35 Rhythm and Language Acquisition: A Temporal Sampling Perspective

USHA GOSWAMI

35.1 Introduction

The importance of rhythm for language processing is a core theme throughout this volume. Both the acoustic structure of the speech signal and the physiology of speech perception and production are intimately tied to rhythm, and so the proposal that rhythm is also at the centre of language *acquisition* is perhaps unsurprising. Rhythm can be understood as a kind of hidden glue that underpins the acoustic structure of human languages and the representation of sensory structure by human brains. Indeed, almost four decades ago, Mehler et al. (1988) proposed that sensitivity to rhythm was a universal precursor to language acquisition by young infants. Here, I adapt temporal sampling (TS) theory (Goswami, 2011) to provide a rhythm-based framework for explaining language acquisition by infants (see also Goswami, 2019, 2020, 2022a). TS theory was originally proposed as a causal framework for integrating the acoustic and motor difficulties with rhythm found in children with dyslexia with an oscillatory model of sensory encoding. The acoustic rhythmic difficulties in dyslexia were indexed primarily by sensory difficulties in discriminating changes in speech energy called amplitude rise times (ARTs, depicted in Figure 35.1), and by perceptual difficulties in discriminating syllable stress patterns and musical rhythm patterns, along with problems in tapping in time with a regular metronome beat (e.g., Thomson and Goswami, 2008; Huss et al., 2011; Goswami et al., 2013). The rhythmic neural encoding aspects of TS theory were focused on low-frequency oscillations in the delta (0.5–4 Hz) and theta (4–8 Hz) neurophysiological bands, thought to play a core role in rhythmic entrainment (Ghitza and Greenberg, 2009; see Chapter 3). These features were integrated into a sensory/neural explanatory system for understanding the phonological difficulties that characterise children with dyslexia, and that also characterise many children with what was previously termed specific language impairment (SLI – now termed developmental language disorder, DLD). Since 2011, TS theory has been developed further in the light of experimental data. Following an outline of TS theory drawing on sensory/neural data from adults and children, two longitudinal studies testing the theory with infant participants will be described in detail.

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

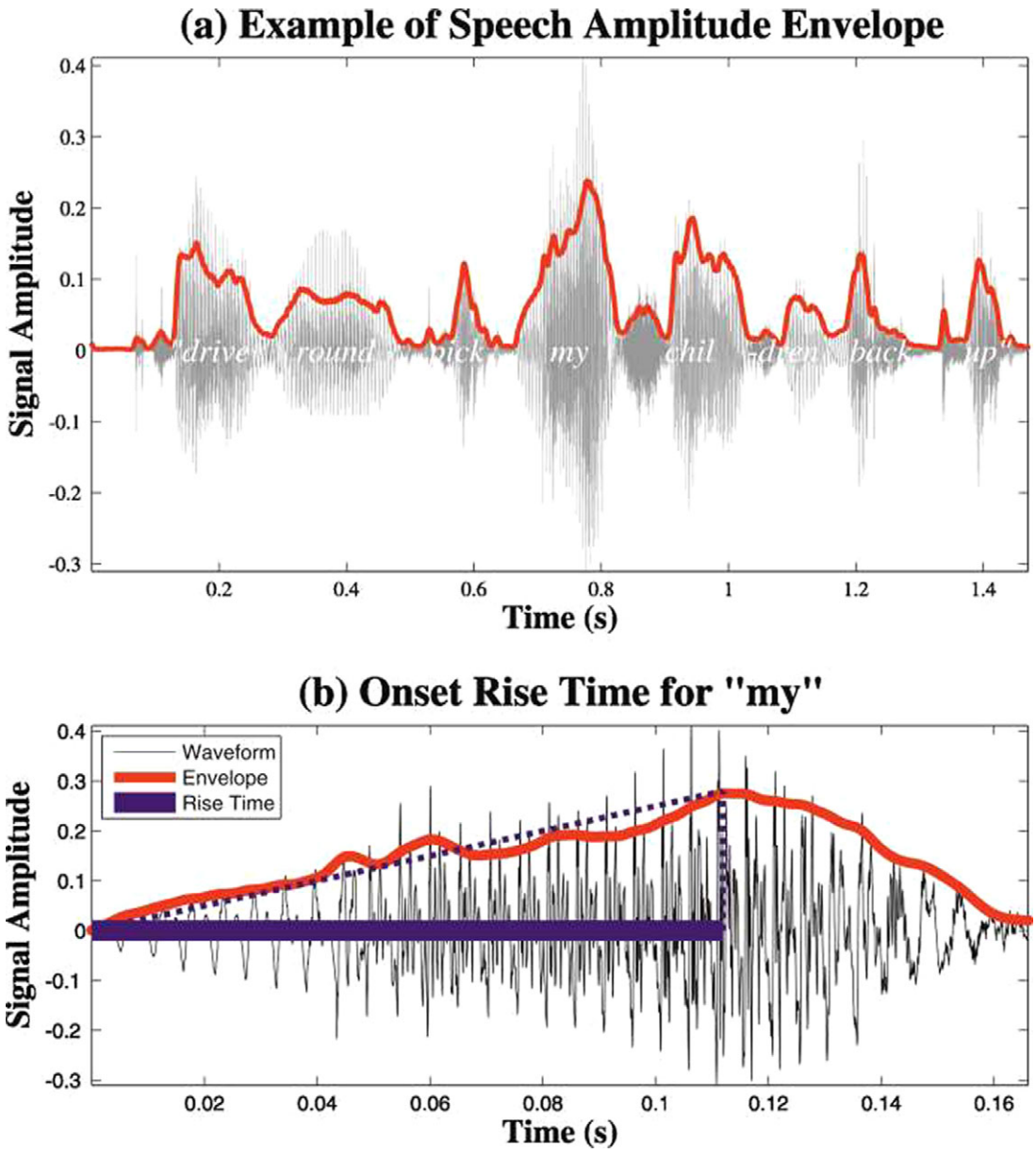


Figure 35.1 A schematic depiction of ARTs.
Panel 1A depicts the amplitude envelope for the sentence 'drive round, pick my children back up', and Panel 1B shows the rise time for the syllable 'my', taken from the longer sentence.
Used with permission of Walter de Gruyter and Company from Goswami and Leong, 2013; permission conveyed through Copyright Clearance Center, Inc.

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

35.2 Temporal Sampling Theory

Temporal 'sampling' refers to the fact that the brain depends on networks of neurons that oscillate between excitation and inhibition to encode sensory information ('brain waves'). Temporal coding is an important aspect of information coding in the brain (Buzsáki and Draghun, 2004), and for encoding the complex speech signal, the synchronous activity of oscillating networks of neurons at different timescales (frequency bands, primarily delta, 0.5–4 Hz, theta, 4–8 Hz, beta, 15–25 Hz, and gamma, 30–80 Hz) is thought to be critical (Giraud and Poeppel, 2012; see also Obleser et al., 2012). Neurophysiology shows that these frequency bands are ubiquitous throughout the brain and encode many aspects of sensory information (Buzsáki, 2006). The acoustic speech signal can be modelled as the summation of a number of frequency bands fluctuating in intensity (amplitude) over time (the amplitude envelope, AE; Drullman, 2006). The auditory system codes amplitude modulations (AMs) in natural sounds across different frequency channels and at different timescales (Joris et al., 2004; Turner, 2010), and there is systematic AM structure in the AE of all natural sounds, not just speech (Turner, 2010; see Daikoku and Goswami, 2022, for AM structure of music, birdsong, wind, and rain). The AE of both child-directed speech (CDS) such as nursery rhymes and infant-directed speech (IDS) or BabyTalk can be analysed in terms of its constituent temporal modulation frequencies (Leong and Goswami, 2015; Leong et al., 2017). This computational modelling shows that the dominant modulation frequency bands for these highly rhythmic speech registers cluster into three AM bands, with centre frequencies around ~2 Hz, ~5 Hz, and ~20 Hz, thereby approximately matching the electrophsyiological bands of delta, theta, and beta/low gamma. These insights concerning the AM structure of CDS and IDS fit well with adult multi-time resolution speech processing models (MTRMs) (see Poeppel, 2003; Poeppel et al., 2008; also see Chapters 3 and 5), models that played a core role in the development of TS theory (see Goswami, 2011).

35.2.1 TS Theory: Sensory Data

Adult MTRMs proposed that the different temporal integration windows characterised by the different oscillatory networks yielded packets of information at different linguistic grain sizes that were matched to stored representations of words in the mental lexicon (Poeppel et al., 2008). Adult MTRMs focused on the linguistic units of the syllable and the phoneme, thought to be captured by temporal integration windows in the theta and gamma bands, respectively. The infant and child speech modelling studies discussed in Section 35.2 showed that the stressed syllable is also an important linguistic grain size, and is captured by temporal integration windows in the delta band (Leong and Goswami, 2015). Leong and Goswami's speech modelling revealed that if one cycle of AM at each temporal rate (delta (~2 Hz), theta (~5 Hz), and beta/low gamma (~20 Hz)) in English nursery rhymes is assumed to match one level of linguistic structure (stressed syllable, syllable, onset-rime), then the AM structure in the AE of CDS provides sufficient sensory information for the brain to extract phonological units of different sizes with over 80–90% efficiency (the spectral-amplitude modulation hierarchy model, S-AMPH, Leong and Goswami, 2015). This suggests that applying an MTRM approach to the simplified rhythmic

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

genres of IDS and CDS, which underpin language acquisition across cultures, could enable 'acoustic-emergent' phonology. When brain rhythms and speech rhythms at these different rates (delta, theta, beta/low gamma) are in temporal alignment, then an emergent phonological system can be extracted from the speech signal by automatic acoustic statistical learning processes (see Goswami, 2022a, for detail; statistical learning is required as speech–brain alignment is only probabilistically accurate in yielding phonological units). This automatic learning can, in principle, enable perceptual organisation of the acoustic speech stream into syllable stress patterns (prosodic feet), syllables, and onset-rime units (m-ash, cr-ash, spl-ash). Further, automatic learning of higher-level acoustic prosodic hierarchies that are also present in the sensory signal could enable the perceptual organisation of acoustic information relevant to extracting syntax and grammar, such as prosodic and intonational phrasing (the 'prosodic phrasing' hypothesis; see Cumming et al., 2015). Perceptual detection of prosodic hierarchies also depends on accurate ART perception (Cumming et al., 2015).

In TS theory, rhythm patterns are hypothesised to be core to these organisational processes. The S-AMPH computational modelling work based on nursery rhymes further revealed that each AM band is in temporal synchronicity with its adjacent band, via phase relations. Phase relations are temporal relations between different cycles of rhythmic timing, such as slower versus faster bands of AM. In IDS and CDS, these phase relations are governed by the delta AM band (Leong et al., 2014). In particular, phase relations between the delta (~2 Hz) and theta (~5 Hz) AM bands in nursery rhymes are critical to perceiving their 'beat' structure or metrical structure (e.g., trochaic versus iambic; Leong et al., 2014, Leong and Goswami, 2015). Delta–theta AM phase alignment is also core to perceiving rhythm in music (Daikoku and Goswami, 2022). Indeed, the finding that the delta AM band sits at the top of the phase hierarchy in CDS and IDS is thought-provoking, as its centre rate of ~2 Hz matches the dominant beat rate of music (120 bpm) (Fraisse, 1982; see also Chapters 26 and 27). Temporal analyses of the lullabies sung by mothers to their infants across cultures reveals a cross-cultural convergence on this beat rate of ~2 Hz (Trehub and Trainor, 1998). Accordingly, AM-based modelling of the speech signal suggests that cultural practices such as BabyTalk and lullabies may facilitate the automatic extraction of phonology and possibly syntax via the AM phase hierarchy, by statistical sensory learning beginning in the cradle. Infant auditory statistical learning is already known to be highly efficient (e.g., Saffran, 2001), and while IDS and CDS are simplified speech registers, they nevertheless encompass more rhythmic categories than trochaic versus iambic, for example anapest rhythms.

TS theory proposes that this automatic statistical learning depends intimately on ART discrimination, as different ARTs are associated with different temporal modulation rates. TS theory (Goswami, 2011) was in part developed because psychoacoustic studies had repeatedly demonstrated impaired ART discrimination by children with dyslexia, across languages (to date English, French, Finnish, Dutch, Spanish, Chinese, and Hungarian, summarised in Goswami, 2015). Difficulties in discriminating ARTs (which index the rate of change of AMs) could be expected to cause impairments in distinguishing the different modulation frequency ranges in speech, which would result in inefficient phase-locking to these frequency ranges by neuronal oscillatory networks (Goswami, 2011). Adult neuroscience studies were able to demonstrate that ARTs (called 'speech edges' in that literature) played a key mechanistic role in the alignment of brain rhythms and speech rhythms (Gross et al., 2013; Doelling et al., 2014; Chapter 8). For

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

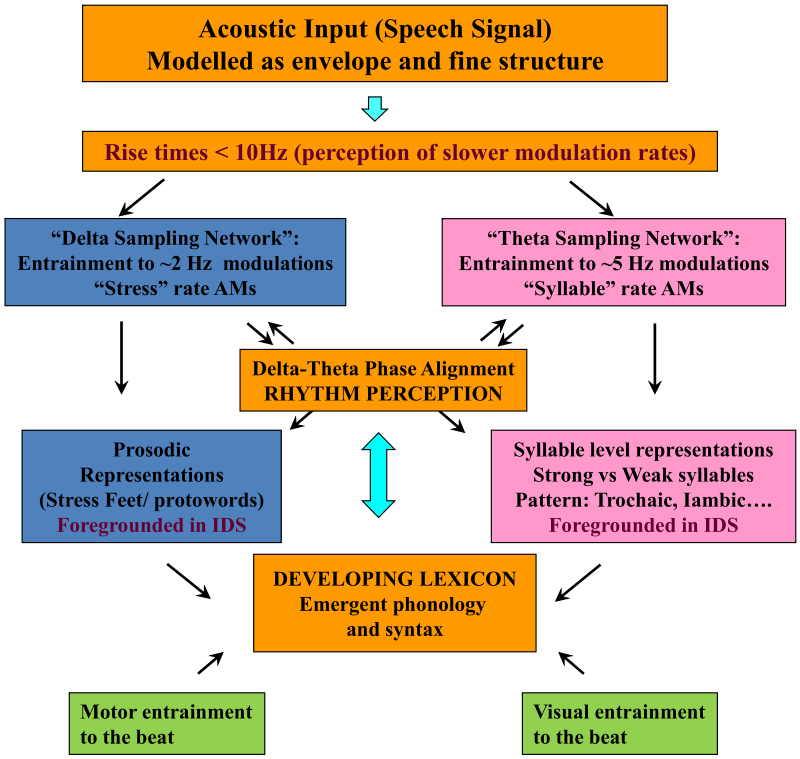


Figure 35.2 TS theory and infant language acquisition.
The schematic depiction of TS theory emphasises the core role of ART and rhythm processing via automatic neural entrainment to AMs at ~2 Hz (delta) and ~5 Hz (theta) rates, sensory-neural processes that support the developing lexicon. Although only briefly discussed in this chapter, individual differences in rhythmic entrainment to visual speech and motor entrainment may also have an impact on the developing lexicon.
Reproduced from Goswami (2022a); licensed under CC BY 4.0.

example, ARTs in the theta band provide sensory landmarks that automatically trigger brain rhythms and speech rhythms into temporal alignment, via phase-resetting ongoing neural activity (Doelling et al., 2014). Further, the oscillatory networks that are responsive to speech inputs are also hierarchically organised, with delta at the top (Gross et al., 2013). Accordingly, prior research with children and adults shows that there is a mechanistic neural hierarchy of oscillations, an AM phase hierarchy in IDS and CDS, and a linguistic hierarchy of phonological units that all match in terms of their intrinsic relational structure. This provides a sensory/neural framework for language acquisition via automatic statistical learning of this relational AM structure, which is depicted in simplified form in Figure 35.2.

35.2.2 TS Theory: Neural Data

Children with dyslexia have phonological impairments at many levels of the linguistic hierarchy, showing difficulties in identifying and manipulating phonological units such as stressed syllables, syllables, onset-rimes, and phonemes in oral tasks (Goswami, 2022b, for a recent

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

review). If automatic statistical learning of AM phase relations supports acoustic-emergent phonology, and given the well-established sensory discrimination difficulties with ARTs found in children with dyslexia, it is logical to expect atypical oscillatory responses to natural speech in these children. The first studies investigating the neural aspects of TS theory in children predicted atypical responding in both the delta and theta bands. These early studies used a rhythmic speech paradigm, in which a 'talking head' repeated the syllable 'ba' at 2 Hz while an electroencephalogram (EEG) was recorded (Power et al., 2012, 2013). Children were listening for oddball syllables that were slightly out of time with the 2 Hz rhythm, and pressed a button when this occurred. Even when children with dyslexia were equated to control children without dyslexia for their performance in the button-press task, their brains were out of time regarding the delta-band response (but not the theta-band response; Power et al., 2013). The key finding was that the oscillatory response in the delta band had a different preferred phase (preferred phase refers to the point in time when most neurons are discharging their electrical potentials). Brain excitation in the delta band was maximal at a different point in time for the group of children with dyslexia compared to control children. By TS theory, this would suggest that the peak delta-band neural response for dyslexic children was aligned with less informative points in the speech stimulus, a phase difference that might be expected to impair the neural representation of phonological information.

Subsequent dyslexia studies testing TS theory in English indicated that the dyslexic brain encoded a less accurate representation of delta-band envelope information (Power et al., 2016; Keshavarzi et al., 2022; Mandke et al., 2022). These investigations recorded either EEG or MEG (magnetoencephalography) during speech listening tasks, and either used a method called the multivariate temporal response function (mTRF) to estimate different features of the heard speech such as the speech AE from the EEG responses (generating a correlational r value; Di Liberto et al., 2015; Crosse et al., 2016), or other methods such as phase-locking values (PLVs) in MEG to estimate the cortical encoding of speech in different frequency bands. These developmental studies showed that neural representation in the delta band was less accurate (exhibiting a significantly lower mTRF r value) in a sentence-listening task for children with dyslexia compared to both age-matched control children and younger reading-level-matched control children using EEG (Power et al., 2016). It was also less accurate in a story-listening task compared to age-matched control children in neuroimaging studies using MEG (exhibiting a significantly lower PLV; Mandke et al., 2022) and EEG (exhibiting a significantly lower mTRF r value; Keshavarzi et al., 2022). In the English EEG studies, neural accuracy was significantly related to measures of phoneme awareness and non-word reading, and in the MEG study, to vocabulary development. Neuroimaging studies in other languages have also found impaired low-frequency envelope encoding for children with dyslexia (in both the delta and theta bands; see Destoky et al., 2020, 2022, French) and atypical speech–brain synchronisation in the delta band (see Molinaro et al., 2016, Spanish). Importantly, speech reconstruction studies with infants, children, and adults (in which EEG recorded during speech listening is used to reconstruct low-frequency speech information <8 Hz using mTRFs) show that delta-band information yields phonetic information as well as prosodic (syllable stress) information (Di Liberto et al., 2015, 2018, 2023). Accordingly, this review of neural data supporting TS theory indicates that low-frequency envelope information is encoded less accurately by children with dyslexia, and that this impaired encoding affects the quality of their phonological development at many levels of the linguistic hierarchy, including the phoneme level.

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

35.2.3 Implications for a TS Theory of Language Acquisition

The sensory and neural data with children gathered to test TS theory have a number of implications for language acquisition. Regarding sensory factors, if ART sensitivity could be measured in infancy, it would be expected to be a significant predictor of linguistic development. As infants get older and become able to participate in metrical rhythm tasks, individual differences in rhythm perception and production might also be expected to relate to ART discrimination and to predict linguistic development. These sensory predictions of TS theory were investigated in the SEEDS project. Regarding neural factors, both the accuracy of continuous speech encoding in the delta and theta bands as measured by mTRFs and the timing of oscillatory responses (such as preferred phase) in the delta band to rhythmic inputs might be expected to predict later language outcomes. These predictions of TS theory were investigated in the Cambridge UK BabyRhythm project. Both were longitudinal infant projects, one with an at-risk sample (SEEDS) and one with a typically developing sample (BabyRhythm), and an overview of their findings is presented here.

35.3 SEEDS: Infants at Family Risk for Dyslexia

The SEEDS (of Literacy) project recruited a group of infants at family risk for dyslexia (FR group) and a group of infants not at family risk for dyslexia (NFR group) in Australia, and followed their development from the age of five months. One aim of SEEDS was to investigate whether infant measures of auditory sensitivity to rhythm would be related to subsequent linguistic and phonological skills and, eventually, to reading achievement (Kalashnikova et al., 2018, 2019, 2020a, 2020b, 2020c, 2021). Children were assigned to the FR and NFR groups at the beginning of the project based on a parent having a pre-existing dyslexia diagnosis and/or their performance on a comprehensive parental screening battery that included language, reading, and cognitive tasks. This targeted sample proved difficult to recruit; hence, most publications are based on samples of 50 infants or fewer.

An infant behavioural procedure for assessing sensitivity to ART was developed based on the psychoacoustic stimuli used to measure ART discrimination in younger children (Corriveau et al., 2010). An infant version of a two-alternative forced choice (2AFC) adaptive threshold procedure was created based on looking preference. Infants sat on their parent's lap facing three monitors in a sound-attenuated infant-testing booth. Once infants fixated the central monitor, the images of a chequerboard appeared on the right and left screens. Infants' fixations to the monitor on one side produced a repeating auditory stimulus with a fixed rise time (e.g., 15 ms, 15 ms, 15 ms, 15 ms . . .), while fixations to the monitor on the other side produced an auditory stimulus with alternating rise times (15 ms, 300 ms, 15 ms, 300 ms . . .). Greater fixation to the alternating stimulus was taken as a measure of discrimination, and step size was reduced accordingly (e.g., 15 ms, 270 ms, 15 ms, 270 ms . . .), enabling a threshold (a just noticeable difference in ART) to be established (see Kalashnikova et al., 2018, for a full description). When the FR and NFR groups were aged 10 months, they already showed a significant difference in ART discrimination in this task (Kalashnikova et al., 2018), as would be expected by TS theory.

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

Early auditory rhythm sensitivity as indexed by the ART measure administered at 10 months was a significant predictor of subsequent vocabulary development (Kalashnikova et al., 2019). When they were aged 36 months, the SEEDS cohort were given a standardised vocabulary measure from the Stanford-Binet (Roid, 2003). They were also given two experimental measures of phonological development at 42 months. These were a mispronunciation detection task (e.g., 'abble' for apple) and a non-word repetition task (the children orally copied items such as 'sep', 'gattom', and 'katapet'). A linear mixed-effects analysis showed that ART discrimination at age 10 months was a predictor of vocabulary at 36 months of age, but not (contrary to hypothesis) of phonological development at 42 months (Kalashnikova et al., 2019). The phonological tasks did not yet show group differences, possibly due to reduced sensitivity at this age. By contrast, infants who were less sensitive to ART had smaller vocabularies, and reduced vocabulary development is known to be an early hallmark of risk for dyslexia. Prospective longitudinal studies of children at family (genetic) risk of dyslexia frequently report deficits in expressive vocabulary (e.g., at 17 months; Koster et al., 2005), and van Viersen et al. (2017) showed that children at risk for dyslexia exhibit delayed growth patterns of both receptive and expressive vocabulary sizes from 17 to 35 months. In the latter study, these early vocabulary scores reliably discriminated between at-risk children who later did and did not develop dyslexia.

One explanation for reduced vocabulary development in FR samples could be that reduced sensitivity to ARTs impairs the learning of new words. This would be predicted by TS theory if infants with poorer rise time sensitivity process the speech signal less effectively. Prior research with children with dyslexia has shown a significant relationship between ART discrimination and the ability to learn novel words (Thomson and Goswami, 2010). Thomson and Goswami reported that children with dyslexia were worse at learning novel words in a paired associate learning (PAL) task compared to both age-matched and younger reading-level-matched control children. Within the dyslexic group, children with better ART sensitivity performed better in the PAL task. To test the possibility that reduced ART sensitivity impairs novel word learning, a 'fast mapping' task was administered when the SEEDS cohort were aged 19 months (Kalashnikova et al., 2020a). Two non-word items, 'lif' and 'wug', were presented repeatedly along with two novel visual objects in a habituation paradigm, and word learning was tested in a subsequent preferential looking paradigm. Both groups of infants showed similar patterns of habituation, and similar engagement and attention during the learning component of the task. However, while the NFR infants showed increased looking time to the correct novel referent at test, the FR infants did not. The FR infants were hence less successful at learning new words early in the language acquisition process, as would be expected on the basis of TS theory.

Of course, it could be argued that the FR infants were simply poorer at PAL. This explanation was ruled out by a subsequent study of PAL when the SEEDS sample were aged 48 and 54 months (Kalashnikova et al., 2020b). In the later PAL task, the SEEDS children were required to learn associations between four novel words and four novel objects, with learning assessed by (1) finding the correct object in response to its label, and (2) producing the novel word when shown the corresponding object. Their letter knowledge was also assessed as a measure of pre-reading skill at both ages, and vocabulary was measured again at 48 months. Kalashnikova et al. (2020b) found no deficit in PAL for the FR children on either measure of

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

learning, and no relationship between PAL and letter knowledge. However, the FR children still had significantly smaller vocabularies than the NFR children. Accordingly, linguistic development is still compromised, even though the difficulties in novel word learning present at 19 months for the FR children could no longer be demonstrated using the later PAL task at 48 and 54 months. Further analyses revealed that it was earlier performance in the non-word repetition task given at 42 months that predicted PAL. This suggests developmental continuity between early auditory difficulties, subsequent early word-learning difficulties, individual differences in later phonological sensitivity, and subsequent vocabulary learning.

By four years of age (48 months), the SEEDS sample were able to complete some of the measures of rhythmic sensitivity used in the TS studies of children with dyslexia. These measures comprised a rhythm discrimination task based on musical beats (following Huss et al., 2011, hereafter musical beat perception) and a measure of rhythm production based on tapping to a 2 Hz metronome beat (following Thomson and Goswami, 2008; see Kalashnikova et al., 2021). By TS theory, both measures of rhythmic processing should be related to phonological development and should be impaired in the FR group. The non-word repetition data gathered at 42 months was thus studied in relation to rhythm discrimination and rhythm production at 48 months. When aged four years, the FR group showed significantly poorer rhythm discrimination (significantly lower 'd' scores) in the musical beat perception task compared to the NFR group (Kalashnikova et al., 2021). There was no group difference in rhythm production (tapping accuracy), although the FR group displayed more variable performance (Kalashnikova et al., 2021; Figure 35.2). As noted above, when aged four years the SEEDS sample also received measures of letter knowledge and vocabulary. The FR and NFR children showed significant group differences in all three linguistic tasks (non-word repetition, letter knowledge, and vocabulary development), with the FR children showing consistently worse performance. Correlation analyses showed that tapping variability was significantly related to performance on all three linguistic tasks, whereas variability in musical beat perception was not. A mediation analysis showed that metronome tapping was a significant predictor of non-word repetition, and non-word repetition was a significant predictor of letter knowledge (Kalashnikova et al., 2021). Accordingly, the link between rhythm and pre-literacy skills (letter knowledge) is mediated by phonological processing (non-word repetition). Again, these longitudinal relations are consistent with TS theory. In future SEEDS analyses, we plan to investigate developmental continuities between ART discrimination, phonological development, and early literacy.

35.4 Cambridge UK BabyRhythm Project: Typically Developing Infants

In order to test neural predictions of TS theory with infants, large samples are required. EEG data collected from infants is quite noisy, and there can be substantial loss of data. The Cambridge UK BabyRhythm project was able to recruit 122 typically developing infants, who were followed from the age of two months to 42 months with very little drop-out. Infants attended for eight EEG sessions at the ages of two, four, five, six, seven, nine, and 11 months, and language outcomes were measured at 12, 15, 18, 24, 30, and 42 months. To date, only

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

language analyses up to 24 months are available (see also Chapter 36). Based on TS theory, three tasks were selected for the EEG recordings: a nursery rhyme task and two rhythmic tasks. Infants watched videos of a 'talking head' singing nursery rhymes in IDS, of a 'talking head' saying 'ta' repeatedly at 2 Hz in IDS, and of a visual presentation of a ball dropping on to a drum accompanied by a metronome beat at 2 Hz. These three audiovisual tasks enabled us to assess cortical tracking of natural speech in the delta and theta bands by applying mTRF analyses to the nursery rhyme EEG, and to generate measures of preferred phase using the EEG recorded to both speech ('ta') and non-speech (drumbeat) rhythmic inputs.

These neural measures were used to predict a range of language outcomes. Although many different experimental measures of phonology and grammar were employed, along with standardised measures such as the UK version of the CDI (communicative development inventory; Alcock et al., 2020), the Covid-19 pandemic had a very disruptive effect on data collection, and only some of the language outcome measures could be used for the neural prediction analyses. The most robust measures were the UK-CDI, in which parents estimated their infants' word comprehension and production, an infant-led vocabulary measure called the CCT (computerised comprehension task; Friend and Keplinger, 2003), in which infants selected a named object on a tablet screen to indicate comprehension, and a non-word repetition task introduced as a game about naming toys. The vocabulary measures were administered from 12 months of age, and the non-word repetition task from 18 months of age (see Rocha et al., 2024, for full analyses).

On the basis of prior TS studies with children with dyslexia, it was expected that the accuracy of encoding of envelope information in the delta band along with individual differences in preferred phase in the delta band could be predictors of language outcomes. Regarding the accuracy of low-frequency envelope encoding, the EEG recorded in response to the sung nursery rhymes presented at the four-, seven-, and 11-month laboratory sessions was analysed using the mTRF method. These longitudinal analyses compared the accuracy of the speech envelope information encoded in the delta (0.5–4 Hz), theta (4–8 Hz), and alpha (8–12 Hz, a control band) bands at each age. Encoding accuracy was estimated for the first 60 infants in the UK BabyRhythm sample (Attaheri et al., 2022), and is currently being replicated with the remaining infants (Attaheri et al., 2024). In Attaheri et al. (2022), delta-band encoding was significantly stronger (significantly larger r values) than theta-band encoding at four months, and continued to be stronger at all subsequent ages, with no evidence for significant alpha-band encoding at any age. Other neural factors related to speech encoding by adults, such as phase-amplitude coupling (PAC) measures, were also present from the earliest measurement point (four months). Adult speech-encoding studies have highlighted theta–gamma PAC as particularly relevant for speech processing (Hyafil et al., 2015). Our infant data showed that delta-led PAC was equally important as theta-led PAC during the nursery rhyme listening task, with delta–beta, delta–gamma, theta–beta, and theta–gamma PAC all significant at all ages (four, seven, and 11 months) (Attaheri et al., 2022). In ongoing analyses (Attaheri et al., 2024), language data collected between 12 and 24 months is being related to the neural data. Currently, only theta–gamma PAC appears to be a significant source of individual differences. Stronger theta–gamma PAC is related to better language outcomes, while delta-related PAC is not (Attaheri et al., 2024). The other neural predictor of better language outcomes as measured

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

by the mTRF analyses was the accuracy of delta-band encoding at 11 months (Attaheri et al., 2024). Worse language outcomes were related to the theta–delta power ratio (higher ratios were associated with worse outcomes), and to the overall amount of theta power (higher theta power spectral density was associated with worse language outcomes; Attaheri et al., 2024). Accordingly, both delta-band and theta-band oscillatory responses and their dynamic interactions are important for language acquisition, as would be expected on the basis of TS theory.

Given the earlier data collected with children with dyslexia (Power et al., 2013), the preferred phase of entrainment in the delta band to rhythmic speech should also relate to individual differences in language outcomes. Preferred phase analyses were conducted for the two rhythmic tasks ('ta' and drumbeat), which were administered at two, six, and nine months in the auditory-only or auditory-visual (AV) domain, and at five and eight months as visual-only (VO) speech. Phase analyses revealed that infants already exhibited phase consistency to rhythmic syllables and drumbeats at two months of age, when most participants were sleeping and hence experienced auditory-only input (Ní Choisdealbha et al., 2023). Individual phase angles in the syllable task at two months were a significant predictor of later vocabulary outcomes as measured by the CCT at 18 months. By the time the infants were aged nine months, phase angles in the syllable task were associated with the number of parent-estimated known words at 24 months on the UK-CDI, for both comprehension and production. Regarding the drumbeat task, phase angles at nine months were predictive of non-word repetition skills at the age of 24 months, an early measure of phonological sensitivity (Ní Choisdealbha et al., 2023).

For VO speech, similar results were obtained (Ní Choisdealbha et al., 2024). At five months, infants already exhibited an increased cortical response to the stimuli at the stimulus frequency (thereby showing cortical tracking at 2 Hz), but phase consistency was variable. At eight months, the infants showed significant phase consistency to VO speech, and oscillatory responses were moving towards a common preferred phase angle. Individual differences in phase angle at eight months were significantly related to later vocabulary at 24 months. Taken together, the analyses reported by Ní Choisdealbha et al. show that individual differences in the phase angle of the neural response to rhythmic speech stimuli are predictive of later language outcomes, for auditory-only, VO, and AV speech. Accordingly, key neural parameters of TS theory explain some of the individual variation in children's later language outcomes.

Ideally, the Cambridge UK BabyRhythm study would have explored whether there was an association between individual differences in ART sensitivity and these different neural outcome measures. However, the 2AFC behavioural paradigm from the SEEDS study did not yield reliable data with the UK BabyRhythm infants. As the infants were still wearing the EEG cap when they received the 2AFC paradigm, instead of using behavioural thresholds we used their ERPs as an index of ART sensitivity (Ní Choisdealbha et al., 2022). The ERP data revealed that the infants showed robust mismatch responses to all the steps on the thresholding staircase, including the ones that we had expected to be below threshold on the basis of the SEEDS data. Even the smallest just noticeable difference of 161 ms in rise time (15 ms versus 176 ms) was discriminated by 85% of the sample at seven months of age, and there was no consistent pattern to the presence or absence of a mismatch response across the 10 values used for the staircase. This finding shows that infants are able to discriminate ARTs with high

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

accuracy, revealing the readiness of the infant brain to process speech rhythm. The ART data also demonstrate the value of neural measures in assessing the parameters of TS theory. Future studies will need to present more fine-grained differences between ART stimuli in order to find individual infant thresholds and relate them to language outcomes.

35.5 Conclusions

Applying TS theory to infant language acquisition appears to be a promising avenue for future research. Referring back to Figure 35.2, the SEEDS and the Cambridge UK BabyRhythm projects have provided evidence for a number of factors theorised to be critical to language acquisition. The UK BabyRhythm project has shown that the infant brain is highly sensitive to ART differences (Ní Choisdealbha et al., 2022), and entrains to delta-band and theta-band speech envelope information from four months of age (Attaheri et al., 2022, 2024). Hence, the delta and theta sampling networks shown in Figure 35.2 are gaining information about AMs at the 'stress' and 'syllable' rates from early in language learning. Computational speech analyses of IDS collected in the SEEDS project showed that delta–theta phase alignment of AMs is foregrounded in BabyTalk (Leong et al., 2017), and although phase-phase analyses have yet to be completed, the UK BabyRhythm neural data showed that phase-amplitude mechanisms are already in play during speech processing (Attaheri et al., 2022). Hence, phase-phase mechanisms may also be online. The effects of individual differences in these sensory/neural variables on the developing lexicon were documented by both projects. The SEEDS data showed that infant rise time sensitivity at 10 months predicted vocabulary development at three years, with behavioural rhythm measures at four years subsequently predicting pre-literacy variables, mediated by phonological sensitivity (non-word repetition). The UK BabyRhythm data suggested that the accuracy of delta-band cortical tracking of speech envelope information at 11 months predicted later vocabulary, as did theta–gamma PAC (Attaheri et al., 2024), while individual differences in neural phase angle to rhythmic speech and non-speech stimuli also predicted both later vocabulary and non-word repetition (Ní Choisdealbha et al., 2023). This suggests that both neural phase alignment in the delta band and the accuracy of speech envelope encoding are critical for later language outcomes (see also Rios-Lopez et al., 2022). Accordingly, data from both infant projects converge in showing that the sensory/neural factors foregrounded by TS theory make important contributions to later language outcomes for English-learning infants.

Box 35.1 Chapter Overview

Summary

This chapter presents a theoretical overview of how rhythm may be important for language development, using the framework of TS theory. Sensory and neural data for TS theory from children are reviewed, and the TS-proposed causal sensory/neural mechanisms are assessed by utilising recent infant longitudinal data.

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

Box 35.1 (cont.)

Implications

Amplitude 'rise time' discrimination and neural entrainment to rhythmic acoustic signals are physical characteristics of the nervous system and not under conscious control, yet they govern in part the efficiency of language acquisition. Future projects in other languages are now required to test the language universality of TS theory and these sensory/neural factors.

Gains

Deeper understanding of the sensory and neural mechanisms that govern individual differences in language acquisition may open the door to novel interventions, possibly based on rhythm, that may enhance language development.

REFERENCES

Alcock, K., Meints, K., and Rowland, C. (2020). *The UK communicative development inventories: Words and gestures*. J&R Press.

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

Attaheri, A., Ní Choisdealbha, A., 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

Buzsáki, G. (2006). *Rhythms of the brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195301069.001.0001

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926–1929. https://doi.org/10.1126/science.1099745

Corriveau, K. H., Goswami, U., and Thomson, J. M. (2010). Auditory processing and early literacy skills in a preschool and kindergarten population. *Journal of Learning Disabilities*, 43, 369–382. https://doi.org/10.1177/0022219410369071

Crosse, M., di Liberto, G., Bednar, A., and Lalor, E. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10, 00604. https://doi.org/10.3389/fnhum.2016.00604

Cumming, R., Wilson, A., and Goswami, U. (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: A new look at a perceptual hypothesis. *Frontiers in Psychology*, 6, 972. https://doi.org/10.3389/fpsyg.2015.00972

Daikoku, T., and Goswami, U. (2022). Hierarchical amplitude modulation structures and rhythm patterns: Comparing Western musical genres, song, and nature sounds to Babytalk. *PLoS One*, 17(10), e0275631. https://doi.org/10.1371/journal.pone.0275631

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

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

Destoky, F., Bertels, J., Niesen, M., et al. (2022). The role of reading experience in atypical cortical tracking of speech and speech-in-noise in dyslexia. *NeuroImage*, 253, 119061. https://doi.org/10.1016/j.neuroimage.2022.119061

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, 2457–2465.

Di Liberto, G. M., Attaheri, A., Cantisani, G., et al. (2023). Emergence of the cortical encoding of phonetic features in the first year of life. *Nature Communications*, 14, 7789. https://doi.org/10.1038/s41467-023-43490-x

Di Liberto, G. M., Peter, V., Kalashnikova, M., et al. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, 175, 70–79. https://doi.org/10.1016/j.neuroimage.2018.03.072

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

Drullman, R. (2006). The significance of temporal modulation frequencies for speech intelligibility. In S. Greenberg and W. Ainsworth (Eds.), *Listening to speech: An auditory perspective* (pp. 39–48). Lawrence Erlbaum Associates.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). Academic Press.

Friend, M., and Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers*, 35, 302–309. https://doi.org/10.3758/BF03202556

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

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517. http://dx.doi.org/10.1038/nn.3063

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, 16, 43–54. http://dx.doi.org/10.1038/nrn3836

Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1453(1), 67–78. https://doi.org/10.1111/nyas.14137

Goswami, U. (2020). Toward realizing the promise of educational neuroscience: Improving experimental design in developmental cognitive neuroscience studies. *Annual Review of Developmental Psychology*, 2(1), 133–155. https://doi.org/10.1146/annurev-devpsych-042320-100040

Goswami, U. (2022a). Language acquisition and speech rhythm patterns: An auditory neuroscience perspective. *Royal Society Open Science*, 9, 211855. https://doi.org/10.1098/rsos.211855

Goswami, U. (2022b). Theories of dyslexia. In M. Skeide (Ed.), *The Cambridge handbook of dyslexia and dyscalculia* (Cambridge Handbooks in Psychology) (pp. 5–24). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108973595.002

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

Goswami, U., Mead, N., Fosker, T., et al. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, 69(1), 1–17. https://doi.org/10.1016/j.jml.2013.03.001

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

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

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, 47, 674–689. https://doi.org/10.1016/j.cortex.2010.07.010

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

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude modulated sounds. *Physiological Reviews*, 84, 541–577. https://doi.org/10.1152/physrev.00029.2003

Kalashnikova, M., Burnham, D., Goswami, U. (2020a). Novel word learning deficits in infants at family risk for dyslexia. *Dyslexia*, 26(1), 3–17. https://doi.org/10.1002/dys.1649

Kalashnikova, M., Burnham, D., and Goswami, U. (2020b). The role of paired associate learning in acquiring letter–sound correspondences: A longitudinal study of children at family risk for dyslexia. *Scientific Studies of Reading*, 25(6), 504–518. https://doi.org/10.1080/10888438.2020.1850732

Kalashnikova, M., Burnham, D., and Goswami, U. (2020c). Infant-directed speech to infants at risk for dyslexia: A novel cross-dyad design. *Infancy*, 25(3), 286–303. https://doi.org/10.1111/infa.12329

Kalashnikova, M., Burnham, D., and Goswami, U. (2021). Rhythm discrimination and metronome tapping in 4-year-old children at risk for developmental dyslexia. *Cognitive Development*, 60, 101129. https://doi.org/10.1016/j.cogdev.2021.101129

Kalashnikova, M., Goswami, U., and Burnham, D. (2018). Mothers speak differently to infants at risk for dyslexia. *Developmental Science*, 21(1), e12487. https://doi.org/10.1111/desc.12487

Kalashnikova, M., Goswami, U., and Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at three years: A significant relationship. *Developmental Science*, 22(6), e12836. https://doi.org/10.1111/desc.12836

Keshavarzi, M., Mandke, K., Macfarlane, A., et al. (2022). Decoding of speech information using EEG in children with dyslexia: Less accurate low-frequency representations of speech, not 'Noisy' representations. *Brain and Language*, 235, 105198. https://doi.org/10.1016/j.bandl.2022.105198.

Koster, C., Been, P. H., Krikhaar, E. M., et al. (2005). Differences at 17 months: Productive language patterns in infants at familial risk for dyslexia and typically developing infants. *Journal of Speech, Language, and Hearing Research*, 48, 426–438. https://doi.org/10.1044/1092-4388(2005/029)

Leong, V., and Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLoS One*, 10(12), e0144411. 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

Leong, V., Stone, M. A., Turner, R. E., and Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America*, 136, 366–381. https://doi.org/10.1121/1.4883366

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

Mehler, J., Jusczyk, P., Lambertz, G., et al. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

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

Ní Choisdealbha, A., Attaheri, A., Rocha, S., et al. (2022). Neural detection of changes in amplitude rise time in infancy. *Developmental Cognitive Neuroscience*, 54, 101075. https://doi.org/10.1016/j.dcn.2022.101075

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press

Ní Choisdealbha, A., Attaheri, A., Rocha, S., et al. (2023). Neural phase angle from 2 months when tracking speech and non-speech rhythm linked to language performance from 12 to 24 months. *Brain & Language*, 243, 105301. https://doi.org/10.1016/j.bandl.2023.105301

Ní Choisdealbha, A., 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(4), e13502. https://doi.org/10.1111/desc.13502

Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: Don't be enslaved by the envelope. *Frontiers in Human Neuroscience*, 6, 250. https://doi.org/10.3389/fnhum.2012.00250

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

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 363, 1071–1086. https://doi.org/10.1098/rstb.2007.2160

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

Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, 7(777), 1–19. https://doi.org/10.3389/fnhum.2013.00777

Power, A. J., Colling, L. J., Mead, N., Barnes, L., and Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, 160, 1–10. https://doi.org/10.1016/j.bandl.2016.06.006

Rocha, S., Ní Choisdealbha, A., Attaheri, A., et al. (2024). Language acquisition in the longitudinal Cambridge UK BabyRhythm cohort. *Collabra: Psychology*, 10(1), 92998. https://doi.org/10.1525/collabra.92998

Rios-Lopez, P., Molinaro, N., Bourguignon, M., and Lallier, M. (2022). Right-hemisphere coherence to speech at pre-reading stages predicts reading performance one year later. *Journal of Cognitive Psychology*, 34(2), 179–193. https://doi.org/10.1080/20445911.2021.1986514

Roid, G. H. (2003). *Stanford-binet intelligence scales: Fifth edition*. Riverside Publishing.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81, 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Thomson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, 102, 120–129. https://doi.org/10.1016/j.jphysparis.2008.03.007

Thomson, J. M., and Goswami, U. (2010). Learning novel phonological representations in developmental dyslexia: Associations with basic auditory processing of rise time and phonological awareness. *Reading & Writing*, 23, 453–469. https://doi.org/10.1007/s11145-009-9167-9

Trehub, S. E., and Trainor, L. (1998). Singing to infants: Lullabies and play songs. In C. Rovee-Collier, L. Lipsitt, and H. Hayne (Eds.), *Advances in infancy research* (Vol. 12) (pp. 43–77). Ablex.

Turner, R. E. (2010). Statistical models for natural sounds. Doctoral dissertation, University College London. Gatsby Computational Neuroscience Unit. www.gatsby.ucl.ac.uk/turner/Publications/turner-2010.html

van Viersen, S., de Bree, E. H., Verdam, M., et al. (2017). Delayed early vocabulary development in children at family risk of dyslexia. *Journal of Speech, Language, and Hearing Research*, 60(4), 937–949. https://doi.org/10.1044/2016_JSLHR-L-16-0031

https://doi.org/10.1017/9781009295888.042 Published online by Cambridge University Press