Speech Rhythm and Language Acquisition:

An Amplitude Modulation Phase Hierarchy Perspective

Usha Goswami

Centre for Neuroscience in Education

Department of Psychology

University of Cambridge

Short Title: Speech Rhythm and Language Acquisition

Corresponding Author:

Professor Usha Goswami

Department of Psychology

Downing St., Cambridge CB2 3EB, U.K.

Tel. 44 1223 767635 Fax. 44 1223 333564

Email: ucg10@cam.ac.uk

Abstract

Language lies at the heart of our experience as humans and disorders of language acquisition carry severe developmental costs. Rhythmic processing lies at the heart of language acquisition. Here I review our understanding of the perceptual and neural mechanisms that support language acquisition, from a novel amplitude modulation perspective. Amplitude modulation patterns in infant-directed and child-directed speech support the perceptual experience of rhythm, and the brain encodes these rhythm patterns in part via neuroelectric oscillations. When brain rhythms align themselves with (entrain to) acoustic rhythms, speech intelligibility improves. Recent advances in the auditory neuroscience of speech processing enable studies of neuronal oscillatory entrainment in children and infants. The 'amplitude modulation phase hierarchy' theoretical perspective on language acquisition is applicable across languages, and cross-language investigations adopting this novel perspective would be valuable for the field.

Traditionally, language acquisition seemed such a remarkable feat for the infant brain to accomplish that it was assumed that special capacities must be at work. These capacities were thought to be distinct from the capacities underpinning broader cognitive development. For example, Chomsky (1957, see Ref. 4) proposed a 'language acquisition device' in the brain, a device with innate knowledge of the general rules that all languages obey, and innate knowledge of permitted rule variations. It was believed that the 'input' for language learning, namely the speech directed towards infants and the language that they overheard, was too 'impoverished' to support the rapid learning that characterises language acquisition. More recently, it has been recognised that the 'input' to infants and young children is actually very rich. Regarding the central role of speech rhythm, it has become clear that auditory perceptual information is replete with statistical patterns, and that statistical learning of these patterns by infants is automatic. Language contains acoustic statistical cues to phonotactics (conditional probabilities concerning the sounds that make up the language and the orders in which they can be combined), acoustic statistical cues to word boundaries and syntactic phrasing (largely carried by speech rhythm and duration cues), and acoustic cues to the emotional content of speech (largely carried by prosodic stress patterns and loudness cues). Regarding these latter rhythm-based cues, infant-directed speech (IDS) has special acoustic modifications to increase perceptual salience. Here I illustrate important developmental links between neural speech encoding, amplitude envelope rise times (for both the neural encoding of speech and for rhythm perception), and the acoustic statistics reflected by patterns of amplitude modulation in infant- and child-directed speech. I will outline a new set of acoustic statistics, amplitude modulation phase hierarchies, that has been uncovered by studying the rhythmic routines of the nursery – English 'nursery rhymes'. Although primarily relevant to the automatic extraction of phonological information, I will review data suggesting that these novel acoustic statistics also contain sensory information relevant to learning inflectional

morphology, grammar and syntax. Accordingly, investigation of the role of amplitude modulation phase hierarchies across languages and learners may reveal new insights into the role of speech rhythm in language acquisition, evolution and diachronic development.

Speech Rhythm as a Precursor to Language Acquisition

It has been known for a long time that newborn infants are sensitive to prosodic and rhythmic patterning in language. Mehler and his colleagues showed that French-exposed infants as young as 4 days could use information about linguistic rhythm and stress to distinguish their native language from other languages (Mehler et al., 1988).[29] In their experiment, recordings were made of a bilingual speaker of French and Russian telling the same story in French and in Russian. Fifteen second segments of these stories were then played to the babies. The babies showed a clear preference for listening to the native language, in this case, French. The researchers then played the tapes backwards. This meant that whereas the absolute parameters of the signal such as voice pitch were preserved, the relative cues such as intonation and melody were modified. With the reversed speech, the babies could no longer tell the difference between French and Russian. Mehler and his colleagues argued that the infants were relying on rhythmic and prosodic cues to distinguish the two languages. In a control experiment in which the speech was filtered so that only the rhythmic cues were preserved (filtered speech sounds a bit like someone speaking under water), the 4-day old infants could again distinguish between Russian and French. Despite this early work, subsequent experimental analyses of infant-directed speech (IDS), the special intonational register used across cultures when speaking to infants, focused on acoustic variables other than rhythm, such as fundamental frequency (Fernald & Kuhl, 1987).[9] Nevertheless, longitudinal studies showed that infants who heard more IDS developed larger vocabularies and also babbled more themselves (e.g., Ramirez-Esparza, Garcia-Sierra and Kuhl, 2014).[38]

In a similar vein, studies exploring infant acoustic statistical learning tended to remove rhythmic and prosodic information from the input, and focus just on conditional probabilities between syllables or phonemes (e.g., Saffran, Aslin & Newport, 1996).[40] In a seminal study, Saffran et al. (1996, see Ref. 40) gave 8-month-old infants novel 'words' to learn on the basis of transitional probabilities. The 'words' were novel 3-syllable units like "bidaku" and "padoti", repeated in random order in a monotonous stream (no stress cues or pauses) by a computer speech synthesiser that sounded like a female voice, for 2 minutes. The transitional probabilities were 1.0 for syllable pairs within words (like bi-da), and 0.33 for syllable pairs that crossed word boundaries (like ku-pa). Accordingly, the transitional probabilities provided statistical information regarding word boundaries (i.e., bidaku padoti golabu bidaku..). Saffran et al. showed that after a 2-minute learning trial, infants could distinguish between familiar words from the learning phase (like "bidaku") and unfamiliar new sequences (like "dapiku"). Studies such as these have shown that powerful acoustic statistical learning capacities are present in infants. These processes are automatic, and are likely to reflect the way that the brain processes incoming sensory information. The brain is always seeking patterns in sensory input, computing various statistics to process incoming sensory information and predict what is coming next (e.g., using Bayesian statistics, Fiser et al., 2010).[10] However, in natural language, the conditional probabilities that support the identification of word boundaries are supplemented by rhythm and stress cues. Further, for language learning, acoustic statistical learning can only be as good as the quality of the sensory input upon which it is based. This point will be revisited below, when we consider the impaired phonological development of children with developmental dyslexia.

Speech Rhythm and Neural Speech Encoding

The mechanisms used by the brain to encode the speech signal are in part also rhythmic, as speech encoding is intimately related to neuroelectric oscillations. In simple

terms, brain cells generate electrical pulses, alternating between generating electrical potentials and recovering from generating potentials. For large networks of cells, this alternation of the network from excitation to inhibition produces an oscillation in electrical rhythm – a brain rhythm or 'brain wave'. The science of electrophysiology (EEG – the electroencephalogram) records these oscillations by placing sensitive electrodes on the scalp, and has documented a range of endogenous oscillatory rates across the cortex, ranging from *delta* (~1 – 3 Hz) through *theta* (~4 – 8 Hz), *alpha* (~ 8 – 12 Hz), *beta* (~15 – 30 Hz), *low gamma* (~30 – 50 Hz) and *high gamma* (~60 Hz +). More recently, cognitive neuroscience is documenting relations between oscillatory phase and human perceptual experience. For example, when detecting visual targets, adults are unaware of visual stimuli that arrive during the trough of an alpha oscillation (the least excitable phase, when fewest neurons are firing) in parietal cortex. Adults are most likely to detect visual targets at the oscillatory *peak* of the ongoing brain rhythm, when the maximum number of cells in the network are discharging electrical potentials (Mathewson et al., 2009).[28] For adults, visual events that arrive 'out of phase' (during the oscillatory trough) do not reach conscious awareness. Regarding speech encoding, studies with adults reveal key roles for neuroelectric oscillations in auditory cortex at four temporal rates, delta, theta, beta and low gamma (see Giraud & Poeppel, 2012,[12] and Poeppel, 2014,[32] for recent summaries). During speech encoding, these different brain electrical rhythms re-calibrate their activity to be in time or in phase with rhythmic energy patterns in the speech signal (chiefly governed by amplitude modulations, as explained below). The accuracy of this neural alignment between brain waves and sound waves is called neural 'entrainment', and aids speech intelligibility.

For example, in the adult brain theta phase patterns reliably discriminate between different sentences (Luo & Poeppel, 2007).[27] Using magnetoencephalography (MEG), Luo and Poeppel were able to show that the phase pattern of the theta oscillation in auditory

cortex (oscillatory cycles of ~200 ms) tracked and discriminated different spoken sentences. In other words, the brain recordings could be used post-hoc to classify the individual sentences heard by the participants. When the sentences were acoustically degraded so that intelligibility declined, the theta phase patterns were not able to classify the sentences as reliably. Accordingly, Luo and Poeppel's (2007, see Ref. 27) data showed that the relative *timing* of oscillatory neural responses (the phase of the responses) was a critical factor related to speech intelligibility. More recently, neuroimaging work by Gross et al. (2013, see Ref. 20) showed that when adults listen to connected speech, cortical oscillations in the delta, theta and gamma bands operate together in an information hierarchy. The slower oscillations govern activity in the faster oscillatory bands, with delta phase controlling theta phase, and theta phase controlling gamma power. Oscillations in the delta band in auditory cortex have also been shown to govern amplitude in the beta band in motor cortex (Arnal et al., 2014).[1] Although linked by Arnal et al. to the motor prediction of speech, this auditory-motor oscillatory link may also be important when deliberately speaking to a rhythm, due to the role of amplitude envelope rise times in speech rhythm, as outlined further below. Meanwhile, quasi-rhythmicity is found in many naturally-occuring sensory inputs, which may support phase entrainment of related neuroelectric oscillations during encoding (e.g., *vision*: gait, wind-blown leaves; *audition*: birdsong, rain). For environmental sounds like rain and wind, Turner (2010, see Ref. 45) showed that Probabilistic Amplitude Demodulation (PAD) provided an effective Bayesian learning approach. Natural sounds are characterised by amplitude (local sound intensity) modulation patterns correlated over long time scales and across multiple frequency bands, such as delta, theta and gamma. Turner (2010, see Ref. 45) showed that these natural sounds can be modelled mathematically by amplitude modulation cascades, or amplitude modulation phase hierarchies (Leong et al., 2014).[22] To explain this important concept in more detail, I next consider the amplitude envelope of speech.

Amplitude Modulation, the Amplitude Envelope of Speech and Rise Times

Amplitude modulation has received little attention in both the developmental language literature and in the linguistic literature related to speech rhythm. In simple terms, amplitude modulation is alternations in sound intensity, or variations in loudness in the speech signal. When someone opens their mouth to speak, they are causing air molecules to move – they are producing a 'sound wave'. The sound wave of speech can be thought of as energy moving through the air. When there is a lot of energy (= loud sound), the signal has greater amplitude than when there is relatively low energy (= soft sound). However, as the speaker opens and closes the vocal tract, there are additional naturally-varying changes or *modulations* in the energy in the overall signal, however loudly or softly overall that particular signal is being produced. These modulations are caused by simultaneous movements of the vocal folds, the tongue and the vocal tract, mouth and lips. These modulations in loudness ('amplitude modulations') are broadly experienced as speech rhythm. These different patterns of amplitude modulation can be thought of as embedded in an overall 'envelope' of sound, the 'amplitude envelope'. The amplitude envelope is the power-weighted average across frequency bands of the modulations that reach the ear and then excite different parts of the cochlea, the nerve endings in the inner ear. It is the changes or modulations in this overall amplitude envelope that is primarily experienced by the listener as speech rhythm (Greenberg, 2006).[19]

Figure 1 about here

The energy profile of the amplitude envelope in speech varies relatively slowly in time, and most of the amplitude fluctuations in the envelope reflect the rising and falling 'arcs' of energy that coincide with syllable production (Greenberg, 2006).[19] As each syllable is produced by a speaker, peak energy is reached as the vowel is produced, and then falls again (see Figure 1). The rising phase of each energy 'arc' corresponds to amplitude 'rise

time', the time taken to reach maximal energy or peak amplitude in a modulation. The energy arc of the syllable peaks at the syllable nucleus, with the production of the vowel. It has been known for a long time that when deliberately speaking to a rhythm, for example alternately saying "sweet" and "seat" to a metronome beat, the speaker times the rise times of the vowels (Scott, 1998).[42] This literature has also been described in terms of the 'perceptual centres' or P-centres of sounds (Morton et al., 1976,[30] and Gordon, 1987).[13] The P-centres of syllables are determined by the rise times of the vowels in the syllables. If counting rhythmically, the speaker times the rise time of the vowel in the stressed syllable of a bisyllabic number word like "seven" with the rise time of the vowel in the single syllable number words ("…five … six … SE-ven"). Another way of conceptualising this is via metrical poetry or via children's nursery rhymes. In the English nursery rhyme "Pussycat pussycat where have you been?", the stressed syllable "PU" in "pussy" is the basis of the rhythmic timing, with the two following syllables compressed into a single beat in order to maintain the rhythm. Indeed, English nursery rhymes span a range of poetic meters, with rhymes like "Cobbler, cobbler mend my shoe" following a trochaic rhythm pattern based on alternating stressed and unstressed syllables, while a rhyme like "As I was going to St Ives" follows an iambic rhythm pattern, based on unstressed and stressed syllables alternating. Stressed syllables also have larger rise times, with a greater increase in amplitude from the onset of the syllable to the peak of the syllable nucleus. This energetic cue is acoustically very salient, and stressed syllables play a core role in prosodic structure (linguistic rhythmic structure) in many of the world's languages.

However, amplitude rise times are also important for another reason. As they correspond to rates of change in energy, they also provide an acoustic guide to different modulation rates. Amplitude rise times turn out to be an important trigger for neural phase re-setting during speech encoding. Amplitude rise times act as acoustic 'edges' that are utilised

by the nervous system to synchronise the timing of ongoing endogenous oscillations in auditory cortex with the timing of amplitude modulation patterns in the speech signal (Gross et al., 2013) [20]. By providing temporal cues to the different rates of amplitude modulation contained in the overall speech envelope, amplitude rise times enable accurate alignment between ongoing neural oscillatory rhythms at different temporal rates and speech rhythm patterns. For example, if the rise times associated with syllable-level modulations are removed from speech, adult listeners can no longer comprehend what is being said (Doelling et al. 2014).[8] If simple clicks are then inserted into the signal in place of these rise times, the speech becomes intelligible again. These and other neural data suggest that amplitude rise times are important mechanistically for the encoding of continuous speech. Perhaps unsurprisingly, rise time discrimination is impaired in two of the language-learning disorders of childhood, developmental dyslexia and oral developmental language disorder (Goswami et al., 2002,[14] and Corriveau et al., 2007).[5] These childhood disorders also present with speech rhythm impairments, as will be discussed below.

The Amplitude Modulation Phase Hierarchy and Speech Rhythm

Nursery rhymes and rhyming games are a ubiquitous part of an English childhood (Opie & Opie, 1987)[31], and rhythmic speech registers are found in the nurseries of many other languages. Accordingly, modelling the amplitude modulation structure of child-directed rhythmic speech may be enlightening regarding the acoustic basis of speech rhythm. Leong et al. (2014, see Ref. 22) capitalised on the modelling work by Turner (2010, see Ref. 45) regarding the amplitude modulation structure of environmental sounds like rain and wind. Leong et al. (2014, see Ref. 22) applied Turner's Probabilistic Amplitude Demodulation modelling approach to the speech signal (see also Leong, 2012).[21] A natural sound like rain is not unstructured, rather it has rhythmic patterning related to loudness patterns that are correlated over long time scales and across multiple different temporal rates. The loudness

patterns are cascade-like because there are statistical dependencies between different temporal rates. In effect, rain or wind contains *nested hierarchies* of amplitude modulations, with slower rates governing faster rates. Given the special rhythmic speech registers used across cultures with young children, rhythmic speech may contain particularly strong cascade-like patterning of amplitude modulations. Leong and Goswami (2015, see Ref. 23) used principle components analysis to uncover these cascade-like patterns in child-directed speech.

Leong and Goswami (2015, see Ref. 23) began by treating the speech signal in the same way as it is filtered by the cochlea in the human ear. They then looked for statistical patterns in the outputs of these filters. The basis for the modelling was 44 English nursery rhymes spoken by six different early years teachers at a natural pace. The modelling showed that the speech energy in the nursery rhymes was clustered into three bands of different temporal rates, a very slow band (amplitude modulations centred on 2 Hz, matching the oscillatory delta band), a slow band (amplitude modulations centred on 5 Hz, matching the oscillatory theta band), and a faster band (amplitude modulations centred on 20 Hz, matching the oscillatory beta band). This last band was fairly wide (12 – 40 Hz). These temporal rates were nested inside each other, so that the timing of the amplitude modulations at ~2 Hz determined the timing of the amplitude modulations at ~5 Hz, and the timing of the modulations at ~5 Hz in turn determined the timing of the faster modulations at ~20 Hz. The ~2 Hz band of amplitude modulations that (by hypothesis) would entrain delta band oscillations were the largest and most dominant in the speech signal, peaking with the nucleus of each stressed syllable. There were also smaller peaks in amplitude modulation correlated with the occurrence of each syllable, whether stressed or unstressed. When a peak in modulation energy in the band of amplitude modulations centred on ~2 Hz (approximately the *stressed syllable* rate) coincided with a peak in modulation energy in the band of

amplitude modulations centred on ~5 Hz (approximately the *syllable* rate), a strong syllable was heard. When a peak in modulation energy in the band centred on ~5 Hz coincided with a trough in modulation energy in the band centred on ~2 Hz, a weaker syllable was heard. This means that a unique acoustic statistic, the *phase alignment* (rhythmic synchronicity) of the very slow and slow rates of amplitude modulation in speech yielded the rhythmic patterning of a nursery rhyme phrase like "Cobbler, cobbler, mend my shoe" or "Jack and Jill went up the hill". The output of the modelling showed that the rhythmic changes carried by low frequency portions of the speech signal (portions lower in pitch, < ~700 Hz) provided particularly salient acoustic clues to rhythmic structure (see red colours in Figure 2). Interestingly, these lower-frequency portions of the speech signal were produced with the largest degree of temporal similarity by the different speakers participating in the study. This suggests that for listening children or infants, these acoustic statistics will be relatively consistent across speakers.

Figure 2 about here

Experiments using tone-vocoded nursery rhymes with adults (Leong et al., 2014)[22] were instrumental in demonstrating the importance of the phase alignment of the slower bands of amplitude modulation for linguistic rhythm perception. Tone-vocoding a speech signal removes the temporal fine structure from the original and then applies the remaining amplitude modulations to a sine tone carrier. The resulting acoustic patterns have clear rhythmic temporal patterning, for example sounding like morse code or flutter, but are unintelligible. Leong et al. (2014, see Ref. 22) reported that adults could reliably identify English nursery rhymes solely on the basis of the phase relations between the amplitude modulations in the two slower bands, corresponding to delta and theta oscillatory bands (the temporal rates in the speech were ~2 Hz and ~4 Hz for these tone-vocoded stimuli). Accordingly, for rhythmic child-directed speech, the slowest band of amplitude modulations,

centred on ~2 Hz, provides information relevant to identifying stressed syllables. The next slowest band of amplitude modulations, centred on ~5 Hz in natural speech, provides information relevant to parsing syllables. Together, these two bands of amplitude modulation can be used to identify the rhythm pattern of a phrase, namely its prosodic structure (for example whether it follows a trochaic rhythm or an iambic rhythm). These two bands of amplitude modulation also identify the prosodic structure of multi-syllabic words. The modelling showed that the faster band of amplitude modulations, centred on ~20 Hz (by hypothesis relevant to beta band oscillatory entrainment) provided information relevant to identifying the onset-rime division of the syllable. The onset-rime division of any syllable is given by dividing at the vowel, as in S-EAT, SW-EET, STR-EET. For the nursery rhymes in the modelling, most syllables were simple consonant-vowel (CV) syllables, with a single phoneme onset and a single phoneme rime. Temporally, the beta band of amplitude modulations tended to pick out the syllable onsets. For the syllables in these nursery rhymes, therefore, this band of AM also helped to identify the constituent *phonemes*. Accordingly, all three rates of amplitude modulation and their phase relations identified by the modelling turned out to hold important statistical clues to phonological structure. This phonological structure can be conceptualised via the *linguistic hierarchy*.

Figure 3 about here

Linguistically, we can think of these different levels of phonology as forming a hierarchy, with larger units (the stressed syllables) at the top of the hierarchy (Liberman & Prince, 1977).[26] At the next level of the hierarchy come the syllables, followed by the onset-rime units. For many world languages, where a simple CV syllable structure is predominant, onset-rime units also represent the constituent phonemes in a syllable. For languages with complex phonology, such as English, both onset and rime units can contain multiple phonemes. Accordingly, phonemes, the smallest units of sound in words, could represent the

end state of the hierarchy (they are typically learned via reading, see Ziegler & Goswami, 2005).[47] When Leong and Goswami (2015, see Ref. 23) applied their spectral-amplitude modulation phase hierarchy (S-AMPH) model to individual nursery rhymes, the model was able to detect the stressed syllables, syllables and onset-rime units quite successfully. For nursery rhymes spoken to a metronome beat, hence with perfect rhythmic timing, the model identified 95% of stressed syllables accurately, 98% of syllables accurately, and 91% of the onset-rime units accurately. For nursery rhymes spoken freely, the model identified 72% of stressed syllables accurately, 82% of syllables accurately, and 78% of the onset-rime units accurately. Accordingly, if encoded accurately by the brain, these AM statistics would yield the linguistic hierarchy over 90% of the time for rhythmically-produced speech. The amplitude envelope of rhythmic child-directed speech thus provides sufficient acoustic information for the automatic neural segmentation of phonological units of different grain sizes (stressed syllables, syllables, onset-rime units). The linguistic routines of the nursery hence provide optimal acoustic statistical input for children's phonological development.

Amplitude Modulation Phase Hierarchies in Infant-Directed Speech

Given the behavioural data showing that infants are sensitive to prosodic and rhythmic patterning in language from birth (Mehler et al., 1988),[29] it is interesting to model IDS from an amplitude modulation perspective. This was done by Leong, Kalashnikova, Burnham and Goswami (2017, see Ref. 25), utilising a corpus of IDS collected longitudinally. Mothers were recorded speaking to their infants in IDS when the infants were aged 7, 9 and 11 months of age. The same mothers were also recorded speaking adult-directed speech (ADS), when they spoke to the experimenters. The S-AMPH model was then applied to both the IDS as produced to these infants at different ages, and to the ADS. Interesting differences were revealed. The temporal modulation structure of IDS turned out to have highly predictable amplitude modulation patterning, similar to that characterising

English nursery rhymes. In IDS the amplitude modulation energy corresponding to oscillations in the delta band (centred on ~2 Hz) was significantly *greater* than in adult-directed speech (ADS). The degree of phase alignment between the amplitude modulations in the two slowest bands (centred on ~2 Hz and ~5 Hz) was also significantly *greater* in IDS compared to ADS. Meanwhile, ADS had significantly more amplitude modulation energy in the AM band centred on ~5 Hz than IDS (corresponding to oscillations in the theta band). ADS also showed significantly stronger phase alignment between the faster bands of amplitude modulation centred on ~5 Hz and ~20 Hz compared to IDS.

The comparison of IDS and ADS shows that acoustically, IDS provides an enhanced and predictable set of statistics (or rhythmic templates), presumably to help the infant brain with successful entrainment to the speech signal. IDS showed more modulation energy in the delta band (0.9 – 2.5 Hz, corresponding to delta oscillations) compared to ADS, reflecting the higher proportion of stressed syllables in IDS compared to ADS. The significantly stronger phase synchronisation between amplitude modulations in the bands centred on ~2 Hz and ~5 Hz respectively in IDS is indicative of greater rhythmic regularity (the temporal statistics facilitate the prediction of the next stressed syllable). These statistical patterns were still present when corrections were made for the different speaker rates that characterise IDS and ADS (IDS is typically slower). These acoustic discoveries suggest that IDS is optimally structured to facilitate neural entrainment to prosodic information by the infant brain, in other words to facilitate the perception of speech rhythm. The stronger phase alignment between the two slower bands of amplitude modulation enhances the novel acoustic statistic described earlier that underpins rhythm perception – the *phase alignment* (rhythmic synchronicity) of the very slow and slow rates of amplitude modulation in speech (by hypothesis, corresponding to oscillatory delta-theta phase alignment). Enhancing the phase alignment of these slowest bands of amplitude modulation would facilitate the extraction of rhythm

patterns (e.g., trochaic versus iambic) from speech, acoustic statistical patterns that are important for language learning.

How does the brain encode these acoustic statistics? The assumption, which has yet to be shown empirically for infants or children regarding the AM hierarchy, is that these patterns of amplitude modulation are encoded automatically by the brain, via oscillatory entrainment. Many pieces of relevant evidence are already in place. It is known that the adult brain uses amplitude rise times to discover the different temporal rates of incoming speech information, and then phase re-sets the activity of the appropriate cortical cell networks so as to align their oscillatory activity with this incoming information (neural entrainment or 'phase alignment' of cell networks, Giraud & Poeppel, 2012).[12] It is known that this process of phase alignment is happening simultaneously in different cell networks in auditory cortex, which are related in an information hierarchy (Gross et al., 2013).[20] There is also 'motor prediction' of speech, with neural coupling of delta-rate activity (phase) in auditory cortex with beta-rate activity (power) in motor cortex (Arnal et al., 2014).[1] Note that in the S-AMPH model, amplitude modulations in the beta band helped to identify the onset-rime division of the syllable, which could mean that delta-beta phase-amplitude coupling identifies the P-centres of the syllables. Delta-beta phase amplitude coupling could thus support rhythmic synchronisation of movement to P-centres in sounds, for example when dancing or tapping in time to the beat. In these multiple ways, the brain is already known to use timing information to form a rich representation of linguistic rhythm, encoding information from multiple modalities and at multiple time scales simultaneously. The multi-time information is then automatically bound together to yield the perception of speech as a seamless single input. There is no reason to think that the infant brain would be any different to the adult brain (Telkemeyer et al., 2011).[44] There is however evidence that children with developmental disorders of written language (developmental dyslexia) and spoken language (developmental

language disorder or DLD, formerly termed specific language impairment or SLI) have impairments both in discriminating amplitude envelope rise times and in perceiving speech rhythm. For developmental dyslexia, there is also evidence for impaired neural entrainment to delta band envelope information in speech.

Amplitude Rise Time Discrimination and Neural Entrainment in Developmental Language Disorders

During the past two decades, a series of studies of children with developmental dyslexia across languages has shown impaired discrimination of amplitude envelope rise times. These acoustic difficulties in rise time discrimination are typically related to impaired phonological development in these different languages, the core cognitive impairment exhibited by children with dyslexia. These phonological difficulties then cause reading difficulties (Goswami, 2015, for a review).[15] Acoustic sensitivity in children can be measured using psychoacoustic threshold estimation tasks. These are listening tasks that measure the 'just noticeable difference' (or *threshold* for noticing a difference) between two simple sounds. For example, one sound might have a higher pitch than another (frequency discrimination), or one sound might be longer than another (duration discrimination). Children with dyslexia consistently show impairments in psychoacoustic threshold estimation tasks measuring amplitude envelope rise times, and they seem to have particular problems with slower rise times (Goswami et al., 2002;[14] Richardson et al., 2004;[39] Stefanics et al., 2011).[43] This acoustic insensitivity to amplitude rise time has been shown in many languages – English, Chinese, Hungarian, Finnish, Dutch, Spanish and French (Goswami, 2015).[15] Furthermore, the rise time difficulties have been correlated with phonological difficulties in these different languages. For example, one study found that poor rise time discrimination was correlated with poor rhyme awareness in English, with poor tone awareness in Chinese, and with poor phoneme awareness in Spanish (Goswami et al., 2011).[17]

In fact, behavioural studies with English children show that individual differences in rise time perception are related to individual differences in a range of rhythmic and prosodic tasks. To measure sensitivity to prosodic structure in children with dyslexia, oral tasks are used, for example all syllables in an utterance are replaced by the single syllable "dee", either stressed ("DEE") or unstressed ("dee"). In this 'DeeDee' task, most of the phonetic information in an utterance is removed while the stress and rhythm patterns of the original words and phrases are retained. Goswami, Gerson, and Astruc (2010, see Ref. 16) created a DeeDee task for children with dyslexia using picture recognition based on celebrity names (e.g., *David Beckham*) and film and book titles (e.g., *Harry Potter*). The child was shown a picture whose name was 'spoken in DeeDees' (for example, "Harry Potter" would be "DEEdeeDEEdee"). Goswami et al. (2010, see Ref. 16) reported that the DeeDee task was performed significantly more poorly by 12-year-old children with developmental dyslexia compared to 12-year-old control children. In a second study, Goswami et al. (2013, see Ref. 18) found that 9-year-old children with dyslexia performed significantly more poorly in the DeeDee task than 7-year-old typically-developing control children – a 'reading-level match' experimental design. The reading-level match design is methodologically important for establishing potential causality, as it holds reading level constant between groups rather than chronological age. The reading-level match design thus approximately equates the children for reading experience, while also giving a mental age advantage to the children with dyslexia. The finding that the children with dyslexia were significantly less accurate than *younger* controls in the DeeDee task suggests that the rhythmic difficulty in dyslexia is a profound one. Individual differences in rise time discrimination were a significant predictor of performance in the DeeDee task even after age and nonverbal IQ were controlled in regression equations. In a series of studies with DLD children, these patterns have been broadly replicated (Richards & Goswami, 2015, 2019;[36-37] Cumming et al., 2015a, b).[6-7] The

children with DLD showed poorer rise time discrimination and poorer speech rhythm abilities, and in addition tended to show poorer duration discrimination as well (which would affect their ability to detect linguistic phrasal grouping, see Richards & Goswami, 2019 [37]).

In order to study neural entrainment to speech by children, a rhythmic speech paradigm was employed while EEG was recorded (Power et al., 2012).[33] Children listened to repetition of the syllable 'ba' presented at a 2 Hz rate ("ba…ba…ba…"). In order to study both auditory and visual entrainment to speech, the children either watched a video of a 'talking head' repeating the syllable (auditory-visual measure, AV), heard the auditory soundtrack without a visual stimulus (auditory measure, A), or watched the talking head producing syllables without hearing the speech (visual measure, V). Power et al. (2012, see Ref. 33) reported both auditory and visual oscillatory entrainment to rhythmic speech by the participating children (13 year olds), with significant entrainment in all conditions at the stimulation rate (2 Hz, delta) and also (for A and AV entrainment) at the theta rate (the 'syllable rate', ~ 5 Hz). In addition, *preferred phase* in the theta band was altered by predictive visual speech information. Preferred phase reflects the point in time during a temporal cycle when most neurons are firing. The information about speech that was provided visually (seeing the person's mouth opening, ready to produce a syllable) thus made the electrical discharge of the auditory cells more temporally accurate. This effect is well-documented in adults (e.g., Schroeder et al., 2008).[41] Visual rhythmic information automatically modulates auditory oscillations to the optimal phase for speech processing, providing an additional (cross-modal) cue for phase-resetting to that provided by amplitude rise times.

Children with dyslexia were then tested in the same rhythmic speech paradigm (Power et al., 2013).[34] In general the dyslexic children's brains showed very similar patterns. There was significant entrainment in all conditions, and visual speech information phase re-

set the auditory cell networks. However, one key difference was found when the auditory modality was involved, namely in the A and AV conditions. The children with dyslexia showed a significant difference in preferred phase in the delta band compared to control children, despite no group differences in power. These data suggest that most delta-rate neurons in the dyslexic brain were firing at a *non-optimal phase* for speech processing. Rather than peak neuro-electrical activity coinciding with the most informative parts of the speech signal, for children with dyslexia the neural response was slightly out of time.

A direct measure of the quality of the neural encoding of speech is offered by envelope reconstruction in EEG. In essence, children's electrical brain responses are reverse-engineered to re-create the amplitude modulation patterns in the sentences that they have been listening to. Power et al. (2016, see Ref. 35) used this technique with children with developmental dyslexia who were listening to semantically unpredictable sentences like "Arcs blew their cough" (in order to prevent successful guessing). The sentences had also been degraded by vocoding. The children had to repeat what they thought they heard. Power et al. (2016, see Ref. 35) reported that the speech envelopes in the 0 – 2 Hz (delta) band were encoded significantly less accurately by the brains of the children with dyslexia. This significant difference was found both in comparisons with age-matched control children and with reading level matched control children, who were 2 years younger in age, even though the children with dyslexia were as accurate in sentence repetition as the reading-level controls. The reading-level match comparison suggests a fundamental encoding deficit for very slow amplitude modulation information in speech for English-speaking children with dyslexia. Individual differences in envelope encoding were also significantly related to individual differences in a speech rhythm task (lexical stress perception).

To date, no studies that I am aware of have studied neural entrainment to speech for DLD children. However, given that children with DLD show impaired rise time

discrimination and impaired perception of speech rhythm, such studies would be very timely. It is possible, for example, that neural entrainment to envelope information would be impaired (as in developmental dyslexia) but in a different oscillatory band (for example, theta or beta/low gamma). Another plausible candidate for atypical entrainment in DLD is impaired delta-theta oscillatory phase alignment. Using the S-AMPH modelling approach, Flanagan and Goswami (2018, see Ref. 11) showed that inflectional morphology in English is cued by changes in the phase synchronisation of delta- and theta-band amplitude modulations. Flanagan and Goswami analysed the plural elicitation task (Berko, 1958),[2] a morphological awareness task in which children are asked to generate the plural forms of nonword items (as in "wug-wugs" and "lun-luns"). Flanagan and Goswami modelled the amplitude modulation structure of both the singular ("wug") and then plural ("wugs") forms of the items in Berko's task using the S-AMPH. The modelling showed that the only acoustic statistic that was systematically associated with the change in inflectional morphology was the magnitude of delta-theta phase synchronisation. As the primary impairment in DLD is grammatical rather than phonological, impaired neural phase synchronisation between delta and theta oscillations in auditory cortex may be a fruitful target for neuroimaging studies with DLD children.

Conclusions

Amplitude modulations in the speech signal and their phase relations play a core role in the perception of speech rhythm. Amplitude modulation phase relations are very salient in rhythmic speech used with children, such as the English nursery rhyme, and are enhanced in speech used with infants. The amplitude modulation structure of IDS is enhanced compared to ADS both in terms of delta band modulation energy and in terms of stronger phase synchronisation between the slower amplitude modulation bands, delta and theta. Children with developmental language disorders (dyslexia and DLD) show impaired perception of

speech rhythm and impaired discrimination of amplitude envelope rise times (a key acoustic cue to amplitude modulation rates). Given the core role of amplitude rise times in the automatic phase-resetting of the cortical oscillations that encode the speech signal, this auditory sensory impairment is likely to impair automatic linguistic learning from AM-based statistics. For example, it is likely to affect the automatic neural extraction of the amplitude modulation phase hierarchy, yielding atypical perceptual effects that have consequences for the acquisition of both phonology and morphology. As many languages use prosodic structure as the bedrock of their phonological systems, such atypical learning in dyslexia and DLD would be expected across languages. However, the perceptual consequences of atypical processing may differ depending on the phonological structures of different languages. For example, languages differ in their rhythmic timing, and this may affect the phase relations between different bands of amplitude modulations. Empirical work is needed. Nevertheless, education can always make a difference developmentally, and this is also true regarding impaired speech rhythm perception. Intervention studies in which children must explicitly match external beat structures to language rhythms, for example via tapping/drumming to poetry or coinciding a musical accompaniment with singing, can confer linguistic benefits (Bhide, Power & Goswami, 2013).[3] Growing understanding of the physiological mechanisms that underpin the neural oscillatory hierarchy may offer further targets for remediation. For example, it may be possible to adapt the speech signal in light of the discoveries about amplitude rise times and the amplitude modulation hierarchy discussed here, so that relevant cues such as amplitude rise times are synthetically amplified or exaggerated. Although technically complex, such adaptations are within the reach of current speech technology, and are already being tried with adults with dyslexia, with promising results (Van Hirtum, Moncada-Torres, Ghesquiere & Wouters, 2019, see Ref. 46). In the future, such adaptations

may prove transformative regarding developmental outcomes for children with dyslexia or DLD.

Acknowledgements.

Usha Goswami's current research is funded by the European Research Council Executive Agency and the Botnar Foundation.

References

1. Arnal, L.H., Doelling, K.B., & Poeppel, D. (2014). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex, 25 (9),* 3077- 3085.
2. Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150-177.
3. Bhide, A., Power, A.J., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind Brain & Education,* 7 (2), 113-123.
4. Chomsky, N. (1957). *Syntactic Structures.* The Hague/Paris : Mouton.
5. Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. *Journal of Speech, Language, and Hearing Research*. 50, 647-666.
6. Cumming, R., Wilson, A., & 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.
7. Cumming, R., Wilson, A., Leong, V., Colling, L.J. & Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments. *Frontiers in Human Neuroscience, 9*, 672.
8. Doelling, K.B., Arnal, L.H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85*, 761-68.
9. Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*, 279–293.

10. Fiser, J., Berkes, P., Orbán, G., Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences. 14 (3),* 119-130.
11. Flanagan, S.A. & Goswami, U. (2018). The role of phase synchronisation between low frequency amplitude modulations in child phonology and morphology speech tasks. *Journal of the Acoustical Society of America, 143 (3),* 1366 – 1375.
12. Giraud, A.L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15,* 511-517.
13. Gordon, J.W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82,* 88 – 105.
14. Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., et al. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 10911-10916.
15. Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience, 16,* 43-54.
16. Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing, 23*, 995-1019.
17. Goswami, U., Wang, H-L., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal sensory deficits in developmental dyslexia: English, Spanish and Chinese. *Journal of Cognitive Neuroscience, 23,* 325-337.
18. Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: a longitudinal study. *Journal of Memory and Language, 69 (1),* 1-17

19. Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In S. Greenberg and W. Ainsworth (Eds.), *Understanding Speech – An Auditory Perspective*, pp. 411-434. Mahwah, NJ: LEA.

20. Gross J., Hoogenboom N., Thut G., Schyns P., Panzeri S., Belin P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752.

21. Leong, V. (2012). "*Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models*". Doctoral dissertation, University of Cambridge. Available online at: http://www.cne.psychol.cam.ac.uk/pdfs/phds/vleong

22. Leong, V., Stone, M., Turner, R.E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America, 136,* 366-81.

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

24. Leong, V., & Goswami, U. (2017). Difficulties in auditory organization as a cause of reading backwardness? An auditory neuroscience perspective. *Developmental Science, 20,* e12457.

25. Leong, V., Kalashnikova, M., Burnham, D. & Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind, 1* (2), 78-90.

26. Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguist Inquirer, 8,* 249-336.

27. Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*, 1001-10.

28. Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To See or Not to See: Prestimulus α Phase Predicts Visual Awareness. *Journal of Neuroscience*. 29 (9) 2725-2732.

29. Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143–178.

30. Morton, J., Marcus, S.M., & Frankish, C. (1976). Perceptual centres (P-centres). *Psychological Review, 83,* 405-8.

31. Opie, I., & Opie, P. (1987). *The Lore and Language of Schoolchildren*. Oxford, UK: Oxford University Press.

32. Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology, 28c*, 142-149.

33. Power, A.J., Mead, N., Barnes, L., & Goswami, U. (2012). Neural entrainment to rhythmically-presented auditory, visual and audio-visual speech in children. *Frontiers in Psychology, 3*, 216.

34. Power, A.J., Mead, N., Barnes, L. & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience,* 7, 777.

35. Power, A.J., Colling, L.C., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain & Language, 160,* 1-10.

36. Richards, S. and Goswami, U. (2015). Auditory processing in Specific Language Impairment (SLI): Relations with the perception of lexical and phrasal stress. *Journal of Speech, Language & Hearing Research, 58,* 1292-1305.

37. Richards, S., & Goswami, U. (2019). Impaired recognition of metrical and syntactic boundaries in children with developmental language disorders. *Brain Sciences, 9,* 33.

38. Ramirez-Esparza, N., Garcia-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17 (6),* 880-891.

39. Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia, 10*, 215-233.

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

41. Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*, 106-113.

42. Scott, S.K., 1998. The point of P-centres. *Psychological Research*, *61*, 4-11.

43. Stefanics, G., Fosker, T., Huss, M., Mead, N., Szűcs, D., & Goswami, U. (2011). Auditory sensory deficits in developmental dyslexia, a longitudinal ERP study. *Neuroimage, 57* (3), 723-32.

44. Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally-modulated sounds in infants: Evidence from a combined NIRS and EEG study. *Frontiers in Psychology, 2*, 62.

45. Turner, R.E. (2010). *Statistical models for natural sounds*. Doctoral dissertation, University College London. http://www.gatsby.ucl.ac.uk/~turner/Publications/Thesis.pdf

46. Van Hirtum, T., Moncado-Torres, A., Ghesquiere, P., & Wouters, J. (2019). Speech envelope enhancement instantaneously effaces atypical speech perception in dyslexia. *Ear and Hearing*, in press.

47. Ziegler, J.C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Figure Legends.

1. Panel (a) shows the energy profile of the sentence fragment "..drive around, pick my children back up.." plotted as signal amplitude against time. The amplitude envelope is shown in red. The rise time cues associated with individual syllables are clearly visible; nevertheless not every rise in amplitude corresponds to a syllable onset. Panel (b) shows the syllable "my" in greater detail, with the rise time plotted. As shown, the rise time does not equate to a single moment in time, and the time taken for the envelope to reach its peak amplitude will vary syllable by syllable. *Figure reproduced with permission from Goswami (2018), Current Directions in Psychological Science, 27, 56-63.*

2. The top panel of the figure shows the energy profile of the raw speech signal for the nursery rhyme "Jack and Jill went up the hill" as a power-weighted average, with the amplitude envelope plotted in red. The middle panel unpacks the averaged envelope to show the amplitude envelopes for different spectral frequency bands, colour-coded from low frequencies (red colours) to high frequencies (blue colours). The energy fluctuations that correspond to the stressed syllables like "Jack" and "hill" are very salient, particularly in the lower spectral frequency bands in the speech signal. The bottom panel shows the same speech information plotted as a speech spectrogram. The salient amplitude changes in the lower frequencies are here depicted via increased shading, a choice which makes their possible salience for the brain less obvious. *Figure reproduced with permission from Goswami (2018), Current Directions in Psychological Science, 27, 56-63.*

3. **Schematic depiction of the linguistic hierarchy, the amplitude modulation (AM) hierarchy nested in children's nursery rhymes, and the oscillatory hierarchy**. The linguistic hierarchy depicted in the centre of the panel shows the phonological units of different grain sizes that are reliably recognised prior to literacy; note that for languages with

simple syllables (comprising single consonants and vowels), the onset and rime units will correspond to single phonemes. The frequencies of the electrophysiological oscillations measurable in the brain and thought to be relevant to perceiving these phonological units are depicted to the left-hand side of the figure (delta, 1 – 3 Hz; theta, 4 – 8 Hz, beta, 15 – 30 Hz). The centre frequencies of the amplitude modulations as extracted by the S-AMPH modelling are depicted to the right-hand side of the figure. The figure shows that the temporal rates for the AMs in speech and for the neuronal oscillations are approximately matched.