36 Neural and Behavioural Rhythmic Tracking during Language Acquisition: Findings, Methods, and Outstanding Issues

KANAD N. MANDKE AND SINEAD ROCHA

36.1 Introduction

Language acquisition is a multimodal phenomenon. Within the womb, the fetus is exposed to the rhythm of their mother's speech via a low-pass filter. They hear the rumbling of their mother talking; they feel her movements. At birth, infants can recognise their mother's voice (Mehler et al., 1978) and show familiarity with stories read to them in utero (Decasper and Spence, 1986). They are not just passive recipients; in their earliest communications, their cries follow the pattern of the language they are exposed to (Mampe et al., 2009). At birth, even with months of auditory experience under their belt, their language system is flexible and open to the input it receives. Young infants can discriminate between sounds in languages they have never been exposed to before, an ability that is lost over the first year of life as the system acquires expertise for its language(s) (Maurer and Werker, 2014). The journey towards adult-like language expertise is long; infants have to learn vocabulary, syntax, and grammar. All these elements have been extensively studied in infants and young children, and we have a wealth of knowledge of key roles of, for example, ostension (Csibra, 2010) or statistical learning (Romberg and Saffran, 2010).

In recent decades, fuelled in part by observations from language disorders, adult speech perception, and music perception, a new contender on the block has emerged as a critical component of linguistic success – rhythm perception. The grossly oversimplified story (discussed with the detail it deserves in other chapters of this edition in Section 6) is that speech is a rhythmic signal, and that efficient processing of the rhythm of speech facilitates language acquisition. The patterning of syllables and stress syllables gives anchors, or perceptual edges, in the speech signal that allow the listener to attend to important information in speech (Doelling et al., 2014). Rhythmic cues give structure to the speech signal for the listener to follow. What is intriguing is that when we speak to infants (see Chapters 23 and 38), we emphasise this rhythm, slowing down and adding greater emphasis. Our voices take on a sing-song quality that is not a reflection of the expert speaker but is attuned to our novice listeners. This phenomenon is known as infant-directed speech (IDS). IDS is linked to enhanced word

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

learning. Infants learn new words better when they are presented in IDS than adult-directed speech (ADS) (Ma et al., 2011), and this benefit is also true for adults learning a new language (Ma et al., 2011). Critically, IDS is not necessary for learning throughout the acquisition journey – once a language has been sufficiently mastered, older infants learn well without it (Ma et al., 2011). Caregivers are therefore responsive to the needs of their infant, modulating the acoustic properties of the IDS they produce according to infant age, and likely reflecting infant attention to different acoustic cues, within the bidirectional and dynamic caregiver–infant speech interactions (Cox et al., 2023). Similarities in prosody have been demonstrated amongst diverse societies (Broesch and Bryant, 2018), and IDS is largely considered universal, at least in form if not quantity (Cox et al., 2023). Adults can distinguish IDS from ADS in non-native languages from short, contextless audio excerpts (Hilton et al., 2022). If the greater rhythmicity of IDS is a critical universal property, we must settle on some core understandings of what we mean by rhythm. In music, rhythm describes a series of temporal intervals (see Chapter 27). It is often characterised by isochrony or equal spacing between event onsets. Whilst naturalistic speech never has the regularity of a metronome or click track, IDS has greater isochrony than ADS. We can consider the proximate mechanisms that may be at play whilst infants are listening to this special rhythmic signal, the most intuitive being that infants are (neurophysiologically) tracking the rhythm of IDS. For this to be the case, we must meet two criteria. First, that the infants can neurally track an auditory rhythm, and second, that the speech contains an auditory rhythm for infants to track.

36.1.1 Criterion 1: Infants Perceive Auditory Rhythm

There is good evidence from the field of music cognition that infants can perceive auditory rhythms. We see this behaviourally, for example in habituation studies where we see that infants discriminate tempo changes (Baruch and Drake, 1997) and metre (Hannon and Johnson, 2005). Through infancy, infants' spontaneous movement behaviour changes in response to music, and whilst infants cannot reliably synchronise to music, they show tempo-flexibility, moving faster to faster auditory tempi and slower to slower tempi (Rocha and Mareschal, 2017; Yu and Myowa, 2021; Zentner and Eerola, 2010). We are also able to measure rhythm perception neurally, with electroencephalographic (EEG) mismatch responses showing that infants detect a missing beat (Winkler et al., 2009) and interpret metrical structure (Flaten et al., 2022). A more direct approach to measuring infant neural responses to musical beats has used steady-state evoked potentials (SSEPs), which reflect the amount of neural energy at different frequencies. An established measure in adult music cognition (Nozaradan et al., 2011), this approach has been used to show that infants have enhanced energy at the perceived beat and metre frequencies of auditory rhythmic patterns (Cirelli et al., 2016; Flaten et al., 2022).

36.1.2 Criterion 2: IDS Contains Auditory Rhythm

If we are therefore happy to proceed with our argument that infants perceive critical timing information in auditory rhythmic stimuli such as repeated tones or real music, the next criterion

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

for rhythm as a key to language acquisition is to show that there is indeed rhythm in the speech signal for infants to track (see Chapter 23). Studies of the acoustic signal of naturalistic IDS show increased amplitude modulations around 2 Hz (Leong et al., 2017). To investigate this, Leong et al. applied a computational model to child-directed speech (CDS) and revealed that the speech is hierarchically organised, known as the spectral-amplitude modulation phase hierarchy (S-AMPH). The approach consists of a set of algorithms that are used to derive underlying spectral characteristics of the speech signal. It uses probabilistic demodulation to model the rhythm patterns in speech, giving a low-dimensional representation of the acoustic and temporal properties of the speech envelope (Goswami and Leong, 2013; Leong and Goswami, 2014, 2015; Leong et al., 2017). This data-driven modelling approach allows us to identify various amplitude modulations corresponding to linguistic boundaries. For example, in the first report on S-AMPH (Leong and Goswami, 2015), the application of the modelling approach to CDS revealed amplitude modulations corresponding to prosodic stress (stress AM ~2 Hz), syllables (~5 Hz), and phoneme rate (~20 Hz). Furthermore, they argued that this nested hierarchy of speech rhythms could be used by an infant to build stimulus-driven phonological maps of a speech system in any given language. Particularly in CDS, these amplitude modulations are exaggerated, and possibly provide the essential acoustic landmarks for children.

Given the above support for our core criteria, it is not surprising that there has been increased focus in recent years on understanding the mechanisms by which rhythmic processing of speech may support typical and atypical language acquisition. The human auditory cortex has been shown to reliably track the amplitude envelope of the speech signal. This is achieved by phase aligning endogenous neural oscillations with the amplitude envelope of the temporally regular auditory information. The speech envelope refers to the amplitude fluctuations over time, typically occurring in low frequencies (< 10 Hz), which help the listener track the speech rhythm. Using magnetoencephalography (MEG), speech tracking of the amplitude envelope was demonstrated in healthy adult listeners (e.g., Gross et al., 2013; Peelle et al., 2013) but since then has been revealed in infant EEG (Attaheri et al., 2022a; Jessen et al., 2021; Menn et al., 2022). There is evidence that this speech envelope-tracking ability develops from childhood to adulthood, and even supports better performance in speech in noise (Destoky et al., 2020; Vander Ghinst et al., 2019). The speech-tracking literature has predominantly used measures such as speech–brain coherence, phase-locking value, and mutual information. These methods essentially measure statistical dependency between the speech signal and underlying neurophysiological data. In the rest of this chapter, we aim to provide an account of the state-of-the-art methods being developed to elucidate the relationship between rhythm and language, summarise where the literature converges and diverges, highlight open questions, and discuss the developments in our field that can enhance understanding of these phenomena.

36.2 A Primer on Neural Measures of Rhythm Processing Suitable for Use with Infants

Neural measurements from the earliest moments in life have been possible for some decades now, including via EEG, MEG, and near-infrared spectroscopy (fNIRS). Most relevant studies

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

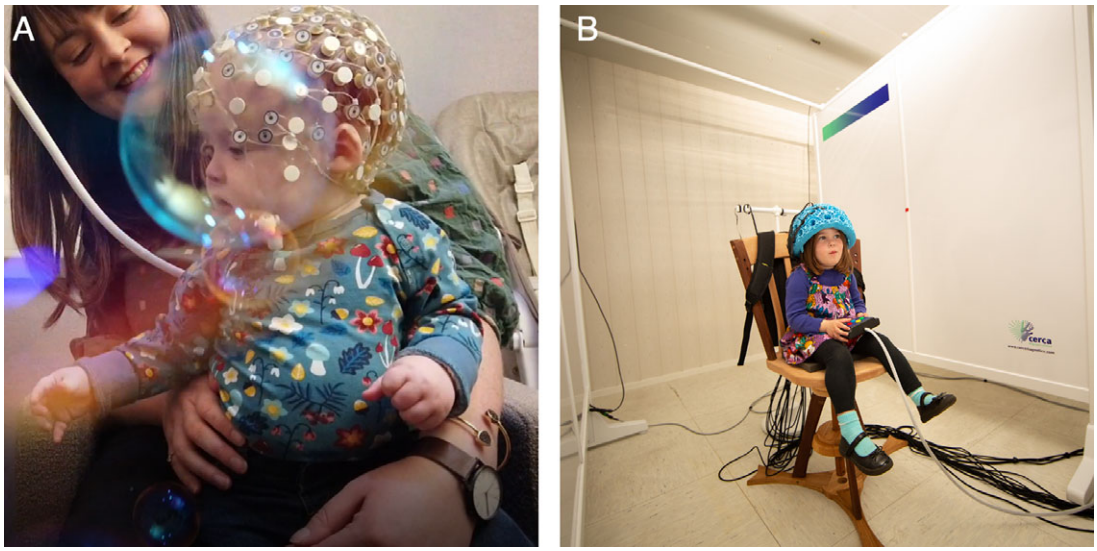


Figure 36.1 Infant neural activity can be measured passively using EEG or MEG systems.
1A. An infant wearing a geodesic sensor net. 1B. MEG adapted with lightweight optically pumped magnetometers.
Picture credits: 1A: Eleanor Smith; 1B: Paul Allen

to this chapter use M/EEG for its excellent temporal resolution. EEG measures spontaneous neuronal activity generated by ensembles of neurons, from the surface of the scalp. MEG on the other hand measures the magnetic components of this underlying neuronal activity. Infant EEG often comprises high-density (64–128-channel) recording using water-based geodesic sensor nets that need little preparation, aiding infant compliance (Figure 36.1a). Modern systems are improving traditional issues with signal-to-noise ratio (SNR), with infant active electrode caps that can be pre-gelled and applied almost as quickly as nets.

Infant EEG has further benefited from technological advances in signal processing post data collection. As it is challenging to ensure infants remain stationary during an experiment, data can suffer from non-canonical movement artefacts, difficult to remove using standard adult-defined techniques. However, recent noteworthy advancements in toolboxes and tutorials specifically for infant EEG data allow greater precision in the analysis of noisy data (Gabard-Durnam et al., 2018; Lopez et al., 2022). These general technological advancements have facilitated the growth in complex methodologies suitable for answering questions on infant speech perception. On the other hand, MEG offers the same temporal resolution as EEG and has a reasonable spatial resolution allowing us to investigate activity between networks of brain regions. A crucial limitation of the traditional cryogenically cooled MEG system is that it has a fixed array of sensors, making head movements a confound in typical experiments. As the sensor array is fixed, any head motion relative to the sensor array can cause changes in the SNR and spatial blurring of the underlying sources. Recognising this limitation, several algorithms are now available to correct head movement artefacts. However, changes in

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

the SNR (as sources move relative to the array) during recording place a limit on the amount of movement that can be compensated (Medvedovsky et al., 2007). The problem of head movement is much more pronounced in the paediatric population, where infants and/or toddlers find it very difficult to stay still in unnatural (i.e., laboratory) environments. This limitation is better overcome by EEG and fNIRs, which involves placing the sensors directly on the participants' heads. Recent exciting developments in MEG hardware have led to the development of room temperature MEG sensors, which involve the use of optically pumped magnetometers (OPMs) (Boto et al., 2017, 2018). The lightweight sensors (OPMs) can be mounted in a helmet, making the scanner a wearable device. This new approach is gaining traction, and early adoption with children demonstrates significant improvements in the SNR with OPMs when testing children with epilepsy (Feys et al., 2022), cortical tracking of speech (de Lange et al., 2021), and hyperscanning during play (Holmes et al., 2023). Being able to place the OPMs directly over a participant's head has two distinct advantages: (1) improved SNR, and (2) improved spatial resolution. This makes a compelling use case in developmental populations, particularly during naturalistic experiments. For example, in a study by Hill et al. (2019), the OPM-MEG system was used to measure somatosensory activity underlying maternal touch in two- and five-year-olds. Therefore, whilst this chapter mostly discusses infant EEG, we see great potential for MEG research in the coming years.

36.3 Methodological Overview

Human speech is intrinsically rhythmic. This is mainly the result of coordinated movement by the oro-musculature involved in speech production. In a stress-timed language such as English, the rhythm in speech typically translates to the occurrence of stress and unstressed syllables in connected speech (Cummins and Port, 1998; Nespor et al., 2011). The speech rhythm (i.e., prosody), indexed by the changes in the amplitude envelope of the signal, offers critical cues for speech segmentation (see Chapter 11 for an alternative perspective). Whilst there are variations in the rate of speech, both within and between speakers, healthy adult listeners change their ongoing neural oscillations to match the incoming speech signal. This is a key mechanism for speech perception. Nevertheless, how the auditory cortex achieves this impressive feat, and the precise oscillatory mechanisms underlying it, remain largely elusive. Moreover, and of particular interest to developmental neuroscientists, there are the questions, what does this mechanism look like in infancy and childhood? (How) does it aid language acquisition? And what happens when these mechanisms break down early in childhood? Progress towards answering these questions has been made through measurement of the associations between the speech signal and ongoing neural oscillations using M/EEG. Owing in large part to these speech-tracking methods, the mechanism of 'neural entrainment' as a basis for speech processing and language acquisition has also received considerable support. Here, we look at speech–brain coherence, phase-locking value (PLV), mutual information (MI), and multivariate temporal response function (mTRF) as examples of methods that have been used to study neural entrainment.

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

36.3.1 Speech–Brain Coherence

Coherence is a statistical measure that is used to identify statistical dependency between two signals, $x(t)$ (e.g., speech time series) and $y(t)$ (e.g., neural time series). It is given by:

$$Coh_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \tag{1}$$

where $S_{xy}(f)$ is cross-spectral density between x and y, and $S_{xy}(f)$ and $S_{yy}(f)$ are the auto-spectral density of x and y, respectively. The spectral densities are estimated using Fourier transform. Values of coherence range between 0 (random coupling) and 1 (perfect synchronisation) (Pascual-Marqui et al., 2011).

36.3.2 Phase-Locking Value (PLV)

PLV measures frequency-specific phase synchronisation between two signals. It is computed by calculating the distribution of phase difference extracted from two source time series $x(t)$ and $y(t)$. It is formally given by:

$$PLV_t = \frac{1}{N}\left|\sum_{n=1}^{N} \exp(j\theta(t,n))\right| \tag{2}$$

where $\theta(t,n)$ gives the phase difference $\phi_1(t,n) - \phi_2(t,n)$. The phase information is typically extracted using the Hilbert transform. PLV provides a summary statistic of the phase difference at t (Lachaux et al., 1999).

36.3.3 Mutual Information (MI)

MI serves as a measure of mutual dependence between two random variables. It is used to quantify the amount of information that can be obtained about one variable by observing the other variable. Unlike speech–brain coherence or PLV, MI captures both linear and non-linear interactions between the two signals. An additional advantage of the method is that the same framework can be extended to study different aspects of the underlying signals (e.g., phase-phase, amplitude-amplitude, phase-amplitude, or cross-frequency coupling). The MI between two random variables X and Y is mathematically given as follows:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\, log\, \frac{P(x,y)}{P(x)P(y)} \tag{3}$$

where $P(x)$ and $P(y)$ are the marginal distributions of variables X and Y, respectively, and $P(x,y)$ is the joint distribution of these variables.

The general steps involved in all three above methods include: (1) band-pass filtering of the neural time series and the speech signal in the same frequency bands; (2) extraction of the relevant quantity (e.g., spectral density, phase, or amplitude information); before (3) subjecting it to the relevant mathematical operation.

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

36.3.4 Multivariate Temporal Response Function (mTRF)

The mTRF is a novel method for investigating the neurophysiological processing of the auditory signal. Unlike the methods mentioned above, the mTRF method involves decoding the patterns of neural activity related to a particular stimulus feature using a set of linear filters, which could include acoustic envelope, spectrogram, phonemes, or phonetic features (Crosse et al., 2016; Di Liberto et al., 2015). These filters are trained on, for example, 80% of the data and then applied to the remaining 20% to generate predictions (or the time course) of the stimulus feature in question. The mTRF approach has some advantages. First, an explicit pre-selection of channels (or ROIs) is not required as data from all the channels is used to create a stimulus reconstruction. Second, the commonly used backward modelling approach can maximise sensitivity to key signal differences between highly correlated sensors. This is achieved by mapping data from all sensor locations simultaneously and by detecting correlations in the data.

36.3.5 Comparison of Approaches

Relevant developmental research in the auditory domain has historically been dominated by the use of non-speech sounds as stimuli, such as amplitude-modulated or frequency-modulated tones, to measure auditory steady-state response (ASSR). These approaches remain very popular because the neural responses to such stimuli are very robust and can reliably be recorded across the lifespan. However, such experiments suffer from a lack of ecological validity and don't allow us to measure the development of neural responses in a naturalistic setting. Experiments with the use of naturalistic, immersive paradigms have recently started to increase, using the methods described above. Such paradigms using audiovisual stories, nursery rhymes, or IDS allow us to study how multiple streams of information are processed by the infant's brain. This gives a clear benefit of increased generalisability of the findings.

All the methods outlined in our chapter (coherence, PLV, MI, and mTRF) generally suffer from the same limitations; that is, developmental studies tend to have smaller sample sizes and noisier data than the adult studies from which these techniques have been developed. The ability of each method to deal with inherent low SNR should be considered by the researcher. Further, all the methods briefly reviewed here (except for MI) rely on linear relationships between the speech signal and the neurophysiological data. This assumption may not be sufficient to fully encapsulate the brain's response to speech stimuli. A further limitation specific to mTRF concerns model selection, as the researcher must define the specific speech parameter that they are interested in studying (e.g., speech envelope, spectrogram, or phonetic features). Choosing the right model for the mTRF can be challenging, and different models may perform differently for different types of stimuli or neural responses. The set of models that generate statistically significant results for one research group may not generalise to other tasks/conditions/datasets. We also think it is important to highlight that the mTRF reconstruction values are often very small. This might be partly to do with the noisy nature of M/EEG signals. Whilst the effects reported in the literature using the mTRF method show statistical significance when compared to a null distribution, their clinical significance remains underexplored, and this will be a critical next step for the field.

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

Finally, it is worth noting that the brain's responses to rhythmic stimulation can be a mixture of series of evoked responses and non-phase-aligned oscillatory (or induced) responses (David et al., 2006). It is important to disentangle the two when studying oscillatory responses in infants as researchers risk attributing oscillatory functions to evoked activity. This can be achieved by removing the averaged evoked response from the data before analysing it or by incorporating computational models (e.g., Doelling et al., 2019) with theoretical models of language acquisition.

36.4 Synthesis of Infant Rhythmic Processing Literature

As identified in Table 36.1, we are now well equipped to ask and answer questions on the neural underpinnings of rhythmic speech processing. The studies outlined below offer a snapshot of 'neural entrainment' research in infants and how this mechanism may aid language acquisition. The precise definition of neural entrainment remains hotly debated (Giraud, 2020; Haegens, 2020; Meyer et al., 2020), and we prefer the term speech tracking. Here, speech tracking is defined as the neural process by which the ongoing neurophysiological activity follows the patterns of the speech signal. However, a causal link has yet to be established.

To our knowledge, the first study to investigate the differential neural substrates of IDS and ADS tracking measured neurophysiological (EEG) responses to recordings of naturalistic IDS and ADS in seven-month-old, pre-verbal infants (Kalashnikova et al., 2018). In this study, spectral analysis revealed that the theta-band power over the left hemisphere was significantly larger than the right hemisphere. The hemispheric differences provide compelling evidence in support of the asymmetric sampling hypothesis (AST) (Hickok and Poeppel, 2007; Poeppel, 2003). It is possible that the functional asymmetry postulated by AST may have origins as early as seven months of age, when infants are at the beginnings of language production, producing babbling. Furthermore, analysis using mTRF showed that theta-band (4–8 Hz) cortical tracking of the speech envelope was greater for IDS than ADS. Here, the authors investigated theta tracking as they were interested in how the exaggerated prosodic features of IDS, such as higher pitch and slower tempo, may enhance the salience of speech sounds for infants and make them

Table 36.1 A summary of techniques that have been used to measure speech tracking developmentally

Method	*Relevant papers*
Speech–brain coherence	Fransen et al. (2015); Mandke et al. (2022); Molinaro et al. (2016); Pascual-Marqui et al. (2011)
Phase-locking value	Lachaux et al. (1999); Peelle et al. (2013)
Mutual information	Gross et al. (2013); Teng et al. (2019); Zan et al. (2019)
Multivariate temporal response function	Attaheri et al. (2022a); Crosse et al. (2016); Di Liberto et al. (2015)
Phase-amplitude coupling	Attaheri et al. (2022a, 2024); Canolty et al. (2006)

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

easier to process. These findings are in line with the literature outlined in our introduction, which suggests that these unique characteristics of IDS may play an important role in early language acquisition in infants. That amplitude envelope was tracked gives a first insight into the idea that it is indeed the rhythm of IDS that is a critical component. However, the choice of investigating theta-band oscillations in response to the envelope reflects the authors' interest in IDS directing the infants' attentional spotlight, and it would be very interesting to understand how different saliency cues drive cortical tracking. Without additional manipulations, it is not possible to know if it specifically or exclusively the enhanced low-frequency rhythms of IDS driving cortical tracking.

In a longitudinal study, Attaheri et al. (2022a) measured cortical tracking of sung speech in infants at four, seven, and 11 months of age using mTRF applied to EEG data in canonical delta, theta, and alpha bands. Audiovisual stimuli were used of a woman performing various British nursery rhymes (e.g., 'Twinkle Twinkle Little Star'). The results revealed that infants had above-chance performance cortical tracking in delta and theta bands across the three time points. They also identified the presence of strong phase-amplitude coupling with delta–theta bands as the drivers. More details on this study and its functional interpretation can be found in Chapter 35. Whilst the current data cannot provide direct evidence for the involvement of this neural process in the extraction of linguistically meaningful information, the data form part of a longitudinal study that continued to track infants into the third year of life with detailed language assessments. Through this design, it is possible to see the extent to which early processing of the amplitude envelope of sung speech predicts language acquisition (Attaheri et al., 2022a, 2024). However, what is also intriguing about these results is a complex developmental pattern. The original longitudinal findings show the strongest mTRF values of cortical tracking at four months of age, with significantly lower levels at 11 months. Attaheri et al. (2022b) also replicated their findings with adults. This study used identical stimuli and the same analysis pipeline to that of the infant study. Here, their findings revealed that adult cortical responses to sung speech reflected very similar underlying processes, showing increased delta- and theta-band tracking, with similar mTRF values for adults as for the infants at the youngest time point tested (four months). The suggested overall trajectory may therefore be an inverted U shape, with younger infants performing similarly to adults, but with weaker tracking in the intervening period. Such interpretation remains speculative, especially as it is not clear from these results whether greater tracking at four months than 11 months is the result of a true developmental characteristic (e.g., increased salience of the stressed syllable amplitude modulation at this early age), or a physical or methodological characteristic (e.g., cleaner EEG data at the earlier age whilst the infant is less mobile).

In another study of how natural IDS facilitates the neural processing of prosody in infants, Menn et al. (2022) used EEG to measure speech–brain coherence in seven–nine-month-old infants. The infants listened to either IDS or ADS presented live by their caregiver. The results showed statistically significant speech–brain coherence for IDS and ADS at prosodic rates. However, the speech–brain coherence was significantly greater for IDS compared to ADS, specifically in the prosodic rates. The authors suggest that natural IDS may facilitate infants' ability to track and learn the rhythmic features of speech, which could in turn support language development. The rhythmic patterns of IDS, which are characterised by exaggerated intonation, slower tempo, and higher pitch, are thought to aid in infant attention and arousal, as well as in the

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

formation of speech representations in the brain. The main contribution of this study is the ecological validity of the naturalistic speech, with the design set up such that the caregiver and infant were communicating as they would at home, in the IDS condition. However, Menn et al. also draw our attention to the fact that the ADS condition was not exactly matched to the IDS, as caregivers were instructed to additionally remove all ostensive cues, such as mutual gaze. It would be of great interest to understand the additive benefit of such cues in future work.

Across the studies and techniques discussed thus far, it is worth noting that most use non-specific linguistic timescales and assume that there is a one-to-one mapping between speech rhythms and canonical neural oscillations (e.g., delta, theta, gamma bands). Of significant importance is the inter-speaker variability or different registers of speech such as ADS or IDS that can produce speech rhythms across a broad range (see chapters in Section 6). Therefore, it is important to first identify the specific linguistic timescales of interest in the speech material before studying the corresponding neural oscillations. This question was first addressed in adults (Keitel et al., 2018) by manually annotating their stimulus material, and by applying data-driven filtering to CDS (Mandke et al., 2022). Both these studies identified prosodic features < 5 Hz. These statistical regularities are noticeably lower than assumptions made in the literature; for example, syllable rate is reflected in the theta band (4–8 Hz) and phoneme rate in the gamma band (> 30 Hz). Constraining the neural oscillations by linguistic boundaries identified in the stimulus material will improve the precision of interpretation, particularly in the language acquisition literature.

Overall, these studies provide valuable insights into the neural mechanisms underlying the processing of prosody in infants and highlight the importance of natural IDS in supporting early language development. However, speech envelope tracking alone may not be sufficient to account for language acquisition, as it oversimplifies the computations undertaken by the infant brain. It fails to consider the role of other features contained in the speech signal, such as phonetic features, formant transitions, temporal fine structure, and so on. For example, Inbar et al. (2023) recently investigated the neurophysiological basis of intonation units (IUs), a fundamental unit of human languages (Inbar et al., 2020). In their naturalistic listening study using EEG, Inbar et al. (2023) demonstrated robust evoked responses to IU in adult listeners. For further details, we direct the reader to Chapter 15. The evidence from the adult speech-tracking literature strengthens proposals that as the acoustic information travels along the auditory pathway, higher-order structures extract more complex representations from the speech signal. This representational hierarchy receives support from the fact that the anatomy of the auditory system is also hierarchically organised. Future work to account for how meaning is assigned to these speech features (e.g., amplitude envelope), and how these are further used by the developing brain in speech production, will be valuable next steps (see Chapters 17 and 18).

36.5 Multimodal Rhythm Perception and Production in Relation to Language Acquisition

In our introduction, we highlighted that speech is a multimodal act, and in this section, we wish to stress that the rhythm in speech is multimodal. Typically, when infants are exposed to

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

speech, they are not only hearing the auditory signal but also gaining rich visual information. For example, when singing to an infant, adults' metrically strong moments involve temporally aligned eye-widening and blinking, in addition to the movement of the mouth (Lense et al., 2022). Infants are receptive to this and their looking at the eyes of the singer is coordinated with these eye movements (Lense et al., 2022). IDS is produced with larger mouth movements than ADS (Green et al., 2010) and more head movements (Smith and Strader, 2014). Eyebrow movements and head nods are particularly useful cues to phrase boundaries and are again more prominent in IDS than ADS (de la Cruz-Pavía et al., 2020). Such inter-sensory redundancies (i.e., synchronous information across modalities) facilitate the detection of changes in prosody above an auditory cue alone (Bahrick et al., 2019). For more details on the multimodal nature of the speech input that infants receive, we direct the reader to Chapter 38.

Aside from the focus on the rich multimodality of the language stimulus directed to infants, it is also critical that we do not forget the multimodality of infants' attempts at language production. The relationship between gross motor actions across limbs and the development of speech is well recognised. Early repetitive motor movements, such as kicking or hand-waving, in which infants can spend 40% of their time, have been described as stereotypies, reflexive or rhythmic actions that precede more deliberately controlled movement (Thelen, 1981). We can think of these rhythmic movements as a 'passive' response to the speech, with seminal studies showing that neonates' earliest movements are associated with the timing of adult speech (Condon and Sander, 1974). However, we can also think further about rhythmic movements whilst infants are actively generating speech sounds themselves. From a dynamic systems theory approach, rhythmic motor actions produced with the mouth and hand may entrain each other, such that the generation of a well-practised action such as hand-banging may 'pull in' the timing of vocalisations (Iverson and Thelen, 1999). It is well documented that fluent speech is preceded by canonical babbling, where the infant produces repetitions of consonant-vowel syllables (Kuhl, 2004). Rhythmic movements such as shaking a rattle reach their peak around the time that infants begin canonical babbling, and drop off once babbling is established (Ejiri, 1998; Iverson et al., 2007). Infant babbling is frequently temporally coordinated with rhythmic movement such as hand-banging, and the vocalisations that co-occur with such movement show more mature properties, which sustain after the movement ends (Ejiri and Masataka, 2001).

Therefore, in addition to considering infants' neural tracking of speech rhythm, we believe it is critical to also consider infants' motoric rhythmic responses to speech. Whilst the cortical tracking methods described above give fine-grained temporal information, the seminal studies of infant movement discussed thus far largely rely on micro-coding of video data, constrained by the frame rate of the video collected. The advancement of motion capture technology now facilitates nuanced analysis of infant movement without the need for the frame-by-frame hand coding of video. Optical 3D motion capture uses reflective markers placed at strategic points on the infant, from which x-, y-, and z-coordinates can be derived. In Figure 36.2a, the infant is wearing rigid bodies (prearranged unique combinations of markers stuck to a firm board), attached to the limbs and head via soft, elasticated, fabric straps. Optical motion capture systems can record infant movement at up to 2,000 frames per second, allowing incredibly high precision measurement. However, these systems are still relatively expensive, requiring the use

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

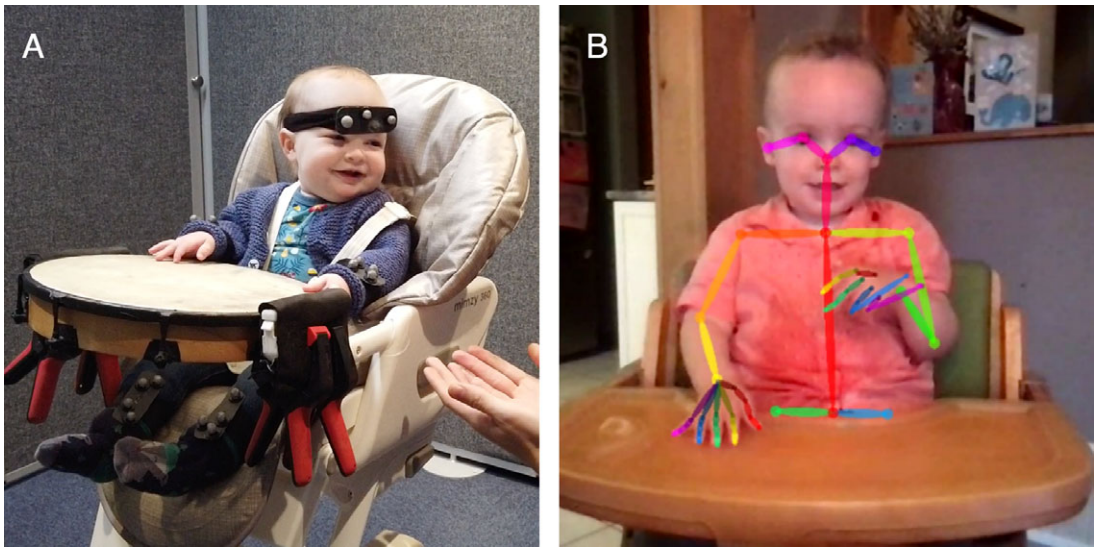


Figure 36.2 Motion capture methods.
2A. Infant wearing rigid body reflective marker arrangements for optical motion tracking. 2B. Infant recorded on home webcam and analysed offline using OpenPose open-source markerless motion capture.
Picture credits: 2A: Eleanor Smith

of multiple near-infrared cameras to measure the reflection of light from the markers. The recent emergence of markerless motion-tracking technology for 2D pose estimation can allow even more naturalistic recording of infant movement via normal video. Markerless motion tracking uses deep-learning models trained on large video datasets to tag key points such as wrist, elbow, or shoulder, and even facial features. Figure 36.2b shows the application of an open-source markerless motion capture model to an infant drumming.

Motion capture studies of infants' rhythmic movement whilst listening to rhythmic stimuli provide interesting insights. Whilst as a group they do not show sensorimotor synchronisation to the rate of auditory presentation at an adult level, over the first two years of life, infants show tempo-flexibility or move faster to faster auditory tempi and slower to slower tempi (Rocha and Mareschal, 2017; Zentner and Eerola, 2010). Case studies show that some infants may show such adaptation to the rate of music as early as three to four months of age (Fujii et al., 2014). Due to the convincing evidence that infants are tracking the amplitude envelope of speech sounds and other rhythmic auditory stimuli (Attaheri et al., 2022a; Menn et al., 2022), it is tempting to conceive of infant rhythmic movement as a reflection of the strength of the tracking of the signal (i.e., to hypothesise that those who are tracking the stimulus well will also show better temporal matching of their movement to that stimulus). However, infants' spontaneous rhythmic movements are found equally in silence as they are to a rhythmic musical stimulus (de l'Etoile et al., 2020; Fujii et al., 2014). Zentner and Eerola (2010) showed equal rates of rhythmic movements for a simple drumbeat as naturalistic music, but less for IDS or ADS. Whilst the relationship between infants' quantity and rate of rhythmic movement is therefore not directly tied to what they are hearing, it is very interesting to consider how the development

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

of rhythmic movement and sensorimotor synchronisation may unfold. For example, in the Cambridge UK BabyRhythm longitudinal study, where infant drumming to speech and non-speech rhythmic stimuli was recorded using motion capture, Rocha et al. (2024) show that infant drumming becomes more rhythmic with age. This was particularly true when infants were drumming in a silent control condition, which mirrors previous findings that infants' spontaneous motor tempo becomes faster and more regular over the first years of life (Rocha et al., 2021). The Cambridge UK BabyRhythm study compared infant drumming in silence to an isochronous 2 Hz drumbeat, an isochronous 2 Hz repetition of the syllable 'ta', and naturalistic sung nursery rhymes. Rocha et al. found that infant drumming in the presence of a drumbeat showed a similar maturation as their spontaneous motor tempo, becoming faster and more rhythmic with age. However, infants did not show the same pattern of becoming more regular with age in the linguistic conditions (repeated syllables and sung nursery rhymes). It is possible that infants are upregulating variability, in response to more complex auditory stimuli, perhaps reflecting a trade-off between greater adaptation to the stimulus with age and rhythmicity.

36.6 Conclusion

This chapter aimed to unpack how rhythm supports language acquisition. Furthermore, we have provided an overview of the methods and highlighted some open questions. It will be of great interest to better understand gross motor rhythmic action in the context of speech perception and production, and the interplay between the seemingly good early cortical tracking, with the emphasis on neural alignment, and less precise or more variable behavioural tracking.

The focus of this chapter on neural tracking of the rhythmic auditory information in speech reflects cutting-edge neuroscience, using ever more sophisticated techniques to drill into the minutiae of perception. The focus is undoubtedly on the auditory modality, but even where this focus is being broadened to consider, for example, visual information present on the face (Lense et al., 2022; Ní Choisdealbha et al., 2024), it is still often measuring the unidirectional impact of features of speech, often presented on a screen, to the neural firing of the infant. In the final section, we make several recommendations as to how we can integrate diverse areas of knowledge and capitalise on the rapid technological developments, to consider the role of rhythm in infant language acquisition more holistically.

We outlined how methodological advancements have provided insight into the way that the infant brain processes language. Recent years have shown that the rhythmic information carried by the amplitude envelope of the speech signal is a core characteristic of IDS that infants are indeed processing. However, it is important to acknowledge other low-level features (e.g., envelope, spectrogram, temporal fine structure) and high-level features (e.g., phonetic features) that are part of the speech signal and are reflected in the EEG signal (Di Liberto et al., 2015). To what extent are these additional timing or landmark cues important? To what extent are these cues more or less important in the special case of IDS? It is important that we do not simply apply our learning from adult speech studies to developmental problems. In the coming years, we can add thorough consideration of the other rhythmic properties of IDS that may be

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

critical, for example, in the visual, touch, or motor domains. In doing so, and without constraining our focus on rhythm to only reflect the amplitude envelope, we can fully consider the breadth of developmental scaffolding that IDS provides.

Finally, in addition to understanding the rich multimodality of spoken language, it is critical to note that outside of the lab, exposure to IDS occurs in bidirectional social interactions (Menn et al., 2023), within a wider conversational context (Golinkoff et al., 2015). Regarding the infant as simply a passive receiver of (auditory) information does a great disservice to how we know that infants develop language. For example, neonates vocalise more when a parent is present (Caskey et al., 2011). Other strands of research into early communication are taking a hyper-scanning approach, where the neural activity of both the infant and the caregiver is recorded simultaneously (e.g., Nguyen et al., 2021a, 2021b, 2021c). The M/EEG methods we have described in this chapter can handle the complexity of this kind of information, and we should attempt to embrace this complexity to further centre the infant as a participator in, rather than the recipient of, IDS.

Box 36.1 Chapter Overview

Summary

The chapter synthesises the current evidence supporting infants' ability to track speech rhythms and underscores the importance of IDS in language acquisition. We advocate for widening the scope of IDS to include visual, somatosensory, and motor rhythms (in addition to auditory), which additionally shape early language acquisition.

Implications

Understanding how infants track incoming sensory information in different modalities has broad implications. The current literature posits rhythm perception as a critical element of language acquisition. Advancements in studying multimodal infant responses will deepen insights into this pivotal aspect of early language development.

Gains

The chapter provides a summary of the state of the art in speech and rhythm processing and how it relates to early language acquisition. We identify avenues for future research and provide a commentary on the suitability of the most popular methods.

REFERENCES

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

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

Attaheri, A., Choisdealbha, Á. N., Rocha, S., et al. (2024). Infant low-frequency EEG cortical power, cortical tracking and phase-amplitude coupling predicts language a year later. *PLoS One*, 19(12), e0313274. https://doi.org/10.1371/journal.pone.0313274

Attaheri, A., Panayiotou, D., Phillips, A., et al. (2022b). Cortical tracking of sung speech in adults vs infants: A developmental analysis. *Frontiers in Neuroscience*, 16, 842447. https://doi.org/10.3389/fnins.2022.842447

Bahrick, L. E., McNew, M. E., Pruden, S. M., and Castellanos, I. (2019). Intersensory redundancy promotes infant detection of prosody in infant-directed speech. *Journal of Experimental Child Psychology*, 183, 295–309. https://doi.org/10.1016/j.jecp.2019.02.008

Baruch, C., and Drake, C. (1997). Tempo discrimination in infants. *Infant Behaviour and Development*, 20(4), 573–577.

Boto, E., Holmes, N., Leggett, J., et al. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, 555(7698), 657–661. https://doi.org/10.1038/nature26147

Boto, E., Meyer, S. S., Shah, V., et al. (2017). A new generation of magnetoencephalography: Room temperature measurements using optically-pumped magnetometers. *NeuroImage*, 149, 404–414. https://doi.org/10.1016/j.neuroimage.2017.01.034

Broesch, T., and Bryant, G. A. (2018). Fathers' infant-directed speech in a small-scale society. *Child Development*, 89(2), e29–e41. https://doi.org/10.1111/cdev.12768

Canolty, R. T., Edwards, E., Dalal, S. S., et al. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793), 1626–1628. https://doi.org/10.1126/science.1128115

Caskey, M., Stephens, B., Tucker, R., and Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, 128(5), 910–916. https://doi.org/10.1542/peds.2011-0609

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

Condon, W. S., and Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, 45(2), 456–462.

Cox, C., Bergmann, C., Fowler, E., et al. (2023). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, 7(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

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

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language*, 25(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Cummins, F., and Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

David, O., Kilner, J. M., and Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage*, 31(4), 1580–1591. https://doi.org/10.1016/j.neuroimage.2006.02.034

Decasper, A. J., and Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, 9(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

de la Cruz-Pavía, I., Gervain, J., Vatikiotis-Bateson, E., and Werker, J. F. (2020). Coverbal speech gestures signal phrase boundaries: A production study of Japanese and English infant- and adult-directed speech. *Language Acquisition*, 27(2), 160–186. https://doi.org/10.1080/10489223.2019.1659276

de Lange, P., Boto, E., Holmes, N., et al. (2021). Measuring the cortical tracking of speech with optically-pumped magnetometers. *NeuroImage*, 233, 117969. https://doi.org/10.1016/j.neuroimage.2021.117969

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

de l'Etoile, S. K., Bennett, C., and Zopluoglu, C. (2020). Infant movement response to auditory rhythm. *Perceptual and Motor Skills*, 127(4), 651–670. https://doi.org/10.1177/0031512520922642

Destoky, F., Bertels, J., Niesen, M., et al. (2020). Cortical tracking of speech in noise accounts for reading strategies in children. *PLoS Biology*, 18(8), e3000840. https://doi.org/10.1371/journal.pbio.3000840

Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

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

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., and Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, 116(20), 10113–10121. https://doi.org/10.1073/pnas.1816414116

Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica*, 55(4), 226–237. https://doi.org/10.1159/000028434

Ejiri, K., and Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4(1), 40–48. https://doi.org/10.1111/1467-7687.00147

Feys, O., Corvilain, P., Aeby, A., et al. (2022). On-scalp optically pumped magnetometers versus cryogenic magnetoencephalography for diagnostic evaluation of epilepsy in school-aged children. *Radiology*, 304(2), 429–434. https://doi.org/10.1148/radiol.212453

Flaten, E., Marshall, S. A., Dittrich, A., and Trainor, L. J. (2022). Evidence for top-down metre perception in infancy as shown by primed neural responses to an ambiguous rhythm. *European Journal of Neuroscience*, 55(8), 2003–2023. https://doi.org/10.1111/ejn.15671

Fransen, A. M. M., van Ede, F., and Maris, E. (2015). Identifying neuronal oscillations using rhythmicity. *NeuroImage*, 118, 256–267. https://doi.org/10.1016/j.neuroimage.2015.06.003

Fujii, S., Watanabe, H., Oohashi, H., et al. (2014). Precursors of dancing and singing to music in three- to four-months-old infants. *PLoS One*, 9(5), e103192. https://doi.org/10.1371/journal.pone.0097680

Gabard-Durnam, L. J., Leal, A. S. M., Wilkinson, C. L., and Levin, A. R. (2018). The Harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, 12, 97. https://doi.org/10.3389/fnins.2018.00097

Giraud, A. L. (2020). Oscillations for all ¯_(ツ)_/¯? A commentary on Meyer, Sun & Martin (2020). *Language, Cognition and Neuroscience*, 35(9), 1106–1113. https://doi.org/10.1080/23273798.2020.1764990

Golinkoff, R. M., Can, D. D., Soderstrom, M., and Hirsh-Pasek, K. (2015). (Baby)Talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, 24(5), 339–344. https://doi.org/10.1177/0963721415595345

Goswami, U., and Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4(1), 67–92. https://doi.org/10.1515/lp-2013-0004

Green, J. R., Nip, I. S. B., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *Journal of Speech, Language, and Hearing Research*, 53(6), 1529–1542. https://doi.org/10.1044/1092-4388(2010/09-0005)

Gross, J., Hoogenboom, N., Thut, G., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Haegens, S. (2020). Entrainment revisited: A commentary on Meyer, Sun, and Martin (2020). *Language, Cognition and Neuroscience*, 35(9), 1119–1123. https://doi.org/10.1080/23273798.2020.1758335

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

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech understanding. *Nature*, 8(5), 393–402. https://doi.org/10.1038/nrn2113

Hill, R. M., Boto, E., Holmes, N., et al. (2019). A tool for functional brain imaging with lifespan compliance. *Nature Communications*, 10(1), 4785. https://doi.org/10.1038/s41467-019-12486-x

Hilton, C. B., Moser, C. J., Bertolo, M., et al. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 6(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Holmes, N., Rea, M., Hill, R. M., et al. (2023). Naturalistic hyperscanning with wearable magnetoencephalography. *Sensors*, 23(12), 5454. https://doi.org/10.3390/s23125454

Inbar, M., Grossman, E., and Landau, A. N. (2020). Sequences of intonation units form a ~ 1 Hz rhythm. *Scientific Reports*, 10, 15846. https://doi.org/10.1038/s41598-020-72739-4

Inbar, M., Genzer, S., Perry, A., Grossman, E., and Landau, A. N. (2023). Intonation units in spontaneous speech evoke a neural response. *Journal of Neuroscience*, 43(48), 8189–8200. https://doi.org/10.1523/JNEUROSCI.0235-23.2023

Iverson, J. M., and Thelen, E. (1999). Hand, mouth and brain: The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, 6(11–12), 19–40. www.imprint-academic.com/jcs

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

Jessen, S., Obleser, J., and Tune, S. (2021). Neural tracking in infants: An analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience*, 52, 101034. https://doi.org/10.1016/j.dcn.2021.101034

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

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, 16(3), 1–19. https://doi.org/10.1371/journal.pbio.2004473

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831–843. https://doi.org/10.1038/nrn1533

Lachaux, J. P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

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

Leong, V., and Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, 8(1), 1–14. https://doi.org/10.3389/fnhum.2014.00096

Leong, V., and Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One*, 10(12), 1–37. https://doi.org/10.1371/journal.pone.0144411

Leong, V., Kalashnikova, M., Burnham, D., and Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind*, 1(2), 78–90. https://doi.org/10.1162/opmi_a_00008

Lopez, K. L., Monachino, A. D., Morales, S., et al. (2022). HAPPILEE: HAPPE in low electrode electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage*, 260, 119390. https://doi.org/10.1016/j.neuroimage.2022.119390

Ma, W., Golinkoff, R. M., Houston, D. M., and Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, 7(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, 19(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Mandke, K., Flanagan, S., Macfarlane, A., et al. (2022). Neural sampling of the speech signal at different timescales by children with dyslexia. *NeuroImage*, 253, 119077. https://doi.org/10.1016/j.neuroimage.2022.119077

Maurer, D., and Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, 56(2), 154–178. https://doi.org/10.1002/dev.21177

Medvedovsky, M., Taulu, S., Bikmullina, R., and Paetau, R. (2007). Artifact and head movement compensation in MEG. *Neurophysiology and Neuroscience*, 29(4), PMID: 18066426.

Mehler, J., Bertoncini, J., Barriere, M., and Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, 7(5), 491–497. https://doi.org/10.1068/p070491

Menn, K. H., Männel, C., and Meyer, L. (2023). Does electrophysiological maturation shape language acquisition? *Perspectives on Psychological Science*, 18(6), 1271–1281. https://doi.org/10.1177/17456916231151584

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

Meyer, L., Sun, Y., and Martin, A. E. (2020). 'Entraining' to speech, generating language? *Language, Cognition and Neuroscience*, 35(9), 1138–1148. https://doi.org/10.1080/23273798.2020.1827155

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., and Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, 37(8), 2767–2783. https://doi.org/10.1002/hbm.23206

Nespor, M., Shukla, M., and Mehler, J. (2011). Stress-timed vs. syllable-timed languages. In M. Oostendorp, C. J. Ewen, E. Hume, and K. Rice (eds.), *The Blackwell Companion to Phonology* (pp. 1147–1159). Wiley-Blackwell.

Nguyen, T., Abney, D. H., Salamander, D., Bertenthal, B. I., and Hoehl, S. (2021a). Proximity and touch are associated with neural but not physiological synchrony in naturalistic mother–infant interactions. *NeuroImage*, 244, 118599. https://doi.org/10.1016/j.neuroimage.2021.118599

Nguyen, T., Schleihauf, H., Kayhan, E., et al. (2021b). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns. *Social Cognitive and Affective Neuroscience*, 16(1–2), 93–102. https://doi.org/10.1093/scan/nsaa079

Nguyen, T., Schleihauf, H., Kungl, M., et al. (2021c). Interpersonal neural synchrony during father–child problem solving: An fNIRS hyperscanning study. *Child Development*, 92(4), e565–e580. https://doi.org/10.1111/cdev.13510

Ní Choisdealbha, Á., Attaheri, A., Rocha, S., et al. (2024). Cortical tracking of visual rhythmic speech by 5- and 8-month-old infants: Individual differences in phase angle relate to language outcomes up to 2 years. *Developmental Science*, 27, e13502. https://doi.org/10.1111/desc.13502

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, 31(28), 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011

Pascual-Marqui, R. D., Lehmann, D., Koukkou, M., et al. (2011). Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1952), 3768–3784. https://doi.org/10.1098/rsta.2011.0081

Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press

Rocha, S., and Mareschal, D. (2017). Getting into the groove: The development of tempo-flexibility between 10 and 18 months of age. *Infancy*, 22(4), 540–551. https://doi.org/10.1111/infa.12169

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

Rocha, S., Attaheri, A., Choisdealbha, A. N., et al. (2024). Infant sensorimotor synchronisation to speech and non-speech rhythms: A longitudinal study. *Developmental Science*, 27, e13483. https://doi.org/10.1111/desc.13483

Romberg, A. R., and Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914. https://doi.org/10.1002/wcs.78

Smith, N. A., and Strader, H. L. (2014). Infant-directed visual prosody. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*, 15(1), 38–54. https://doi.org/10.1075/is.15.1.02smi

Teng, X., Cogan, G. B., and Poeppel, D. (2019). Speech fine structure contains critical temporal cues to support speech segmentation. *NeuroImage*, 202, 116152. https://doi.org/10.1016/j.neuroimage.2019.116152

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology*, 17(3), 237–257. https://doi.org/10.1037/0012-1649.17.3.237

Vander Ghinst, M., Bourguignon, M., Niesen, M., et al. (2019). Cortical tracking of speech-in-noise develops from childhood to adulthood. *Journal of Neuroscience*, 39(15), 2938–2950. https://doi.org/10.1523/JNEUROSCI.1732-18.2019

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

Yu, L., and Myowa, M. (2021). The early development of tempo adjustment and synchronization during joint drumming: A study of 18- to 42-month-old children. *Infancy*, 26(4), 635–646. https://doi.org/10.1111/infa.12403

Zan, X. P., Presacco, A., Anderson, S., and Simon, J. Z. (2019). Sensory processing mutual information analysis of neural representations of speech in noise in the aging midbrain. *Journal of Neurophysiology*, 122, 2372–2387. https://doi.org/10.1152/jn.00270.2019

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

https://doi.org/10.1017/9781009295888.043 Published online by Cambridge University Press