermekkori Viselkedési Kérdőív Available at: https://research.bowdoin.edu/rothbart-temperament-questionnaires/instrument-descriptions/the-early-childhood-behavior-questionnaire/

Law, J., Clegg, J., Rush, R., Roulstone, S., and Peters, T. J. (2019). Association of proximal elements of social disadvantage with children's language development at 2 years: an analysis of data from the children in focus (CiF) sample from the ALSPAC birth cohort. *Int. J. Lang. Commun. Disord.* 54, 362–376. doi: 10.1111/1460-6984.12442

Limperopoulos, C., Soul, J. S., Gauvreau, K., Huppi, P. S., Warfield, S. K., Bassan, H., et al. (2005). Late gestation cerebellar growth is rapid and impeded by premature birth. *Pediatrics* 115, 688–695. doi: 10.1542/peds.2004-1169

Lubsen, J., Vohr, B., Myers, E., Hampson, M., Lacadie, C., Schneider, K. C., et al. (2011). Microstructural and functional connectivity in the developing preterm brain. *Semin. Perinatol.* 35, 34–43. doi: 10.1053/j.semperi.2010.10.006

Ma, Q., Wang, H., Rolls, E. T., Xiang, S., Li, J., Li, Y., et al. (2022). Lower gestational age is associated with lower cortical volume and cognitive and educational performance in adolescence. *BMC Med.* 20:424. doi: 10.1186/s12916-022-02627-3

MacKay, D. F., Smith, G. C. S., Dobbie, R., and Pell, J. P. (2010). Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. *PLoS Med.* 7:e1000289. doi: 10.1371/journal.pmed.1000289

Marinopoulou, M., Billstedt, E., Lin, P. I., Hallerbäck, M., and Bornehag, C. G. (2021). Number of words at age 2.5 years is associated with intellectual functioning at age 7 years in the SELMA study. *Acta Paediatr.* 110, 2134–2141. doi: 10.1111/apa.15835

Matte, T. D., Bresnahan, M., Begg, M. D., and Susser, E. (2001). Influence of variation in birth weight within normal range and within sibships on IQ at age 7 years: cohort study. *BMJ* 323, 310–314. doi: 10.1136/bmj.323.7308.310

Nivins, S., Kennedy, E., McKinlay, C., Thompson, B., and Harding, J. E.Children with Hypoglycemia and Their Later Development (CHYLD) Study Team (2023). Size at birth predicts later brain volumes. *Sci. Rep.* 13:12446. doi: 10.1038/s41598-023-39663-9

Özçalişkan, Ş., and Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Dev. Sci.* 13, 752–760. doi: 10.1111/j.1467-7687.2009.00933.x

Pérez-Pereira, M., Fernández, P., Resches, M., and Gómez-Taibo, M. L. (2016). Does temperament influence language development? Evidence from preterm and full-term children. *Infant Behav. Dev.* 42, 11–21. doi: 10.1016/j.infbeh.2015.10.003

Posner, M. I., Rothbart, M. K., and Voelker, P. (2016). Developing brain networks of attention. *Curr. Opin. Pediatr.* 28, 720–724. doi: 10.1097/MOP.0000000000000413

Putnam, S. P., Gartstein, M. A., and Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: the early childhood behavior questionnaire. *Infant Behav. Dev.* 29, 386–401. doi: 10.1016/j.infbeh.2006.01.004

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., and Leerkes, E. (2014). Development and assessment of short and very short forms of the infant behavior questionnaire–revised. *J. Pers. Assess.* 96, 445–458. doi: 10.1080/00223891.2013.841171

Putnam, S. P., Rothbart, M. K., and Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant Child Dev.* 17, 387–405. doi: 10.1002/icd.582

Roben, C. K. P., Cole, P. M., and Armstrong, L. M. (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Dev.* 84, 891–905. doi: 10.1111/cdev.12027

Rogers, C. E., Lean, R. E., Wheelock, M. D., and Smyser, C. D. (2018). Aberrant structural and functional connectivity and neurodevelopmental impairment in preterm children. *J. Neurodev. Disord.* 10, 1–13.

Rothbart, M. K. (2007). Temperament, development, and personality. *Curr. Dir. Psychol. Sci.* 16, 207–212. doi: 10.1111/j.1467-8721.2007.00505.x

Rothbart, M. K., and Bates, J. E. (2007). Temperament. In *Handbook of child psychology*. (eds.) W. Damon and N. Eisenberg Vol. 3. Social, emotional, and personality development (New York: Wiley), pp. 105–176.

Rothbart, M. K., and Derryberry, D. (1981). "Development of individual differences in temperament" in Advances in Developmental Psychology. eds. M. E. Lamb and A. L. Brown, vol. *1* (Hillsdale, NJ: Lawrence Erlbaum), 37–86.

Salley, B. J., and Dixon, W. E. Jr. (2007). Temperamental and joint attentional predictors of language development. *Merrill-Palmer Q.* 53:7. doi: 10.1353/mpq.2007.0004

Shaw, J. C., Berry, M. J., Dyson, R. M., Crombie, G. K., Hirst, J. J., and Palliser, H. K. (2019). Reduced Neurosteroid exposure following preterm birth and its' contribution to neurological impairment: A novel avenue for preventative therapies. *Front. Physiol.* 10:599. doi: 10.3389/fphys.2019.00599

Shiner, R., Buss, K., McClowry, S., Putnam, S., and Saudino, K. (2012). What is temperament now? Assessing Progress temperament research on the twenty-fifth anniversary of Goldsmith et al. *Child Dev. Perspect.* 6, 436–444. doi: 10.1111/j.1750-8606.2012.00254.x

Shokrkon, A., and Nicoladis, E. (2022). The directionality of the relationship between executive functions and language skills: a literature review. *Front. Psychol.* 13:848696. doi: 10.3389/fpsyg.2022.848696

Snijders, V. E., Bogicevic, L., Verhoeven, M., and van Baar, A. L. (2020). Toddlers' language development: the gradual effect of gestational age, attention capacities, and maternal sensitivity. *Int. J. Environ. Res. Public Health* 17:7926. doi: 10.3390/ijerph17217926

Stene-Larsen, K., Brandlistuen, R. E., Lang, A. M., Landolt, M. A., Latal, B., and Vollrath, M. E. (2014). Communication impairments in early term and late preterm children: A prospective cohort study following children to age 36 months. *J. Pediatr.* 165, 1123–1128. doi: 10.1016/j.jpeds.2014.08.027

Stolt, S., Haataja, L., Lapinleimu, H., and Lehtonen, L. (2009). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *J. Commun. Disord.* 42, 107–123. doi: 10.1016/j.jcomdis.2008.10.002

Tang, A., Crawford, H., Morales, S., Degnan, K. A., Pine, D. S., and Fox, N. A. (2020). Infant behavioral inhibition predicts personality and social outcomes three decades later. *Proc. Natl. Acad. Sci. USA* 117, 9800–9807. doi: 10.1073/pnas.1917376117

Thomas, A., and Chess, S. (1977). Temperament and development. New York: New York University Press.

Yang, H., Yang, S., and Isen, A. M. (2013). Positive affect improves working memory: implications for controlled cognitive processing. *Cognit. Emot.* 27, 474–482. doi: 10.1080/02699931.2012.713325

Ye, Z., and Zhou, X. (2009). Executive control in language processing. *Neurosci. Biobehav. Rev.* 33, 1168–1177. doi: 10.1016/j.neubiorev.2009.03.003

frontiers | Frontiers in Human Neuroscience

TYPE Original Research
PUBLISHED 05 July 2024
DOI 10.3389/fnhum.2024.1370572

Check for updates

OPEN ACCESS

EDITED BY
Marcela Pena,
Pontificia Universidad Católica de Chile, Chile

REVIEWED BY
Alessandro Tavano,
Max Planck Society, Germany
Sari Ylinen,
Tampere University, Finland

*CORRESPONDENCE
Maria Clemencia Ortiz-Barajas
✉ mariac.ortizb@gmail.com

RECEIVED 14 January 2024
ACCEPTED 11 June 2024
PUBLISHED 05 July 2024

CITATION
Ortiz-Barajas MC (2024) Predicting language outcome at birth.
Front. Hum. Neurosci. 18:1370572.
doi: 10.3389/fnhum.2024.1370572

Predicting language outcome at birth

Maria Clemencia Ortiz-Barajas*

CNRS, IKER (URM 5478), Bayonne, France

Even though most children acquire language effortlessly, not all do. Nowadays, language disorders are difficult to diagnose before 3–4 years of age, because diagnosis relies on behavioral criteria difficult to obtain early in life. Using electroencephalography, I investigated whether differences in newborns' neural activity when listening to sentences in their native language (French) and a rhythmically different unfamiliar language (English) relate to measures of later language development at 12 and 18 months. Here I show that activation differences in the theta band at birth predict language comprehension abilities at 12 and 18 months. These findings suggest that a neural measure of language discrimination at birth could be used in the early identification of infants at risk of developmental language disorders.

KEYWORDS

EEG, theta activity, newborns, language development, predictability

1 Introduction

Most children acquire their native language(s) rapidly and effortlessly during the first years of life regardless of culture (Kuhl, 2004). However, this is not always the case. Around 7% of kindergarten children (5–6 years) (Tomblin et al., 1997) are identified as having specific language impairment (SLI, also known as developmental language disorder, DLD), a disorder characterized by the difficulty to understand and produce spoken language in the absence of other cognitive deficits. Another 5 to 17% of school children suffer from dyslexia (Shaywitz, 1998), a specific deficit in reading acquisition not attributable to low IQ, poor education or neurological damage (Ramus and Ahissar, 2012). If untreated, these disorders can have an impact on many aspects of the child's life (social, behavioral, academic), which can persist until adulthood. Nowadays, language disorders are difficult to diagnose before 3–4 years of age (Cristia et al., 2014), because diagnosis relies on behavioral criteria that are difficult to obtain early in life. However, children with learning or reading disabilities typically show deficits in speech perception earlier than when their disorder is diagnosed (Kuhl et al., 2005). Identifying measures that could allow their earlier detection is fundamental for the design of earlier interventions.

Previous research has shown that phonological deficits are often found in individuals with dyslexia and/or SLI (Ramus, 2003; Schulte-Körne and Bruder, 2010; Leonard, 2014). However, whether these deficits are speech-specific or related to basic auditory perception is still under debate (Lorusso et al., 2014; Cantiani et al., 2016). Furthermore, deficits processing auditory information in early infancy/childhood have been shown to relate to poorer later language and literacy skills in school (Molfese, 2000; Leppänen et al., 2010; Van Zuijen et al., 2013; Schaadt et al., 2015; Cantiani et al., 2016; Lohvansuu et al., 2018). Molfese (2000) found that the amplitude and latency of ERPs recorded at birth while infants listened to speech and non-speech sounds, could predict with 81% accuracy whether at 8 years of age children would be identified as normal, poor or dyslexic readers. In another newborn study, Leppänen et al.

(2010) showed that children with familial risk for dyslexia exhibited atypical processing of sound frequency at birth, as evidenced by their ERP response to tones varying in pitch. Additionally, these early differences in auditory processing were related to phonological skills and letter knowledge before school age, as well as to phoneme duration perception, reading speed and spelling accuracy in the second grade of school (Leppänen et al., 2010). Similarly, Cantiani et al. (2016) investigated Rapid Auditory Processing (RAP) abilities in 6 months-olds at risk for Language Learning Impairment (LLI), by assessing their discrimination of pairs of tones varying in frequency and duration. They found their ERPs to be atypical and to be predictive of their expressive vocabulary at 20 months (Cantiani et al., 2016). More recently, Mittag et al. (2021) used magnetoencephalography (MEG) to investigate auditory processing of white noise in 6 and 12-month-olds. They found atypical auditory responses in infants at risk for dyslexia, which predicted syntactic processing between 18 and 30 months, and as well as word production at 18 and 21 months. However, this predictive relation was not found for the control infants.

Other studies have also investigated whether early speech perception abilities relate to later language acquisition. This is supported by the native language neural commitment (NLNC) hypothesis (Kuhl, 2000, 2004) which proposes that early linguistic experience with the native language produces dedicated neural networks that influence the brain's ability to learn language. This hypothesis suggests that infants' early skills in native-language phonetic perception should predict infants' later language abilities (Kuhl, 2004). Tsao et al. (2004) tested this hypothesis by performing one of the first studies exploring the link between speech perception and language acquisition before the age of 2 years. They used the conditioned head-turn task to test 6-month-old infants on a speech discrimination task (a vowel contrast perceived by adults as native), and found significant correlations between their speech perception skills at 6 months and vocabulary measures (words understood, words produced and phrases understood) at 13, 16 and 24 months. In a follow-up study, Kuhl et al. (2005) tested a similar paradigm on 7 month-olds, this time with two conditions: one contrast from their native language, and one from a non-native language. They found that both native and non-native phonetic perception abilities were related to later measures of language outcome but in opposite directions: better native-language discrimination at 7 months was positively correlated to expressive vocabulary at 18 and 24 months, whereas better non-native-language discrimination was negatively correlated to expressive vocabulary at 18 and 24 months. These findings were supported by an electrophysiological study comparing ERP responses in 11-month-olds to native and foreign speech contrasts (Rivera-Gaxiola et al., 2005). They showed that infants who exhibited larger (more positive) P150-250 amplitudes to the foreign deviant with respect to the standard produced more words at 18, 22, 25, 27, and 30 months, than those who displayed larger (more negative) N250-550 amplitudes to the foreign deviant with respect to the standard, at the same ages. A later ERP study from the same team showed that ERP responses to native and non-native contrasts at 7 months also related to later language outcomes, again in opposing directions: greater negativity of the MMN (mismatch negativity) to native language phonetic contrasts at 7 months was associated with a larger number of words produced at 18 and 24 months, whereas more negative MMNs to non-native language phonetic contrasts at 7 months predicted fewer words produced at 24 months (Kuhl et al., 2008). Kuhl et al. (2008) suggest that increased sensitivity in the perception of native phonetic contrasts is indicative of neural commitment to the native language, whereas sensitivity to non-native contrasts reveals uncommitted neural circuitry. The ERP responses shown in these studies seem to be a reflection of this level of neural commitment, which in turn predicts language scores at later ages (Rivera-Gaxiola et al., 2005; Kuhl et al., 2008).

Previous linguistic studies focused on the discrimination of phonetic contrasts as the early measure of speech perception that could predict later language skills. A recent electroencephalography (EEG) study explored whether neural tracking of sung nursery rhymes during infancy could predict language development in infants with high likelihood of autism (Menn et al., 2022). Autistic children often show delay in language acquisition (Howlin, 2003), which is why identifying measures that could predict later language skills is relevant for this population. Menn et al. (2022) found that infants with higher speech-brain coherence in the stressed syllable rate (1–3 Hz) at 10 months showed higher receptive and productive vocabulary (words understood and words produced) at 24 months, but no relationship with later autism symptoms. They suggest that these results could reflect a relationship between infants' tracking of stressed syllables and word-segmentation skills (Menn et al., 2022), which in turn predict later vocabulary development (Junge et al., 2012; Kooijman et al., 2013). Similarly, a recent study investigating word learning at birth revealed that neonates can memorize disyllabic words so that having learnt the first syllable they can predict the word ending, and the quality of word-form learning predicts expressive language skills at 2 years (Suppanen et al., 2022).

To my knowledge, most studies investigating infant speech perception abilities as possible predictors of later language development have tested infants using phonetic contrasts (Tsao et al., 2004; Kuhl et al., 2005; Rivera-Gaxiola et al., 2005; Kuhl et al., 2008), bi-syllabic pseudo-words (Suppanen et al., 2022), and nursery rhymes (Menn et al., 2022). However, perception abilities of natural speech have rarely been used as predictors. Here, I explore the potential of using EEG measures at birth in response to naturally spoken sentences in the native language (prenatally heard) and a rhythmically different unfamiliar language as predictors of later language skills in typically-developing infants.

At birth, infants are equipped with a rich set of speech perception abilities that help them acquire language from the get-go. Some of these are universal, broad-based abilities, in place independently of what language they heard *in utero* (Ortiz Barajas and Gervain, 2021). For instance, newborns can recognize speech, and show preference for it over equally complex speech analogs (Vouloumanos and Werker, 2007). They are also able to discriminate two languages, even if they are unfamiliar to them, on the basis of their different rhythms (Mehler et al., 1988; Nazzi et al., 1998; Ramus et al., 2000), but they are unable to discriminate them if their rhythms are similar (Nazzi et al., 1998; Ramus et al., 2000). Interestingly, newborns also exhibit speech perception abilities shaped by prenatal experience with the language(s) spoken by their mother during the last trimester of pregnancy. Newborns' prenatal experience with speech mainly consists of language prosody, i.e., rhythm and melody, because maternal tissues filter out the higher frequencies, necessary for the identification of individual phonemes, but preserve the low-frequency components that carry prosody (Pujol et al., 1991). On the basis of this experience, newborns are able to recognize their native language, and prefer it over other languages (Mehler et al., 1988; Moon et al., 1993). Furthermore, it has been shown that recognizing the language heard *in utero*, goes beyond simply discriminating it from an unfamiliar one, as monolingual and bilingual newborns exhibit different patterns when presented with the same pair of rhythmically different languages: monolinguals, who are familiar with one of the languages being contrasted, discriminate them, and prefer the familiar language; while bilinguals, who are familiar with both languages being contrasted, discriminate them and show equal preference for both languages (Byers-Heinlein et al., 2010).

Building up on previous research showing that the discrimination of native/non-native phonetic contrasts predicts later language skills (Kuhl et al., 2005; Rivera-Gaxiola et al., 2005; Kuhl et al., 2008), here I explore whether newborns' ability to discriminate languages on the basis of their different rhythms could relate to language development. It has been suggested that individuals with dyslexia have difficulty extracting stimulus regularities from auditory inputs (Daikhin et al., 2017), therefore a rhythmic discrimination task, which requires detecting regularities in speech rhythm, represents a good predictor candidate for this population.

The neural mechanisms that support rhythmic discrimination in infants are not fully understood (Ortiz Barajas and Gervain, 2021). Previous infant studies have shown that low-frequency neural activity (delta and/or theta band) reflect language discrimination at birth (Ortiz-Barajas et al., 2023) and at 4.5 months (Nacar Garcia et al., 2018). Since rhythm is carried by the low-frequency components of the speech signal (Rosen, 1992), specifically the syllabic rate, it is reasonable for rhythm to be encoded by the low-frequency oscillations delta and theta. In adults, theta activity has been claimed to support the processing of syllables. This claim has mainly been based on two facts: (1) the syllabic rate of speech, roughly 4-5 Hz (Ding et al., 2017; Varnet et al., 2017), corresponds to the frequencies of the theta band (Giraud and Poeppel, 2012), and (2) brain responses in the theta band have been shown to synchronize to the speech envelope, corresponding to the slow overall amplitude fluctuations of the speech signal over time, with peaks occurring roughly at the syllabic rate (Gross et al., 2013; Molinaro et al., 2016; Vander Ghinst et al., 2016; Zoefel and VanRullen, 2016; Pefkou et al., 2017; Song and Iverson, 2018). Furthermore, newborns' neural activity has been found to track (synchronize to) the speech envelope of familiar and unfamiliar languages equally well, suggesting that envelope tracking at birth represents a basic auditory ability that helps newborns encode the speech rhythm of familiar and unfamiliar languages, supporting language discrimination (Ortiz Barajas et al., 2021; Ortiz Barajas and Gervain, 2021).

To explore the use of a neural measure of language discrimination at birth as a predictor of language outcome, I recorded EEG data from 51 full-term, healthy newborns (mean age: 2.39 days; range: 1–5 days; 20 females), born to French monolingual mothers, while they listened to naturally spoken sentences in three languages: their native language, i.e., the language heard prenatally, French, a rhythmically similar unfamiliar language, Spanish, and a rhythmically different unfamiliar language, English (Figure 1A illustrates the study design). As infants were tested within their first 5 days of life, their experience with speech was mostly prenatal. Based on the above mentioned speech perception abilities, it is reasonable to assume that participants should be able to discriminate and prefer the prenatally heard language French (syllable-timed) from English (stress-timed) based on their different rhythms, but not from Spanish (syllable-timed), as they are rhythmically similar. Given that stimuli are presented in 7 min blocks, and languages are not contrasted closely, I hypothesize that for language recognition to take place, the newborn brain compares each language to the long-term representation it has formed from prenatal experience, in order to recognize familiar features. This hypothesis is supported by one recent study from my team investigating the role of prenatal experience on long-range temporal correlations (LRTC) using a superset of the EEG dataset used here (Mariani et al., 2023), revealing that the newborn brain exhibits stronger correlations in the theta band after being exposed to the native language (French) than to the rhythmically similar (Spanish) and the rhythmically different (English) unfamiliar languages, indicating the early emergence of brain specialization for the native language. These findings support the hypothesis that participants from this study did recognize the prenatally heard language, and that such recognition is reflected by theta activity.

I assessed language rhythmic discrimination at birth as the neural activation difference between the native language (French) and the rhythmically different unfamiliar language (English). I expect this discrimination measure to reflect neural commitment to the native language and in turn to predict language scores at later ages as follows: higher discrimination measures should predict higher language scores, reflecting commitment to the native language, whereas lower discrimination measures should predict lower language scores, reflecting uncommitted neural circuitry. Spanish sentences were presented in this experiment as part of a larger project investigating speech perception at birth. However, here I do not present results for Spanish, as I focus on the rhythmic discrimination of the native language (French) and the rhythmically different unfamiliar language (English).

To explore the potential use of this neural discrimination measure as a predictor of language outcome, participants were followed longitudinally in order to describe their developmental trajectory, and to look at their individual variability. Figure 2 displays the timeline of the longitudinal study: EEG data were recorded at birth, followed by the collection of information about the participants' vocabulary size at 12 and 18 months using the MacArthur-Bates Communicative Developmental Inventory (CDI) questionnaires. The participants' receptive and expressive vocabulary sizes were estimated from the CDI questionnaires, in order to track their language development, and relate it to their neural measures at birth. To assess the predictive role of language discrimination at birth on later language abilities, I conducted a path analysis including newborns' performance at discriminating the native language (French) from a rhythmically different unfamiliar one (English), and their measures of vocabulary size at 12 and 18 months (number of words understood and number of words produced). A total of 51 infants contributed neural data at birth, and 35 of them contributed with at least one CDI questionnaire at the subsequent ages. Vocabulary data were collected from 27 participants at 12 months, and 30 participants at 18 months (Supplementary Table S1).

2 Materials and methods

The EEG data from this study were acquired as part of a larger project that aimed to investigate speech perception during the first two years of life. One previous publication presented a superset of the current dataset (47 participants) evaluating speech envelope tracking

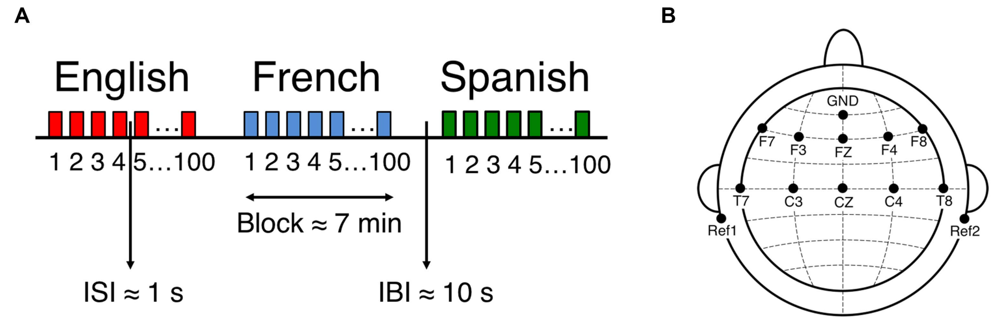


FIGURE 1
EEG experimental setup and design. **(A)** Experiment block design. ISI: Interstimulus interval, IBI: Interblock interval. **(B)** Location of recorded channels according to the international 10–20 system. Figure adapted from Ortiz Barajas et al. (2021).

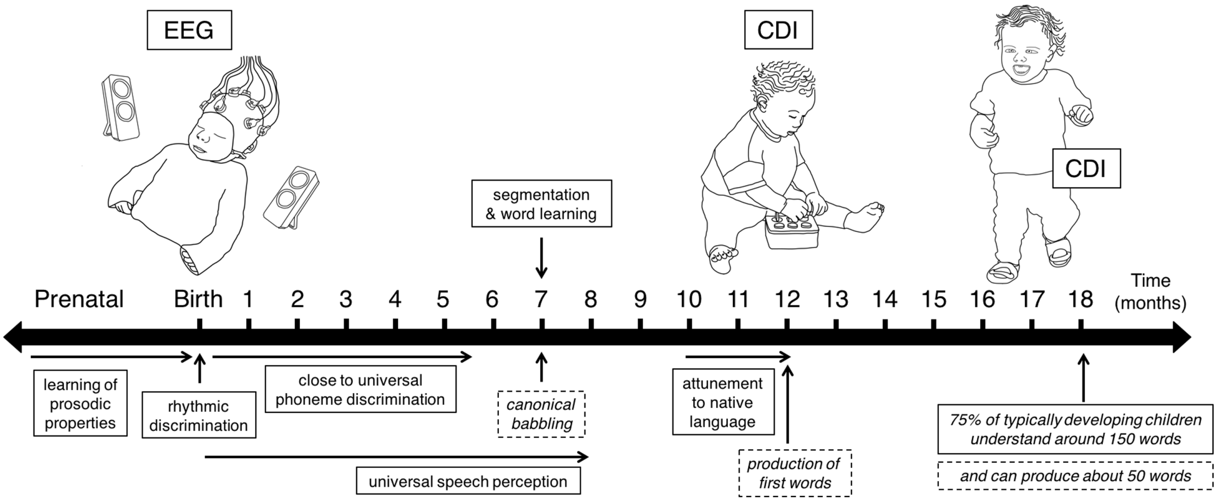

EEG
CDI
CDI
segmentation & word learning
Prenatal
Birth
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Time (months)
learning of prosodic properties
rhythmic discrimination
close to universal phoneme discrimination
canonical babbling
universal speech perception
attunement to native language
production of first words
75% of typically developing children understand around 150 words
and can produce about 50 words

FIGURE 2
Study timeline indicating the time points when longitudinal data were collected, and displaying some of the developing speech perception (solid boxes) and production (dashed boxes) abilities children exhibit during the first 18 months of life.

in newborns and 6-month-olds (Ortiz Barajas et al., 2021). A second publication, evaluating the role of neural oscillations during speech processing at birth, presented a subset (40 participants) of the initial publication (Ortiz-Barajas et al., 2023). A third publication, exploring changes in neural dynamics at birth, presented a subset (33 participants) of the initial publication (Mariani et al., 2023). These three publications evaluated different hypotheses, therefore analyzing different aspects of the data, which explains the differences in sample size. The EEG dataset used in this manuscript (29 participants) represents a subset of that used in the previous publications, as not all participants contributed with vocabulary measures at 12 and 18 months The processed EEG data that support the findings of this study have been deposited in the OSF repository https://osf.io/4w69p.

2.1 Participants

The protocol for this study was approved by the CER Paris Descartes ethics committee of the Paris Descartes University (currently, Université Paris Cité). All parents gave written informed consent prior to participation, and were present during the testing session.

For the first measure of the study, newborns were recruited at the maternity ward of the Robert-Debré Hospital in Paris, where they were tested during their hospital stay. The inclusion criteria were: (i) being full-term and healthy, (ii) having a birth weight > 2,800 g, (iii) having an Apgar score > 8, (iv) being maximum 5 days old, and (v) being born to French native speaker mothers who spoke this language at least 80% of the time during the last trimester of the pregnancy according to self-report. A total of 54 newborns took part in the EEG experiment. However, 3 participants failed to complete the recording due to fussiness and crying ($n = 2$), or technical problems ($n = 1$); and were thus excluded from the longitudinal study. The remaining **51 newborns** (20 girls, 31 boys; age 2.39 ± 1.17 d; range 1–5 d) were followed longitudinally by means of the CDI questionnaires.

For the second and third measures of the study, parents of the infants who contributed with EEG data at birth were requested to fill out vocabulary questionnaires when their children turned 12 and 18 months. As it is often the case in longitudinal studies, some of the participants did not contribute measures to all the assessments. A total of **35 participants** contributed at least one vocabulary questionnaire (at 12 and/or 18 months), of which 27 participants contributed CDI data at 12 months, and 30 participants at 18 months. Supplementary Table S1 presents the list of participants and the data points that they contributed longitudinally.

From the 35 participants who contributed EEG recordings and vocabulary data, 6 participants were excluded due to bad EEG data quality in at least one of the language conditions of interest (French and English). Therefore, a final sample of **29 participants** contributed good quality EEG data at birth, and were included in the prediction analyses: a subset of **22 participants** contributed CDI data at 12 months, while a subset of **27 participants** contributed CDI data at 18 months.

2.2 Procedure

Figure 2 presents a timeline highlighting the three ages when data were collected: EEG data at birth, and CDI data at 12 months and 18 months.

For the first measure of the study, newborns were presented with naturally spoken sentences in three languages while their neural activity was simultaneously recorded using EEG. The recording sessions were conducted in a dimmed, quiet room at the Robert-Debré Hospital in Paris, while newborns were comfortably asleep or at rest in their hospital bassinets. The stimuli were delivered bilaterally through two loudspeakers positioned on each side of the bassinet (Figure 2, EEG recording at birth) using the experimental software E-Prime. The sound volume was set to a comfortable conversational level (~65–70 dB). Participants were divided into 3 groups, where each group heard a different set of sentences: 17 newborns heard set1, 17 newborns heard set2, and 17 newborns heard set3. Supplementary Table S2 presents the three sets of sentences used in the study. Participants were presented with one sentence per language (French, English, Spanish), which was repeated 100 times to ensure sufficiently good data quality. The experiment consisted of 3 blocks, each block containing the 100 repetitions of the test sentence in a given language, each block thus lasted around 7 min. An interstimulus interval of random duration (between 1 and 1.5 s) was introduced between sentence repetitions, and an interblock interval of 10 s was introduced between language blocks (Figure 1A). The order of the languages was pseudo-randomized and approximately counterbalanced across participants. The entire recording session lasted about 21 min.

For the second and third measures of the study, parents were requested to fill out the French version of the MacArthur-Bates Communicative Developmental Inventory (CDI) questionnaires (Kern, 2007) when their child turned 12 and 18 months. In each case they were asked to return the questionnaire before their child turned 13 and 19 months respectively, to ensure that the measurement would not exceed these age limits. In order to make it easier for parents to complete the questionnaires, I provided them with the short version of the CDI, which is one page long. The short version CDI has been shown to be as reliable as the original version for the English CDI (Floccia et al., 2018). For the measurement at 12 months I used the *Words and Gestures* CDI, which inquires about the child's babbling skills, provides a list of 83 words for parents to indicate whether the child understands them and spontaneously produces them, and a list of 25 gestures for them to indicate if the child makes them (e.g., shake the head to say no). For the measurement at 18 months I used the *Words and Sentences* CDI, which provides a list of 97 words for parents to indicate whether the child understands them and spontaneously produces them, and inquires whether the child has started to combine words together.

2.3 Stimuli

At birth, I presented infants with sentences in three languages: their native language (French), a rhythmically similar unfamiliar language (Spanish), and a rhythmically different unfamiliar language (English). The stimuli consisted of sentences taken from the story Goldilocks and the Three Bears. Sentences were divided in three sets, where each set comprised the translation of a single utterance into the 3 languages (English, French and Spanish). For instance, set 1 contained the following three sentences: *The bears lived all together in a beautiful house* (English); *Les ours habitaient tous ensemble dans une maison* (French); *Los osos vivían juntos en una casa* (Spanish). The translations were slightly modified by adding or removing adjectives (or phrases) from certain sentences in order to match the duration and syllable count across languages within the same set. All sentences were recorded in mild infant-directed speech by a female native speaker of each language (a different speaker for each language), at a sampling rate of 44.1 kHz. There were no significant differences between the sentences in the three languages in terms of minimum and maximum pitch, pitch range and average pitch. Supplementary Table S2 presents detailed information about the 9 sentences used as stimuli (i.e., duration, syllable count, pitch), and Supplementary Figures S1, S2 display the sentences' time-series, and frequency spectra, respectively. Additionally, the amplitude and frequency modulation spectra as defined by Varnet et al. (2017) are presented in the Supplementary Figure S3. Utterances were found to be similar in every spectral decomposition. The intensity of all recordings was adjusted to 77 dB.

2.4 EEG data acquisition

EEG data were recorded at birth with active electrodes and an acquisition system from Brain Products (actiCAP & actiCHamp, Brain Products GmbH, Gilching, Germany). A 10-channel layout was used to acquire cortical responses from the following scalp positions: F7, F3, FZ, F4, F8, T7, C3, CZ, C4, T8 (Figure 1B). These recording locations were chosen in order to include those where auditory and speech perception related neural responses are typically observed in infants (Stefanics et al., 2009; Tóth et al., 2017) (channels T7 and T8 used to be called T3 and T4 respectively). An additional electrode was placed on each mastoid for online reference, and a ground electrode was placed on the forehead. Data were referenced online to the average of the two mastoid channels, and they were not re-referenced offline. Data were recorded at a sampling rate of 500 Hz, and online filtered with a high cutoff filter at 200 Hz, a low cutoff filter at 0.01 Hz and an 8 kHz (−3 dB) anti-aliasing filter. The electrode impedances were kept below 140 kΩ.

2.5 EEG data analysis

The EEG data were processed using custom Matlab® scripts. To extract the low-frequency activity of interest (delta and theta), the continuous EEG signals were band-pass filtered between 1 and 8 Hz with a zero phase-shift Chebyshev filter. The filtered signals were then segmented into a series of 2,560-ms long epochs. Each epoch started 400 ms before the utterance onset (corresponding to the pre-stimulus baseline), and contained a 2,160 ms long post-stimulus interval (corresponding to the duration of the shortest sentence). All epochs were submitted to a three-stage rejection process to exclude the contaminated ones: (1) Epochs with peak-to-peak amplitude exceeding 150 μV were rejected. (2) Epochs with a standard deviation (SD) higher than 3 times the mean SD of all non-rejected epochs, or lower than one-third the mean SD were rejected. (3) The remaining epochs were visually inspected to remove any residual artifacts. Participants who had less than 20 remaining epochs in a given condition after epoch rejection were excluded. From the 35 participants who contributed EEG and CDI data, 6 were excluded due to bad data quality resulting in an insufficient number of non-rejected epochs in one of the language conditions of interest (French and English). Therefore, 29 participants contributed good quality EEG data for the French and English conditions (Supplementary Table S1). The included participants contributed on average 41 epochs (SD: 13.14; range: 20–79) for French, and 35 epochs (SD: 10.69; range: 20–62) for English. The number of non-rejected epochs from the 29 participants were submitted to a paired samples t-test (two-tail), and it yielded no significant differences between the two language conditions [$p = 0.082$].

The non-rejected epochs were subjected to time-frequency analysis to uncover stimulus-evoked oscillatory responses using the Matlab® toolbox 'WTools' (Parise and Csibra, 2013). With this toolbox, a continuous wavelet transform of each non-rejected epoch was performed using Morlet wavelets (number of cycles 3.5) at 1 Hz intervals in the 1–8 Hz range. The full pipeline is described in detail in (Csibra et al., 2000; Parise and Csibra, 2013). Briefly, complex Morlet Wavelets are computed at steps of 1 Hz with a sigma of 3.5. The real and the imaginary parts of the wavelets are computed separately as cos and sin components, respectively. The signal is then convoluted with each wavelet. The absolute value of each complex coefficient is then computed. This process resulted in a time-frequency map of spectral amplitude values (not power) per epoch.

Time-frequency transformed epochs were then averaged for French and English separately. To remove the distortion introduced by the wavelet transform, the first and last 200 ms of the averaged epochs were removed, resulting in 2,160 ms long segments, including 200 ms before and 1,960 ms after stimulus onset. The averaged epochs were then baseline corrected using the mean amplitude of the 200 ms pre-stimulus window as baseline, subtracting it from the whole epoch at each frequency. This process resulted in a time-frequency map of spectral amplitude values per condition and channel, at the participant level. The group mean (29 participants) of these time-frequency maps for channel F4 is presented in Figure 3A as an example.

Language discrimination between French (the native language) and English (the rhythmically different unfamiliar language) was assessed by submitting the spectral amplitude values from their time-frequency responses to paired-samples t-tests (two-tailed). Figure 3B displays the *P*-map for this analysis in channel F4, and Figure 3C highlights the time-frequency regions where the absolute *T*-values for this comparison exceed the critical threshold ($|T\text{-value}| > 2.048$). Cluster-level statistics were calculated, and nonparametric statistical testing was performed by calculating the *p*-value of the clusters under the permutation distribution (Maris and Oostenveld, 2007), which was obtained by permuting the language labels in the original dataset 1,000 times. The sample size for these analyses was 29 participants.

Once significant clusters, i.e., time-frequency regions where neural responses to French and English are significantly different, had been identified (Figure 3C), the mean spectral amplitude in the cluster's region was computed for each language separately. A neural measure of language discrimination was obtained by calculating the mean amplitude difference between the two language conditions (French – English) in the region of the significant cluster. This process yielded one discrimination measure per participant, which represents the candidate predictor of later language skills.

2.6 Predicting language outcome

Measures of language development were obtained from the CDI questionnaires collected at 12 and 18 months. Receptive vocabulary was assessed as the number of *words understood*, and expressive vocabulary was assessed as the number of *words produced* at each given age. Data from one infant were removed from analysis because expressive vocabulary at 12 months was larger than 3 SDs above the mean of the same score in the group.

To investigate the predictive role of language discriminations at birth on later language development, I conducted a path analysis considering the neural activation difference between French and English as the independent variable, and vocabulary measures at 12 and 18 months (words understood and words produced) as dependent variables. Figure 4 depicts the relationships that were assessed. Additionally, to evaluate whether CDI data reliably tracks infants' vocabulary growth, the predictive role that vocabulary measures at 12 months have on vocabulary measures at 18 months was also evaluated. Three hypothesis were tested here: (i) neural data at birth can predict vocabulary skills at 12 months; (ii) neural data at birth can predict vocabulary skills at 18 months; (iii) vocabulary skills at 12 months can predict vocabulary skills at 18 months. Two comparisons evaluated each hypothesis: one predicting the number of words understood and another one the number of words produced. The Bonferroni correction was applied to adjust the original alpha value ($\alpha = 0.05$) and correct for the multiple comparisons evaluating the same hypothesis ($n = 2$). This resulted in the adjusted alpha value ($\alpha = 0.025$), which was used to evaluate the obtained results. To test for outliers, data's residuals and influential cases were investigated. Residuals were evaluated by assessing heteroskedasticity with the White test and the Breusch-Pagan test. To identify possible influential cases, Cook's distance and leverage values were computed.

Additionally, to assess the relationship between language skills at a given age, Pearson's correlation coefficients (two-tailed) were computed between the number of words understood and the number of words produced at 12 and 18 months, separately. All statistical analyses were carried out with SPSS 29 (IBM).

3 Results

3.1 EEG data analysis

A time-frequency response to French and English was obtained for the 29 participants who contributed at least 20 non-rejected epochs per condition. Figure 3A presents the group mean time-frequency

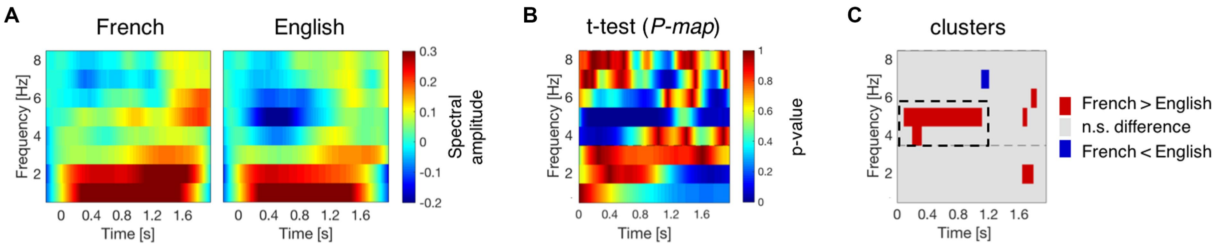


FIGURE 3
Neural activation during speech processing at birth. **(A)** Average time-frequency response to French and English at channel F4. The time-frequency maps illustrate the mean spectral amplitude per condition from 1 to 8 Hz. The color bar to the right of the figure shows the spectral amplitude scale of the maps. **(B)** P-map obtained by submitting the time-frequency responses to French and English to paired-samples t-tests (two-tailed). **(C)** Time-frequency regions where the absolute *T*-values exceed the critical threshold (|*T*-value| > 2.048). Red regions indicate higher activation for French, while blue regions indicate higher activation for English. The dashed rectangular box indicates the cluster exhibiting significant differences between French and English at channel F4.

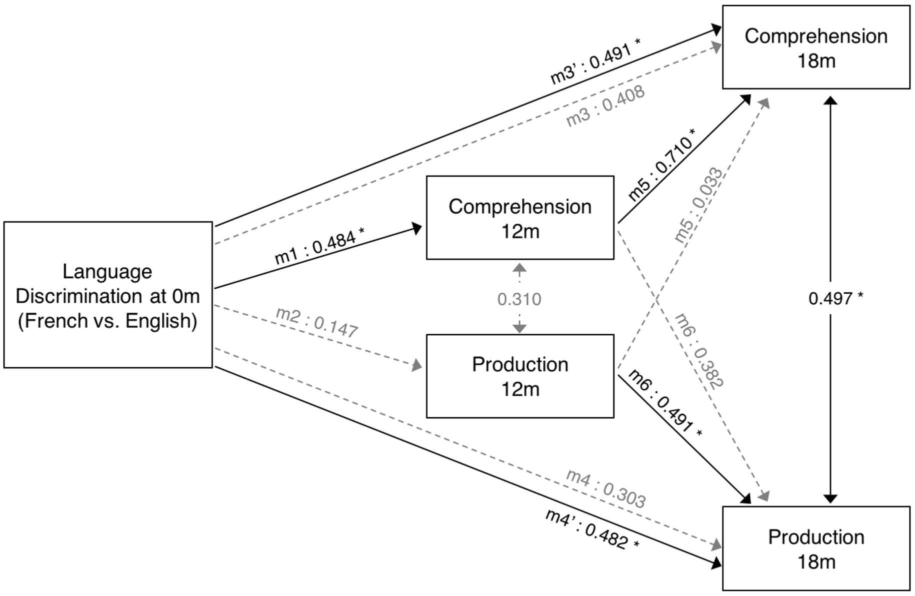


FIGURE 4
Diagram of path analysis assessing the relationship between language measures from birth to 18 months. Models 1 to 4 assess the predictive role of language discrimination at birth on vocabulary measures at 12 and 18 months. Models 5 and 6 assess the predictive relationship between vocabulary measures at 12 and 18 months. The single-ended arrows represent the predictive relationships under evaluation, and the double-ended arrows illustrate the non-causal relationships between variables (correlation). The solid black arrows illustrate significant relationships, while dashed gray arrows illustrate non-significant relationships.

maps for the two conditions at channel F4. Neural activation differences between French and English were assessed by submitting their time-frequency responses to permutation testing involving paired-samples t-tests (two-tailed). Figure 3B presents the P-map for this comparison, and Figure 3C highlights the time-frequency regions where differences take place at channel F4. Supplementary Figure S4 presents the results for all channels.

A significant cluster revealing neural activation differences between French and English was found at channel F4 ranging from 4 to 5 Hz [t (28) = 862,17; p = 0.02]. In the cluster region, neural responses exhibit higher activation for French (the native language) than for English (the rhythmically different unfamiliar language), mainly at 5 Hz, during the first half of the sentences. The maximum effect size, partial eta-squared (n_p^2), for this significant cluster in channel F4 is 0.9794. These results were obtained for a subset of participants (n = 29) from the original publication (n = 40) investigating neural oscillations at birth (Ortiz-Barajas et al., 2023), therefore they reveal the same findings: theta activity in the human newborn brain is sensitive to rhythmic differences across languages as it can successfully distinguish between the rhythmically different languages, English, a stress-timed language, and French, a syllable-timed language (Ramus et al., 1999).

The language discrimination measure, defined as the difference in neural activation between French and English, ranged from −0.422 to 0.896 (mean = 0.204, SD = 0.298). Supplementary Table S3 presents the language discrimination measure (Discrimination_Theta_F4_0m) for the 29 included participants.

3.2 Predicting language outcome

Measures of language development were obtained by collecting information about children's receptive and expressive vocabulary at 12 and 18 months. Supplementary Table S3 presents the vocabulary measures (words understood, words produced) for the 29 included participants.

A path analysis was conducted to evaluate the predictive relationship between language discrimination at birth and language skills at 12 and 18 months. Additionally, the predictive relationship between language measures at 12 and 18 months was also evaluated to assess infant's vocabulary growth. Table 1 presents the results of the linear regression models, and Figure 4 depicts the standardized estimates of the path coefficients.

When evaluating the predictive role of language discrimination at birth, a significant path coefficient was found for language comprehension at 12 months (Beta = 0.484, $p = 0.023$, model 1). This significant linear relationship is illustrated in Figure 5A. In contrast, language discrimination at birth did not predict production skills at 12 months (Beta = 0.147; $p = 0.513$, model 2), nor language comprehension at 18 months (Beta = 0.408; $p = 0.035$, model 3), nor language production at 18 months (Beta = 0.303; $p = 0.124$, model 4). Figures 5B–D illustrate the non-significant linear regressions evaluated for language production at 12 months, as well as for language comprehension and production at 18 months, respectively. When evaluating for outliers, the model assessing the prediction of production skills at 18 months (model 4) exhibited heteroskedasticity according to Breusch-Pagan test (Chi-Square = 4.924, $p = 0.026$). Additionally, 3 influential cases were identified in the models assessing the prediction of language skills at 18 months (models 3 and 4) due to having leverage values greater than twice the average (leverage values = 0.21, 0.16, and 0.19; average value = 0.07). Supplementary Table S3 highlights the influential cases in red, and Figures 5C,D identifies them with blue circles. As post-hoc analyses, the 3 influential cases were removed and regression models were re-calculated (Table 2, models 3′ and 4′). Language discrimination at birth was found to significantly predict language comprehension (Beta = 0.491; $p = 0.015$, model 3′) and language production (Beta = 0.482; $p = 0.017$, model 4′) at 18 months, after the 3 influential cases were removed. Figures 5E,F illustrate how these linear regressions, excluding the influential cases, predict language skills at 18 months (models 3′ and 4′).

To assess whether language abilities at 12 months are representative of the developmental path that language acquisition will follow, I assessed the predictive relationship between vocabulary measures at 12 and 18 months. The results show that language comprehension at 18 months is significantly predicted by language comprehension at 12 months (Beta = 0.710; $p = 0.001$, model 5), but not by language production at 12 months (Beta = 0.033; $p = 0.859$, model 5). Similarly, language production at 18 months is significantly predicted by language production at 12 months (Beta = 0.491; $p = 0.015$, model 6), but not by language comprehension at 12 months (Beta = 0.382; $p = 0.049$, model 6). Table 2 presents post-hoc regression analyses (models 5′ and 6′) removing the non-significant predictors from models 5 and 6 (Table 1). Figures 6A,B illustrate the significant linear relationship between vocabulary measures at 12 and 18 months. Furthermore, Figures 6C,D illustrate the developmental trajectories for word comprehension and word production respectively, exhibiting a vocabulary growth that is consistent across participants. These results confirm that CDI questionnaires provided reliable measures of language growth in this sample.

When evaluating the relationship between language skills at 12 and 18 months, a significant positive correlation was observed between language comprehension and production at 18 months ($r = 0.497$, $p = 0.008$, $n = 27$), but not between vocabulary measures at 12 months ($r = 0.310$, $p = 0.160$, $n = 22$). These non-predictive relationships are depicted in Figure 4 with double-sided arrows.

4 Discussion

The current study investigated whether a neural measure of language discrimination at birth, defined as the neural activation difference found when processing the prenatally heard language (French) and a rhythmically different unfamiliar language (English), could be used as predictor of language outcome. Results revealed that differences in theta activity at birth, claimed to reflect rhythmic discrimination of French and English predict language comprehension at 12 months. Furthermore, post-hoc analyses after removing 3 outliers from the vocabulary data at 18 months revealed that language discrimination at birth also predicts language comprehension and production at 18 months. These findings suggest that the ability to recognize the native language and discriminate it

TABLE 1 Regression models assessing the prediction of language skills at 12 and 18 months.

Model	Dependent variable	*R*	*R* square	df	*F*	Sig	Independent variable	Beta	Sig	Sample size
1	Comprehension_12m	0.484	0.234	1	6.110	**0.023***	Discrimination_0m	**0.484***	**0.023***	22
2	Production_12m	0.147	0.022	1	0.444	0.513	Discrimination_0m	0.147	0.513	22
3	Comprehension_18m	0.408	0.167	1	4.996	0.035	Discrimination_0m	0.408	0.035	27
4	Production_18m	0.303	0.092	1	2.534	0.124	Discrimination_0m	0.303	0.124	27
5	Comprehension_18m	0.725	0.525	2	9.398	**0.002***	Comprehension_12m; Production_12m	**0.710**; 0.033	**0.001***; 0.859	20
6	Production_18m	0.741	0.549	2	10.366	**0.001***	Comprehension_12m; Production_12m	0.382; **0.491**	0.049; **0.015***	20

Models 1 to 4 use language discrimination at birth as predictor, while models 5 and 6 explore the predictive relationship between vocabulary measures at 12 and 18 months. The alpha value for these tests is $\alpha = 0.025$.

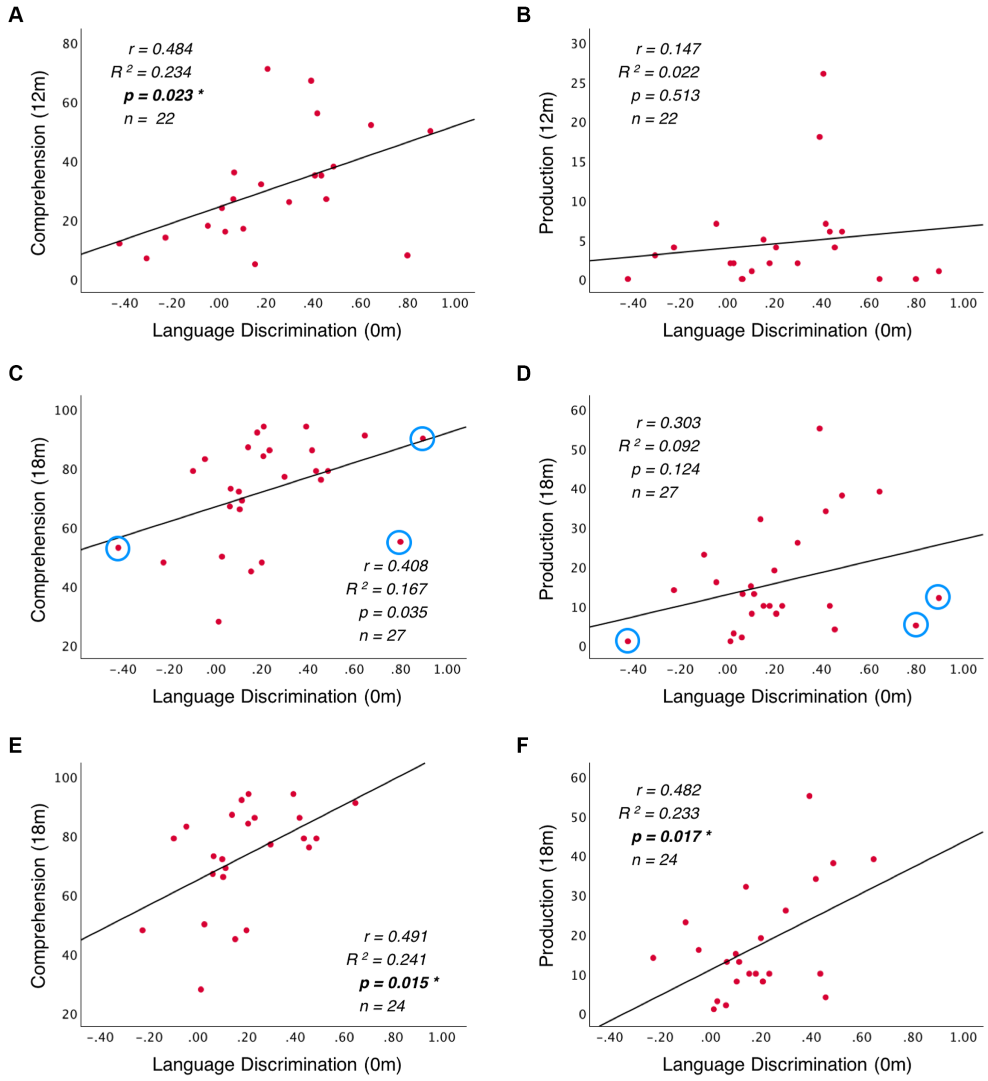


FIGURE 5

Linear regression between language discrimination at birth and: **(A)** language comprehension at 12 months, **(B)** language production at 12 months, **(C)** language comprehension at 18 months, and **(D)** language production at 18 months. Panels **(E)** and **(F)** illustrate the linear regressions from **(C)** and **(D)** respectively, after removing the outliers highlighted with blue circles.

TABLE 2 *Post-hoc* regression models.

Model	Dependent variable	*R*	*R* square	df	*F*	Sig	Independent variable	Beta	Sig	Sample size
3′	Comprehension_18m	0.491	0.241	1	6.988	**0.015***	Discrimination_0m	**0.491**	**0.015***	24
4′	Production_18m	0.482	0.233	1	6.676	**0.017***	Discrimination_0m	**0.482**	**0.017***	24
5′	Comprehension_18m	0.724	0.524	1	19.830	**<0.001***	Comprehension_12m	**0.724**	**<0.001***	20
6′	Production_18m	0.656	0.431	1	13.622	**0.002***	Production_12m	**0.656**	**0.002***	20

Models 3′ and 4′ represent post-hoc linear regressions of models 3 and 4 (Table 1) after removing 3 influential cases.
Model 5′ and 6′ represent post-hoc linear regressions of models 5 and 6 (Table 1) after removing the non-significant predictors.
The alpha value for these tests is $\alpha = 0.025$.

from a rhythmically different unfamiliar language at birth can predict later language development.

When newborns discriminate their native language from a rhythmically different unfamiliar language, they perform two tasks: (1) they discriminate the acoustic features that differentiate both rhythmic classes, and (2) they recognize the features of their native language (heard *in utero*). Therefore, a language discrimination task involving the native language, is different from a discrimination task involving two unfamiliar languages (Byers-Heinlein et al., 2010). Here, newborns discriminated their native language (French) from

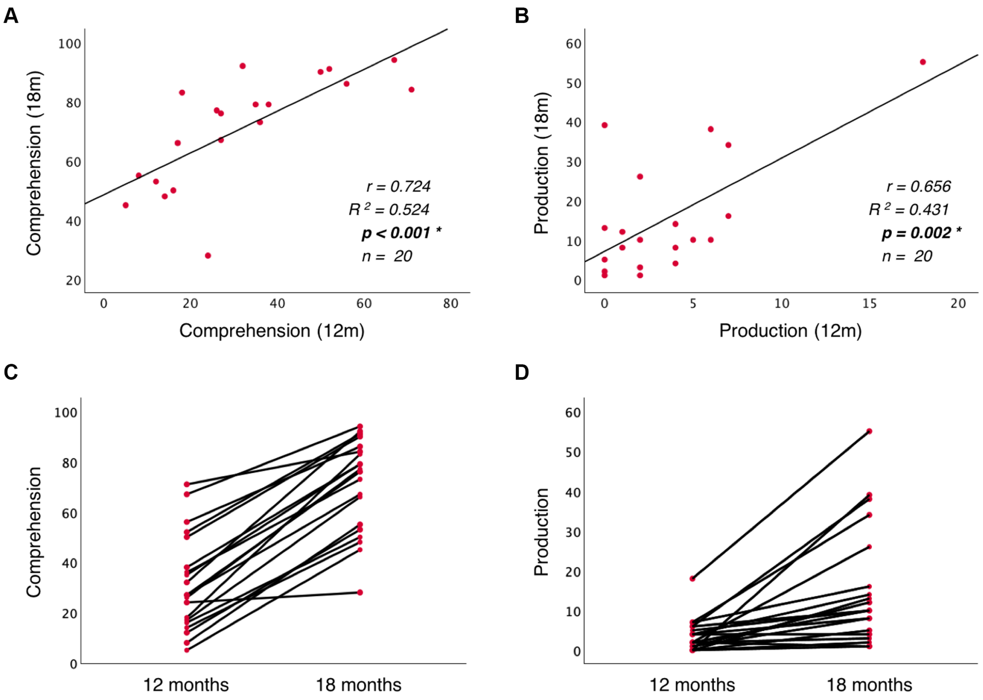


FIGURE 6
Panels **(A)** and **(B)** illustrate the linear relationship between vocabulary measures at 12 and 18 months, while **(C)** and **(D)** illustrate the trajectory of vocabulary growth from 12 to 18 months at the participant level.

a rhythmically different unfamiliar language (English). This discrimination was reflected by activation differences in the theta band such that, at the group level, higher theta activation was exhibited for French that for English. Such activation differences could have been originated from different activation profiles: (i) activation for French and no activation for English, (ii) no activation for French and suppression for English, or (iii) activation for French and English, with higher activation for French. Findings from my previous study investigating neural oscillations during speech processing at birth, using a superset of the current dataset (Ortiz-Barajas et al., 2023), revealed that theta activity during French and English processing was higher than at rest, pointing in the direction of situation (iii), where activation for both languages takes place, and differences originate from higher activation to French. This supports the hypothesis that the modulation of theta activity might be one way for the newborn brain to encode speech rhythm (regardless of language familiarity), aiding in the discrimination of rhythmically different languages, and the recognition of the native language (Ortiz Barajas et al., 2021; Ortiz Barajas and Gervain, 2021; Ortiz-Barajas et al., 2023).

Furthermore, theta activity in the newborn brain has also been found to exhibit increased long-range temporal correlations after stimulation with the prenatally heard language, indicating the early emergence of brain specialization for the native language (Mariani et al., 2023). If stronger theta activation for French (as compared to English) reflects brain specialization for the native language, a discrimination measure reflecting this activation difference should predict infants' later language abilities. Results from the current study revealed that larger discrimination measures at birth predict higher vocabulary measures at 12 and 18 months, while lower discrimination measures predict lower later language skills. These findings suggest that language discrimination at birth represents an early measure of neural commitment to the native language that predicts its later developmental trajectory. Theta activity has been argued to support the processing of syllabic units in adults (Ghitza and Greenberg, 2009). Findings from infant studies point in the same direction, as theta activity has been found to underlie language discrimination (Nacar Garcia et al., 2018; Ortiz-Barajas et al., 2023), suggesting that it might encode speech rhythm. Additionally, theta activity in the infant brain has also been found to synchronize to the speech envelope (Ortiz Barajas et al., 2021), and the speech envelope carries rhythm (Rosen, 1992). Since both the speech envelope and rhythm correlate with syllabic rate (Varnet et al., 2017; Zhang et al., 2023), it is reasonable to suggest that theta activity might encode syllabic units, and rhythm, by extracting relevant features from the speech envelope already at birth. If this is the case, the predictive power of the language discrimination measure at birth could be due to theta activity favoring the encoding of syllables in French (a syllable-timed language), which in turn would favor later word learning. This claim is supported by previous studies showing that tracking of stressed syllables at 10 month (Menn et al., 2022) and learning of disyllabic words at birth (Suppanen et al., 2022) predict language abilities at 2 years. These results taken together suggest that syllable encoding supports word-segmentation and word learning, which in turn support language development. Newborns have been shown to have a universal sensitivity to syllables (Sansavini et al., 1997; Ortiz Barajas and Gervain, 2021), however, it cannot be established whether the larger theta activity observed here on prenatally French-exposed newborns reflects good encoding of syllabic units due to this inherent (universal) ability, or whether prenatal experience with French (a syllable-timed language) has strengthened this sensitivity. Future research testing the same stimuli on prenatally English-exposed newborns (English being a stress-timed language) will shed light on the role of theta activity on syllable encoding at birth.

When exploring the predictive role of language discrimination at birth on later language skills, a significant linear relationship was found with language comprehension at 12 months, as well as with language comprehension and production at 18 months (after removing outliers). These results (Figure 4) depict a language trajectory that is coherent and consistent along development: language scores at any given age predict language scores at a subsequent age. However, one exception was found for language production at 12 months, which was not predicted by language discrimination at birth. This could be because at 12 months, language production is at its very beginning (Figure 2) and individual variability is low (Figures 5B, 6D). This suggests that measuring language production at 12 months is too early to describe the language developmental trajectory of each individual. This is supported by the fact that language production at 12 months is not correlated with language comprehension at the same age, which on the contrary, does describe the language trajectory of participants. However, language production undergoes an accelerated growth around 18 months (vocabulary spurt) (Kuhl, 2007), and becomes a better indicator of the language trajectory, as it correlates with language comprehension at the same age, and it can be predicted by language discrimination at birth.

In summary, the current study revealed a predictive relationship between a measure of theta activity during language discrimination at birth and later language outcome that merits further exploration and confirmation in future studies. These results point toward a developmental scenario in accordance with theoretical predictions as well as empirical findings: prenatal experience with speech mainly consists of language prosody, as maternal tissues filter out the higher frequencies, but preserve the low-frequency components that carry prosody (Pujol et al., 1991). Having experience with the prosody of their mother's language, allows newborns to identify it and discriminate it from other rhythmically different languages at birth. Low frequency neural activity (delta and theta) has been found to support speech processing at birth, and to reflect rhythmic language discrimination, suggesting that it reflects the processing of prosody (Ortiz-Barajas et al., 2023). Considering the relevance of low frequency neural activity in speech processing at birth, as well as in adulthood (Giraud and Poeppel, 2012; Meyer, 2018), it is reasonable to hypothesize that it has a central role in language acquisition, as not only it describes speech processing at the time of measurement, it also seems to describe the language developmental trajectory a child might follow.

Data availability statement

The processed EEG data that support the findings of this study have been deposited in the OSF repository https://osf.io/4w69p.

Ethics statement

The studies involving humans were approved by the CER Paris Descartes ethics committee of the Paris Descartes University (currently, Université Paris Cité). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

MCO-B: Investigation, Formal analysis, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work and publication have received the financial support of the AcAMEL project, financed by the Nouvelle Aquitaine Region and the CAPB (Communauté d'Agglomération du Pays Basque).

Acknowledgments

I would like to thank the Robert Debré Hospital for providing access to the newborns, and all the families and their babies for their participation in this study. I sincerely acknowledge Lucie Martin and Anouche Banikyan for their help with infant testing; as well as Judit Gervain for valuable comments on the analysis and the manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2024.1370572/full#supplementary-material

References

Byers-Heinlein, K., Burns, T. C., and Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychol. Sci.* 21, 343–348. doi: 10.1177/0956797609360758

Cantiani, C., Riva, V., Piazza, C., Bettoni, R., Molteni, M., Choudhury, N., et al. (2016). Auditory discrimination predicts linguistic outcome in Italian infants with and without familial risk for language learning impairment. *Dev. Cogn. Neurosci.* 20, 23–34. doi: 10.1016/j.dcn.2016.03.002

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., and Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Dev.* 85, 1330–1345. doi: 10.1111/cdev.12193

Csibra, G., Davis, G., Spratling, M. W., and Johnson, M. H. (2000). Gamma oscillations and object processing in the infant brain. *Science* 290, 1582–1585. doi: 10.1126/science.290.5496.1582

Daikhin, L., Raviv, O., and Ahissar, M. (2017). Auditory stimulus processing and task learning are adequate in dyslexia, but benefits from regularities are reduced. *J. Speech Lang. Hear. Res.* 60, 471–479. doi: 10.1044/2016_JSLHR-H-16-0114

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

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., et al. (2018). I: INTRODUCTION. *Monogr. Soc. Res. Child Dev.* 83, 7–29. doi: 10.1111/mono.12348

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

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

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

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: implications for the differentiation between autism and Asperger syndrome. *J. Autism Dev. Disord.* 33, 3–13. doi: 10.1023/A:1022270118899

Junge, C., Kooijman, V., Hagoort, P., and Cutler, A. (2012). Rapid recognition at 10 months as a predictor of language development. *Dev. Sci.* 15, 463–473. doi: 10.1111/j.1467-7687.2012.1144.x

Kern, S. (2007). Lexicon development in French-speaking infants. *First Lang.* 27, 227–250. doi: 10.1177/0142723706075789

Kooijman, V., Junge, C., Johnson, E. K., Hagoort, P., and Cutler, A. (2013). Predictive brain signals of linguistic development. *Front. Psychol.* 4, 1–13. doi: 10.3389/fpsyg.2013.00025

Kuhl, P. K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci.* 97, 11850–11857. doi: 10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533

Kuhl, P. K. (2007). Cracking the speech code: how infants learn language. *Acoust. Sci. Technol.* 28, 71–83. doi: 10.1250/ast.28.71

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Phil. Trans. Royal Soci. B* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Kuhl, P., Conboy, B., Padden, D., Nelson, T., and Pruitt, J. (2005). Early speech perception and later language development: implications for the critical period. *Lang. Learn. Dev.* 1, 237–264. doi: 10.1207/s15473341lld0103&4_2

Leonard, L. B. (2014). Children with specific language impairment. Cambridge, Massachusetts, U.S.: The MIT Press.

Leppänen, P. H. T., Hämäläinen, J. A., Salminen, H. K., Eklund, K. M., Guttorm, T. K., Lohvansuu, K., et al. (2010). Newborn brain event-related potentials revealing atypical processing of sound frequency and the subsequent association with later literacy skills in children with familial dyslexia. *Cortex* 46, 1362–1376. doi: 10.1016/j.cortex.2010.06.003

Lohvansuu, K., Hämäläinen, J. A., Ervast, L., Lyytinen, H., and Leppänen, P. H. T. (2018). Longitudinal interactions between brain and cognitive measures on reading development from 6 months to 14 years. *Neuropsychologia* 108, 6–12. doi: 10.1016/j.neuropsychologia.2017.11.018

Lorusso, M. L., Cantiani, C., and Molteni, M. (2014). Age, dyslexia subtype and comorbidity modulate rapid auditory processing in developmental dyslexia. *Front. Hum. Neurosci.* 8:313. doi: 10.3389/fnhum.2014.00313

Mariani, B., Nicoletti, G., Barzon, G., Ortiz Barajas, M. C., Shukla, M., Guevara, R., et al. (2023). Prenatal experience with language shapes the brain. Science. *Advances* 9:eadj3524. doi: 10.1126/sciadv.adj3524

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

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

Menn, K. H., Ward, E. K., Braukmann, R., Van Den Boomen, C., Buitelaar, J., Hunnius, S., et al. (2022). Neural tracking in infancy predicts language development in children with and without family history of autism. *Neurobiol. Lang.* 3, 495–514. doi: 10.1162/nol_a_00074

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur. J. Neurosci.* 48, 2609–2621. doi: 10.1111/ejn.13748

Mittag, M., Larson, E., Clarke, M., Taulu, S., and Kuhl, P. K. (2021). Auditory deficits in infants at risk for dyslexia during a linguistic sensitive period predict future language. *NeuroImage* 30:102578. doi: 10.1016/j.nicl.2021.102578

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain Lang.* 72, 238–245. doi: 10.1006/brln.2000.2287

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

Moon, C., Cooper, R. P., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500. doi: 10.1016/0163-6383(93)80007-U

Nacar Garcia, L., Guerrero-Mosquera, C., Colomer, M., and Sebastian-Galles, N. (2018). Evoked and oscillatory EEG activity differentiates language discrimination in young monolingual and bilingual infants. *Sci. Rep.* 8:2770. doi: 10.1038/s41598-018-20824-0

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 756–766. doi: 10.1037/0096-1523.24.3.756

Ortiz Barajas, M. C., and Gervain, J. (2021). "The role of prenatal experience and basic auditory mechanisms in the development of language" in Minnesota Symposia on child psychology. eds. M. D. Sera and M. Koenig. *1st* ed (Hoboken, New Jersey, U.S: Wiley), 88–112.

Ortiz Barajas, M. C., Guevara, R., and Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Dev. Cogn. Neurosci.* 48:100915. doi: 10.1016/j.dcn.2021.100915

Ortiz-Barajas, M. C., Guevara, R., and Gervain, J. (2023). Neural oscillations and speech processing at birth. *iScience* 26:108187. doi: 10.1016/j.isci.2023.108187

Parise, E., and Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-month-old infants. *PLoS One* 8:e72360. doi: 10.1371/journal.pone.0072360

Pefkou, M., Arnal, L. H., Fontolan, L., and Giraud, A.-L. (2017). θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *J. Neurosci.* 37, 7930–7938. doi: 10.1523/JNEUROSCI.2882-16.2017

Pujol, R., Lavigne-rebillard, M., and Uziel, A. (1991). Development of the human cochlea. *Acta Otolaryngol.* 111, 7–13. doi: 10.3109/00016489109128023

Ramus, F. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain* 126, 841–865. doi: 10.1093/brain/awg076

Ramus, F., and Ahissar, M. (2012). Developmental dyslexia: the difficulties of interpreting poor performance, and the importance of normal performance. *Cogn. Neuropsychol.* 29, 104–122. doi: 10.1080/02643294.2012.677420

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

Ramus, F., Nespor, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265–292. doi: 10.1016/S0010-0277(99)00058-X

Rivera-Gaxiola, M., Klarman, L., Garcia-Sierra, A., and Kuhl, P. K. (2005). Neural patterns to speech and vocabulary growth in American infants. *Neuroreport* 16, 495–498. doi: 10.1097/00001756-200504040-00015

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, –373

Sansavini, A., Bertoncini, J., and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33, 3–11. doi: 10.1037/0012-1649.33.1.3

Schaadt, G., Männel, C., Van Der Meer, E., Pannekamp, A., Oberecker, R., and Friederici, A. D. (2015). Present and past: can writing abilities in school children be associated with their auditory discrimination capacities in infancy? *Res. Dev. Disabil.* 47, 318–333. doi: 10.1016/j.ridd.2015.10.002

Schulte-Körne, G., and Bruder, J. (2010). Clinical neurophysiology of visual and auditory processing in dyslexia: a review. *Clin. Neurophysiol.* 121, 1794–1809. doi: 10.1016/j.clinph.2010.04.028

Shaywitz, S. E. (1998). Dyslexia. *N. Engl. J. Med.* 338, 307–312. doi: 10.1056/NEJM199801293380507

Song, J., and Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition* 179, 163–170. doi: 10.1016/j.cognition.2018.06.001

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., and Winkler, I. (2009). Newborn infants process pitch intervals. *Clin. Neurophysiol.* 120, 304–308. doi: 10.1016/j.clinph.2008.11.020

Suppanen, E., Winkler, I., Kujala, T., and Ylinen, S. (2022). More efficient formation of longer-term representations for word forms at birth can be linked to better language skills at 2 years. *Dev. Cogn. Neurosci.* 55:101113. doi: 10.1016/j.dcn.2022.101113

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245

Tóth, B., Urbán, G., Háden, G. P., Márk, M., Török, M., Stam, C. J., et al. (2017). Large-scale network organization of EEG functional connectivity in newborn infants. *Hum. Brain Mapp.* 38, 4019–4033. doi: 10.1002/hbm.23645

Tsao, F., Liu, H., and Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Dev.* 75, 1067–1084. doi: 10.1111/j.1467-8624.2004.00726.x

Van Zuijen, T. L., Plakas, A., Maassen, B. A. M., Maurits, N. M., and Van Der Leij, A. (2013). Infant ERPs separate children at risk of dyslexia who become good readers from those who become poor readers. *Dev. Sci.* 16, 554–563. doi: 10.1111/desc.12049

Vander Ghinst, M., Bourguignon, M., Op De Beeck, M., Wens, V., Marty, B., Hassid, S., et al. (2016). Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *J. Neurosci.* 36, 1596–1606. doi: 10.1523/JNEUROSCI.1730-15.2016

Varnet, L., Ortiz-Barajas, M. C., Guevara Erra, R., Gervain, J., and Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *J. Acoust. Soc. Am.* 142, 1976–1989. doi: 10.1121/1.5006179

Vouloumanos, A., and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* 10, 159–164. doi: 10.1111/j.1467-7687.2007.00549.x

Zhang, Y., Zou, J., and Ding, N. (2023). Acoustic correlates of the syllabic rhythm of speech: modulation spectrum or local features of the temporal envelope. *Neurosci. Biobehav. Rev.* 147:105111. doi: 10.1016/j.neubiorev.2023.105111

Zoefel, B., and VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage* 124, 16–23. doi: 10.1016/j.neuroimage.2015.08.054

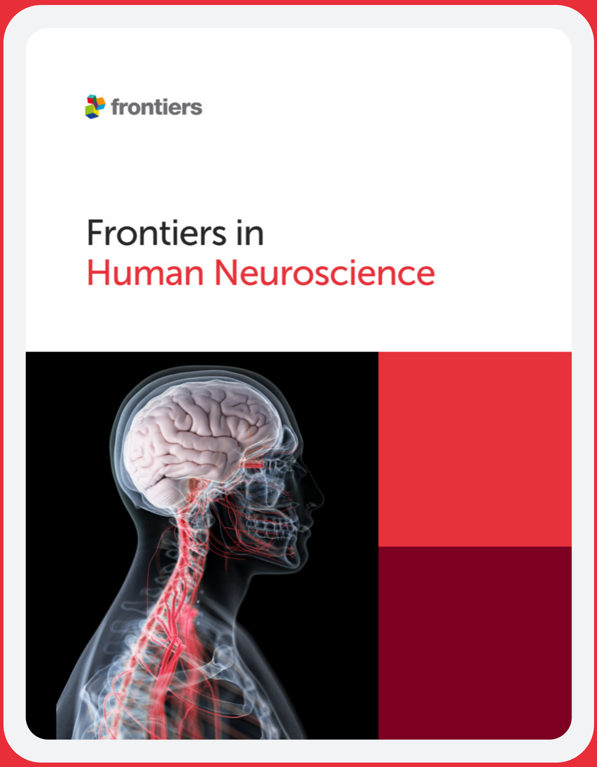
Frontiers in
Human Neuroscience

Bridges neuroscience and psychology to understand the human brain

The second most-cited journal in the field of psychology, that bridges research in psychology and neuroscience to advance our understanding of the human brain in both healthy and diseased states.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Human Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact