

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/254862759>

# Speech rhythm and temporal structure: Converging perspectives?

Article in *Laboratory Phonology* · March 2013

DOI: 10.1515/lp-2013-0004

CITATIONS

60

READS

1,450

2 authors:



Usha Goswami

University of Cambridge

260 PUBLICATIONS 16,110 CITATIONS

[SEE PROFILE](#)



Victoria Leong

University of Cambridge

53 PUBLICATIONS 588 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Infant EEG methods [View project](#)



BabyRhythm project [View project](#)

## Speech Rhythm and Temporal Structure: Converging Perspectives?

USHA GOSWAMI and VICTORIA LEONG

*Centre for Neuroscience in Education, University of Cambridge*

### 1. Introduction

In this paper, we present data from an overlapping research field to laboratory phonology, namely phonological processing in developmental dyslexia. We suggest that dyslexia data can offer a novel yet converging perspective on temporal structure and speech rhythm to data from laboratory phonology. The dyslexia data address the acoustic basis of rhythm perception and of rhythmic synchronisation, considering the potential role of the amplitude modulation (AM) structure of the speech envelope. In particular, we argue that a consideration of the perceptual effects of the *rise times* in the amplitude envelope and the *phase relations* between different temporal rates of AM may be theoretically productive (Leong 2012). Recent work in auditory neuroscience suggests that amplitude modulation is also important for the neural entrainment of *oscillatory neuronal networks*, which play a role in speech encoding (Giraud and Poeppel 2012). As neuronal oscillations entrain to amplitude modulation in the speech signal at different temporal rates, entrainment to the slower temporal modulations (focused on AM rates around 2 Hz and 5 Hz) may be critical for rhythmic perception and rhythmic synchronisation. Here we use *synchronisation* to refer to behavioural rhythmic alignment, for example, speaking in time with another speaker, or tapping in time with an external rhythm, and *entrainment* to refer to the alignment ('phase-locking') of neuronal oscillations with temporal regularities -in an external sensory input such as an auditory signal.

We also consider the conceptual convergence between the data from developmental dyslexia, the “foot oscillators” identified by Saltzman, Nam, Krivokapic, and Goldstein (2008; see also Krivokapic this volume), and the “supra-syllabic periodicities” clustering ~2 Hz identified by Tilsen and Arvaniti (submitted), cited in Arvaniti and Rodriquez (this volume). The papers by Arvaniti and Rodriquez (this volume) and Krivokapic (this volume) are considering quite different aspects of rhythmic speech, but they highlight important areas of theoretical convergence with respect to the wider field. First, speech rhythm cannot be captured successfully by models based on durational timing. This is clearly demonstrated by Arvaniti and Rodriquez’s careful analysis of the basis of rhythm discrimination in six different languages. Second, something systematically related to rhythm is nevertheless being perceived, as speakers of languages with different “rhythm types” converge unconsciously in their rhythmic production when speaking together. Indeed, Krivokapic suggests that when speakers synchronise with each other, it is variations in the production of prosodic “feet” that carry the effect. We finish by highlighting some key future questions for the field. These include the

contribution of articulatory mechanisms and oscillatory processes to rhythmic timing in speech production, the role of individual differences, and the role of development.

## **2. The perspective from developmental dyslexia: Some background**

### *2.1. Discrimination of amplitude modulation and rise time in dyslexia*

Developmental dyslexia is usually defined as a specific difficulty in reading and spelling that cannot be accounted for by low intelligence, poor educational opportunity, or obvious sensory/neurological damage. The accepted core cognitive deficit, across languages, involves specific problems with phonological representations and processing (e.g., Ziegler and Goswami 2005). This “phonological deficit” encompasses all levels of phonology, and is not simply a segmental phonological deficit – indeed, Ziegler and Goswami’s (2005) review of reading acquisition and dyslexia across languages showed that phoneme awareness emerges largely as a *consequence* of learning to read an alphabetic orthography. Further, learning to read “re-maps” phonology in the brain. Frith (1998) wrote that acquiring the alphabetic code was like catching a virus: “This virus infects all speech processing, as now whole word sounds are automatically broken up into sound constituents. Language is never the same again” (p. 1051).

Other research has shown that children with dyslexia who are learning to read transparent orthographies like German acquire age-appropriate phonemic awareness by around 10 years (e.g., Wimmer 1993). Nevertheless, they remain dyslexic, and their reading is slow and effortful, albeit accurate. Hence the “phonological deficit” in dyslexia is no longer conceptualised as a segmental one. In turn, this analysis suggests that any acoustic basis for the cross-language “phonological deficit” is unlikely to depend solely on quickly-varying acoustic cues to the identity of phonetic segments (such as formant transitions). Rather, the deficit may relate to acoustic problems at the syllabic and supra-syllabic level (Goswami et al. 2002). The acoustic cues specifying the temporal pattern of these larger speech units are primarily found in the slowly-varying amplitude envelope of the acoustic signal. During the past decade, we have been exploring the role of amplitude envelope perception in the “phonological deficit” in dyslexia, across languages (our studies encompass English, French, Spanish, Hungarian, Chinese and Finnish; e.g., Goswami et al. 2002; Richardson et al. 2004; Hämäläinen et al. 2009; Surányi et al. 2009; Goswami et al. 2011a).

Accurate perception of the amplitude envelope, which is the slow-varying energy profile of the acoustic waveform that the ear receives (amplitude variation over time; Houtgast and Steeneken 1985), is crucial for speech intelligibility. An example of the amplitude envelope for the sentence “...drive round, pick my children back up” is shown in Figure 1a. As described by Giraud and Poeppel (2012), rhythmic structure in the envelope is given by regular modulations of signal energy over time, which for speech peak at a rate of 3–5 Hz, the “syllable rate” (Greenberg et al. 2003). The onsets of successive syllable-related modulations in the amplitude envelope and their rates of change (rise

times) are critical linguistic perceptual events, as these rise times will vary with the phonetic properties of the syllable (e.g., plosive versus glide) and will be larger when a syllable is stressed. The rise time for the syllable “my” taken from the original sentence is shown in Figure 1b. As illustrated, syllable rise time does not equate to a single moment in time, such as the peak associated with the vowel. The time taken for the envelope to reach its highest amplitude will vary syllable by syllable, and it is the precision of the discrimination between these rate of change measures that appears to be impaired in developmental dyslexia. From Figure 1a, it may be noted that not all local rises in amplitude correspond to the onsets of whole syllables. For example, there is a smaller rise time associated with the affricate consonant /tʃ/ in “children”, located prior to the main onset of the syllable “chil”. These brief consonantal rise times are distinguishable from the larger and longer-lasting syllable vowel-centred rise times via their spectral and temporal characteristics (see Leong 2012). Therefore, it is the perceptual discrimination of different rise times occurring at different modulation rates and at different acoustic frequencies, rather than sensitivity to one particular rise time in the envelope, that is likely to be important for phonological development. Nevertheless, the modulation spectrum plotted in Figure 2 shows that across all acoustic frequency bands, the peak lies consistently between 3–5 Hz. Hence the most prominent rise times (energy changes) in the amplitude envelope will be those corresponding to syllables. By hypothesis, these prominent changes at low frequency rates within the speech envelope are perceived less well by individuals with dyslexia.

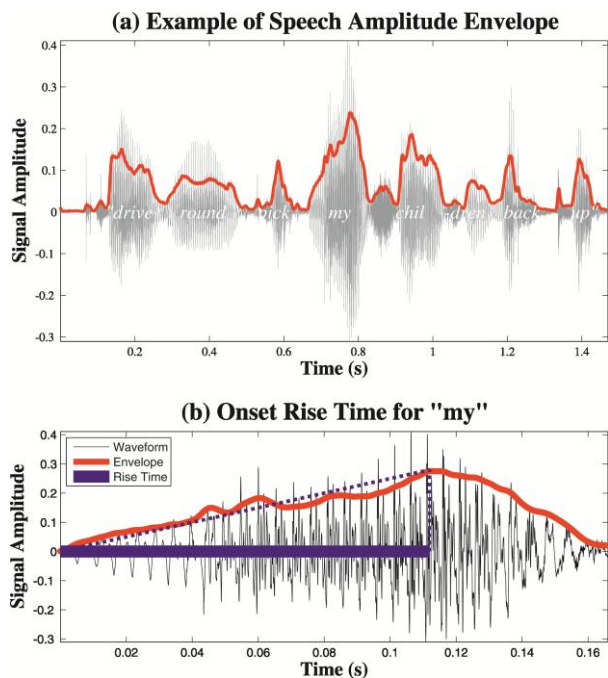


Figure 1. *The amplitude envelope of speech, syllable rise times and the modulation spectrum. (Left top, (a)) Example of a short excerpt of conversational speech where the amplitude envelope is shown in red, overlaid on the original waveform of the acoustic signal in grey. (Left bottom, (b)) Illustration of the onset rise time for the syllable "my", taken from the speech sample in (a).*

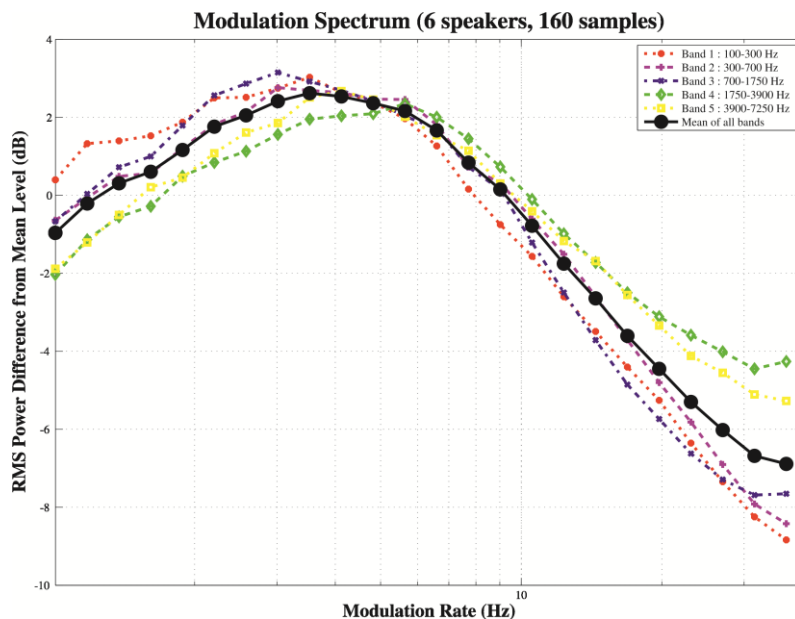


Figure 2. Averaged long-term modulation spectra of 160 conversational speech samples from 6 different speakers. Speech samples were between 24s-34s in length. The modulation spectra for 5 frequency bands are shown, as well as the average modulation spectrum across the 5 frequency bands. Notice that the highest power (peak) in the modulation spectrum consistently lies between 3-5 Hz for all speech frequency bands. This dominant amplitude modulation rate corresponds to the syllable rate of utterance. Therefore the most prominent energy changes (e.g. onset rises) in the amplitude envelope convey information about syllable pattern.

In studies using non-speech amplitude envelopes with different rise times, we have found that children with dyslexia are indeed impaired compared to their same-age peers in discriminating rise times in 6 languages (varying in rhythm type, see Goswami 2011 for a summary). Independent work in Dutch has found similar results (Poelmans et al. 2011), although a negative result was reported in Greek (Georgiou et al. 2010). Brain imaging (EEG) studies suggest that the brains of children with dyslexia respond differently to *slower* rise times (90ms but not 15 ms, see Stefanics et al. 2011). Similarly, in studies using a slow (300 ms) rise time as a standard (e.g., Richardson et al. 2004), we found that children with dyslexia could discriminate very rapid rise times from this standard (e.g., 15 ms), but could not discriminate between slower rise times. Children with dyslexia aged 9 years could only discriminate rise times of 15 – 60 ms from a rise time of 300 ms. In a recent longitudinal study, we found that rise time discrimination did develop in children with dyslexia, but not age-appropriately (Goswami et al. 2012). By the age of 12 years, children with dyslexia were showing significantly poorer rise time discrimination in comparison to younger reading level matched control children aged 10 years. A persistent difficulty in discriminating between different rise times is likely to affect the syllabic parsing of the speech stream, as the shape of the amplitude modulation characterising different syllables (such as plosives versus sonorants) will not be well-distinguished. Indeed, behavioural data suggest that the rise time difficulties found in developmental dyslexia are linked with rhythmic and prosodic difficulties in linguistic tasks.

## 2.2. Speech rhythm and syllable stress perception in dyslexia

Behavioural studies with English children have found that individual differences in rise time perception are related to individual differences in a range of rhythmic and prosodic tasks (see Goswami 2011 for overview). For example, sensitivity to prosodic patterns in dyslexia can be

measured using a reiterant speech task, the “DeeDee” task (Whalley and Hansen 2006). Here each syllable in a word is converted into the same syllable (DEE), removing most phonetic information while retaining the stress and rhythm patterns of the original words and phrases. Goswami, Gerson, and Astruc (2010) created two novel DeeDee measures for children in a picture recognition task, one based on celebrity names (e.g., *David Beckham*) and one based on film and book titles (e.g., *Harry Potter*). In the first task, the words were “spoken in DeeDees”, and hence retained the metrical phrase-level structure of the originals. In the second task, this phrase-level information was removed by utilising 4 synthesised tokens, “DEE” and “dee” in initial and final position, which served to emphasise syllable stress (strong or weak). The selected film and book titles were then created by combining the synthetic “Dees” in the appropriate strong-weak syllable sequence. Goswami et al. (2010) reported that both tasks were performed more poorly by 12-year-old children with developmental dyslexia compared to 12-year-old controls. In unpublished data, we have found that 9-year-old children with dyslexia perform significantly more poorly in the synthetic DeeDee task than 7-year-old typically-developing controls – a “reading level match” experimental design (Goswami et al. submitted). This finding is methodologically important, as the reading level match research design holds reading level constant rather than chronological age and thus gives a mental age advantage to the dyslexic children. As the dyslexics were significantly less accurate than *younger* controls in the prosodic task, this experimental result suggests that the prosodic difficulty in dyslexia is a profound one.

Direct measures of stress perception yield similar results with dyslexic participants. Leong, Hämäläinen, Soltész, and Goswami (2011) designed a direct syllable stress perception task based on 4-syllable words that had either first syllable primary stress (2000 stress template) or second syllable primary stress (0200 stress template). Participants were required to make a same-different judgement about pairs of words that either shared a stress template (e.g., both 2000) or did not (e.g., 0200 versus 2000). Highly compensated adults with dyslexia (undergraduate students at the University of Cambridge) showed significantly lower sensitivity to syllable stress ( $d'$  measure) than adults without dyslexia (other Cambridge students) in this task. Reduced stress sensitivity was found *both* when different lexical templates had to be compared (e.g., “*maternity-ridiculous*”), and when the same word repeated twice had to be compared (e.g., “*difficulty-difficulty*”). The unique auditory predictor of individual differences in the direct stress perception task was rise time. Again, we have recently found similar impairments in the direct perception of syllable stress in children with developmental dyslexia (Goswami et al. 2012).

We have also developed a “metrical musical task” to test whether perceptual difficulties with metrical rhythm occur in tasks based on musical notes instead of speech syllables. Huss et al. (2011) developed a “chime bar” task based on short “tunes” played on the note of G that were 6 – 15 notes in length, were in either 4/4 time or 3/4 time, were based on an isochronous beat structure of 2 Hz (500 ms), and had different metrical structures conveyed by increasing the intensity of the accented note in

a bar. Disruptions to metrical structure were introduced by increasing the duration of the accented note by either 100 ms or 166 ms. Ten-year-old children with and without dyslexia and younger reading level controls were tested. Huss et al. found that the children with dyslexia were significantly worse at perceiving changes in metrical structure compared to typically-reading same age controls (chronological age matched controls). At the age of 10 years, the dyslexic children performed at the same level as 8-year-old reading level matched controls. Performance in the metrical musical task explained over 60% of the variance in reading in the age-matched sample along with age and IQ. Indeed, performance in the musical task showed stronger associations with reading than traditional phonological awareness measures (like rhyme awareness).

The same musical task was re-administered a year later, when the children with dyslexia were aged 11 years (Goswami et al. submitted). Despite the task being familiar, the dyslexic children were now performing more poorly than the reading level matched controls (now 9-year-olds), suggestive of a profound perceptual difficulty. Additional analyses suggested that a key difficulty was perceiving the patterns of beat distribution in the different metrical arrangements, which also involved sensitivity to auditory grouping. Finally, rhythmic synchronisation as measured by tapping in time with a metronome beat is impaired in developmental dyslexia. Children with developmental dyslexia are much more variable in tapping on the beat in comparison to non-dyslexic children (significant differences were found at temporal rates of 2.5Hz and 2 Hz, see Thomson and Goswami 2008). These dyslexic impairments in rhythmic synchronisation are still present in adulthood, as Cambridge university undergraduates with dyslexia are also significantly poorer at synchronising to the beat (Thomson et al. 2006). Indeed, children with speech and language impairments (SLI) also show rhythmic synchronisation deficits (tapping to a beat at 1.5 Hz and 2 Hz, Corriveau and Goswami 2009).

### **3. Amplitude modulation, rise time and the speech signal**

The behavioural and acoustic data from individuals with developmental dyslexia suggest that the accurate perception of amplitude modulation and rise time play an important role in the phonological representation of syllable structure and speech rhythm. In recent work, we have begun to model the syllabic and prosodic structure of speech in terms of the dominant amplitude modulations in the speech signal and their *phase relationships* (Amplitude Modulation Phase Hierarchy [AMPH] models, see Leong 2012). In one of Leong's AMPH models, a multi-dimensional representation of the speech envelope is used, simultaneously capturing the amplitude modulation patterns occurring at 5 different acoustic frequencies and at 3 different modulation rates. This spectro-temporal decomposition of the envelope enables the visualisation of slower amplitude modulations related to prosodic stress and syllable structure, as well as faster amplitude modulations generated, for example, by brief plosive bursts in stop consonants (Figure 3). Figure 3 shows two 4-syllable English words, "*comfortable*" and

"debatable", and the amplitude changes in each of the 5 frequency bands at 3 different temporal rates within the envelope. These frequency bands and temporal rates were determined by a principal components analysis of the spectro-temporal characteristics of a multi-speaker speech corpus (see Leong, 2012, for detail), and are reproduced here for illustrative purposes. While "COMfortable" has first-syllable lexical stress, "deBAteable" has second-syllable stress (as indicated in capitals). The figures show that at the slowest AM rate (below 2.5 Hz, designated the *Stress AM* rate), the location of primary lexical stress is marked by a large peak in amplitude modulation (indicated as a dark red colour). For example, the stressed first syllable "COM" in "COMfortable" clearly elicits a large amplitude peak right at the beginning of the word across all the frequency bands. For "deBAteable", this initial increase in amplitude is more gradual so that the peak in amplitude modulation across frequencies is reached later, at the second stressed syllable "BAY" rather than the first unstressed syllable "de". At the second AM rate (2.5 – 12 Hz, designated the *Syllable AM* rate), there are 4 peaks in amplitude modulation for each 4-syllable word, showing clearly that amplitude modulation at this rate relates to syllable structure. These amplitude modulations are more dominant in the lower frequency bands, which typically contain higher power. At the third temporal rate, designated the *Phoneme AM* rate (12 – 40 Hz), more transient and localised amplitude peaks are visible, with peaks corresponding to the production of individual phonemes within syllables, typically onset consonants like /t/ and /k/. These peaks are more variable in their positioning across the different frequency bands, reflecting their different spectral content.

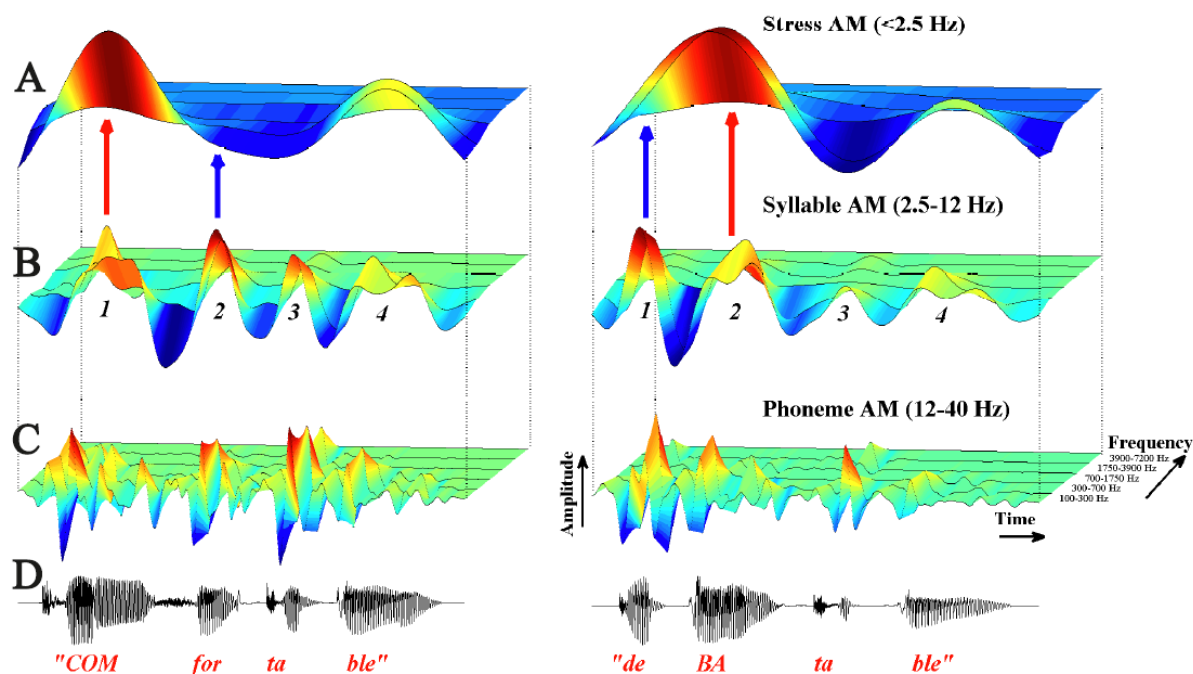


Figure 3. Spectro-temporal representation of the speech amplitude envelope for the words "COMfortable" (left column) and "deBAteable" (right column). 3 dominant amplitude modulation (AM) rates within the envelope are labelled as A (*Stress A*), B (*Syllable AM*) and C (*Phoneme AM*) respectively. For each



*subplot, the change in amplitude (y-axis) over time (x-axis) within each of 5 frequency bands (z-axis) is shown. High amplitude is indicated in red, low amplitude is indicated in blue. Notice that the modulation pattern of the Syllable AM (middle row) reveals the 4 syllables contained within each word (numbered from 1-4). The difference in stress patterning between the two words is reflected as a shift in the timing of the initial peak of the Stress AM toward the stressed syllable (red arrow), and away from the unstressed syllable (blue arrow). The original acoustic waveform for the word, D, s is shown at the bottom in black.*

However, there is another important feature of these AMs at different modulation rates, and that is their temporal alignment with each other. Taken individually, neither Stress AMs nor Syllable AMs contain sufficient information to determine the exact prosodic pattern of stressed and unstressed syllables within each word. The Stress AM indicates the presence of prosodic stress, but not its syllabic location. The Syllable AM indicates the number and location of syllables, but not their prosodic status. However, when considered together, the relative alignments of the Syllable and Stress AM reveal the syllable stress pattern of each utterance. Hence for "COMfortable", the peak in the Stress AM is temporally aligned with the first peak in the Syllable AM (marked by a red arrow), indicating first syllable stress. Meanwhile, for "deBAtable", the Stress AM peak is temporally-shifted so that it is now aligned with the *second* peak in the Syllable AM instead (also indicated by a red arrow; the blue arrows denote unstressed syllables), indicating second syllable stress. This temporal shift between the two AMs can also be expressed in terms of a change in their relative oscillatory *phase*. Hence the phase relations between the AM-peaks at the Stress AM rate and the Syllable AM rate seem likely to contribute to the perceptual experience of strong versus weak syllables. These phase relationships could be important non-durational acoustic cues that contribute to the perceptual experience of speech rhythm and prosody, suggesting potentially new avenues for research.

#### **4. Amplitude Modulation Phase Hierarchies and the Perception of Rhythm**

If the phase relations between the Stress AM rate and the Syllable AM rate are indeed related to prosodic experience and whether we experience syllables as weak versus strong, then changing the *phase relations* between the Syllable and Stress AM rates should change the perception of speech rhythm. In other words, by physically changing these phase relations, we can test systematically whether the phase relations between the AM-peaks at the Stress AM rate and the Syllable AM rate correspond to the perceptual experience of strong versus weak syllables. In recent work, we extracted a temporal hierarchy of AMs at five different modulation rates from nursery rhyme sentences with distinctive trochaic or iambic metrical rhythm patterns (Leong et al. 2011). The temporal hierarchy is shown in Figures 4 and 5, taking as examples the trochaic English nursery rhyme "Mary Mary quite contrary, how does your garden grow" versus the iambic nursery rhyme "As I was going to St Ives I met a man with seven wives". Leong and colleagues played the AM tiers back to listeners either singly or in paired combination to see which AM rates or combinations would elicit the best rhythm judgment. The 5 tiers in the AM hierarchy were designated respectively as the 'Slow' tier (<1 Hz), the

‘Stress’ tier ( $\sim 2$  Hz), the ‘Syllable’ tier ( $\sim 5$  Hz), the ‘Sub-beat’ tier ( $\sim 12$  Hz), and the ‘Fast’ tier ( $\sim 35$  Hz). In this 5-tier AM hierarchy, individual Syllable AM cycles correspond to individual spoken syllables (8 for each sentence in Figures 4 & 5), while individual Stress AM cycles correspond to individual prosodic stress feet (4 for each sentence in Figures 4 & 5). The phase relationship between the Syllable and Stress AMs provides the syllable stress pattern of Strong (S) and weak (w) syllables. This phase relationship is shown with vertical red lines in Figures 4b & 5b. Syllable AM cycles that occur near the peak of the Stress AM are considered ‘Strong’ (S). Syllable AM cycles that occur near the trough of the Stress AM are considered ‘weak’ (w). To make the AMs audible, we used them to modulate a sine-tone carrier (i.e. single channel tone-vocoding). Note that the phonetic fine structure of the signal was intentionally discarded, and only AMs derived from the amplitude envelope were used to modulate the sine-tone carrier. The stimuli sounded like Morse-code or fast fluttering, depending on the AM tier used for vocoding. In each case, participants were asked to identify the original nursery rhyme using only the rhythm pattern that they heard.

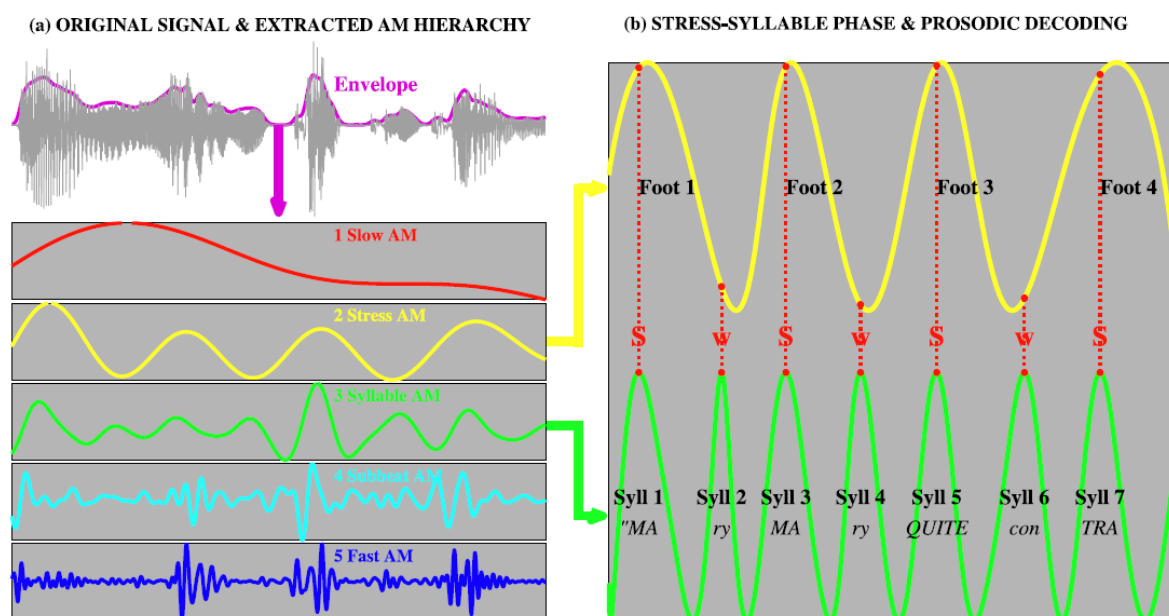


Figure 4. *Encoding of syllable stress patterns by amplitude modulations in the envelope, illustrated using a nursery rhyme with a trochaic ('S-w') syllable stress pattern.* (Left) The amplitude envelope is shown in purple, superimposed onto the original signal waveform. The envelope is filtered into 5 modulation bands, forming the 5 tiers of an AM hierarchy: (1) 'Slow' AM tier (0.5-0.8 Hz); (2) 'Stress' AM tier (0.8-2.3 Hz), (3) 'Syllable' AM tier (2.3-7 Hz), (4) 'Sub-beat' AM tier (7-20 Hz), and (5) 'Fast' AM tier (20-50 Hz). (Right) The phase pattern of the 'Stress' (yellow) and 'Syllable' (green) AM tiers is shown.

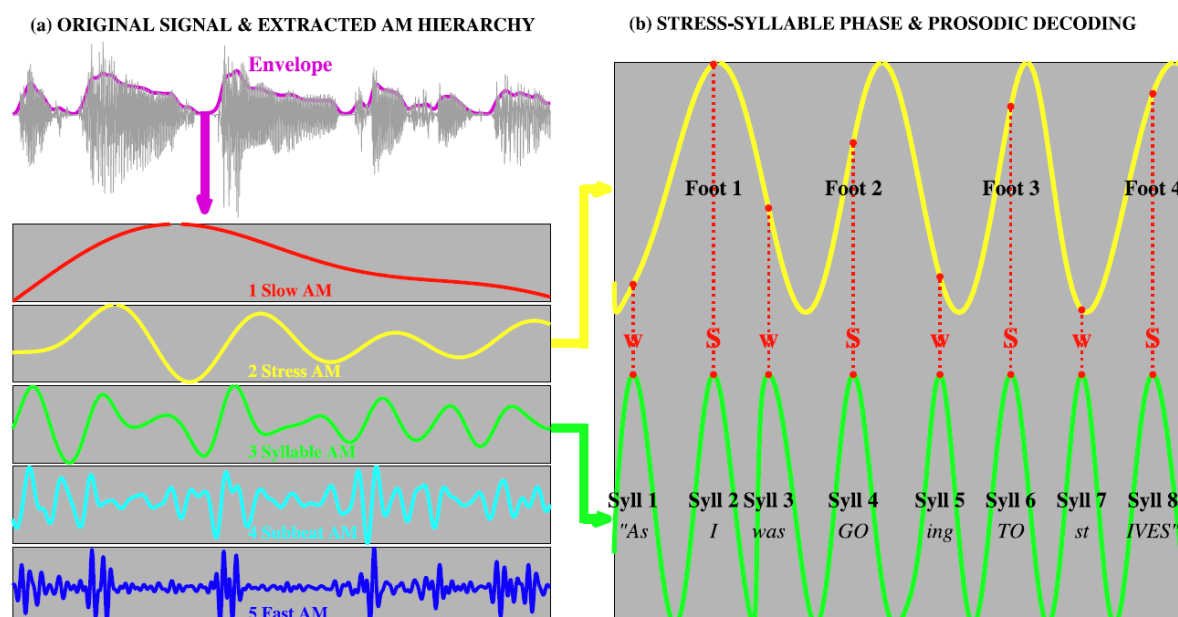


Figure 5. Encoding of syllable stress patterns by amplitude modulations in the envelope, illustrated using a nursery rhyme with an iambic ('w-S') syllable stress pattern.

Since poetic meter is described in terms of patterns of stressed and unstressed syllables, we predicted that the putative AM correlates (Stress and Syllable AMs) would contain more metrical information than the other AM tiers or their combinations, and hence result in the best rhythm discrimination when presented together. Second, if the specific pattern of strong and weak syllables perceived by listeners depends on the oscillatory phase relationship between the Stress AM and the Syllable AM, this phase relationship could provide the statistics underlying metrical foot patterns, constituting a 'phase code' for metrical rhythm. To test this hypothesis, we systematically shifted the phase-relationship between Stress and Syllable AM tiers to see if this would alter the perceived metrical pattern of the nursery rhyme sentences. We expected that listeners would base their metrical perception on temporal phase-relationships between Stress and Syllable AMs, even when these phase-relationships were artificially re-aligned to create novel 'phase-shifted' variants.

This was essentially what we found in Leong et al. (2011). Rhythm discrimination using Stress+Syllable AMs exceeded rhythm discrimination using other tiers or tier combinations from the AM hierarchy. Furthermore, the rhythm pattern that listeners perceived systematically tracked the phase relationship between Stress and Syllable AM tiers. For example, an originally trochaic sentence was judged as iambic by listeners when the Stress AM was shifted with respect to the Syllable AM by  $1\pi$  radians (half an oscillatory cycle). Therefore, listener judgements supported the proposal that metrical rhythm patterns in speech could be cued acoustically by hierarchical AM patterns in the amplitude envelope. Theoretically, these AM patterns could represent linguistic units (such as syllables and prosodic feet) on the one hand, and drive (entrain) neural activity on the other hand (see Section 7). Figure 3 depicts the hypothesised relationship between Stress and Syllable AM tiers and prosodic structure (syllables and feet). In addition, the data showed that this hierarchical

representation yielded an important new temporal statistic - the *phase relationship between AM tiers*. The Stress-Syllable phase relationship was an important perceptual cue to metrical rhythm pattern in the experiment, suggesting that human perception of speech rhythm may depend on such temporal modulation statistics.

One significant advantage of coding rhythm via AM phase patterns is that such coding would be robust to random durational variations when the speaker speeds up or slows down - a major contributor to anisochrony in speech, and a major challenge for accounts of speech encoding based on oscillatory models of neuronal entrainment (see Section 7). Phase relationships between Stress and Syllable AMs within a particular stress foot can remain constant even if the overall duration of the foot varies from foot to foot. This is because once phase-locked, both Stress and Syllable oscillatory cycles will compress or stretch in synchrony, yielding the same rhythmic pattern (coded in the phase relationship) even if foot duration changes. On Leong et al.'s (2011) data, rhythm discrimination is therefore particularly reliant on the phase relationships between Stress and Syllable rate amplitude modulations in the speech envelope. In "stress-timed" languages such as English, the Stress-Syllable AM phase relationship could be an important contributor to speech rhythm, whereas in "syllable-timed" or "mora-timed" languages, faster rates of amplitude modulation could play a larger role

## 5. Cautionary Notes

The perceptual rhythm experiments described briefly above (Leong et al. 2011; see also Leong 2012 for detail) suggest that speech rhythm may in part depend on the hierarchical nature of AMs in the envelope, with the temporal statistics critical to capturing speech rhythm being the *phase relationships* between AM tiers. Our proposal is based on our behavioural data related to phonology and dyslexia, rather than to data related to speech processing and speech comprehension per se. Nevertheless, in the wider speech processing field, it has been argued that the importance of the speech envelope is being over-emphasised by authors such as Giraud and Poeppel (2012) (see, for example, Obleser, Herrmann, and Henry 2012). Obleser et al. (2012) argued that spectral content can be as important as temporal envelope information for speech intelligibility, and noted that the speech envelope can be easily obscured in noisy environments. Further, they pointed out that *frequency* modulation in the speech signal, for example in the 3 Hz range, can also entrain the neuronal oscillations discussed by Giraud and Poeppel (2012). Meanwhile, Cummins (2012) has argued that the claim that the amplitude envelope of speech carries information about syllables should be interpreted with caution. Cummins (2012) argued that "quasi-cyclic jaw wagging" does not modulate the amplitude envelope in a way that provides unambiguous syllabic information. Rather, as he pointed out, the amplitude envelope is related to the movement of all the articulators in complex ways. Cummins (2012) further argued that the assumption that the syllable can be conceptualised as an oscillatory system based on a quasi-periodic mean syllable rate of around 5 Hz is too lax.

Cummins' core argument was that the speech signal is being mis-characterised in oscillatory accounts, since speech is not sufficiently periodic to sustain true oscillatory entrainment.

In our view, there are currently too few data to enable a systematic assessment of the role of envelope-driven neural oscillatory entrainment in either speech comprehension or phonological representation (although see Sections 7 and 8). Nevertheless, amplitude modulation and rise time discrimination are clearly related to individual differences as measured behaviourally in phonological development and rhythmic processing (as outlined in Section 2 above). Furthermore, individual differences in synchronisation (e.g., in tapping to a beat) are also related to individual differences in phonology and reading development. Accordingly, we find an AM-driven oscillatory perspective (e.g., Giraud and Poeppel 2012) attractive theoretically, while acknowledging that systematic exploration of this perspective is currently lacking. In our view, an AM phase hierarchy perspective (Leong 2012) helps to specify how speech rhythm and acoustic temporal structure may be related, and enables investigation of neuronal entrainment (see Sections 7 and 8). Since neuronal oscillations may be entrained by other forms of statistical regularities apart from simple periodic 'beats' (Obleser et al. 2012), this opens the possibility that other acoustic statistical regularities could help to support neuronal oscillatory entrainment, even in 'aperiodic' speech. For example, the sequence of strong-weak syllable stress patterns in a *metrically-regular* sentence generates a robust predictive neural response in the listener that facilitates their semantic processing of future words (Rothermich et al. 2012). This neural response is generated even when the sentence is not spoken isochronously (i.e., it is naturally 'aperiodic') and does not contain exaggerated stress emphasis. The AMPH method is one way to quantify these statistical regularities in prosodic patterning.

Similarly, while individual differences in children's discrimination of amplitude envelope rise time appear to be systematically related to individual differences in phonological and reading development, this does not mean that rise time *per se* is the core construct. Rise time sensitivity may be a marker of something else, such as the efficiency of auditory oscillatory systems. Alternatively, an unidentified third factor, which covaries with both rise time discrimination and phonological development, may underlie the behavioural relationships that we have documented to date. We noted in Section 2 that it is likely to be the perceptual effects of different rise times and their phase relations, rather than sensitivity to one particular rise time in the envelope, that is likely to be important for phonological development. As Scott and McGettigan (2012) have pointed out, not all signal rise times are equal in their perceptual effects. Indeed, Scott and McGettigan (2012) argued that perceptual centres rather than syllable-specific rise times may turn out to be more important in explaining the patterns in the dyslexia data, and this is quite possible (see Goswami et al. 2002; 2012). On the AMPH perspective described here, the child's perception of global rhythm would depend on how efficiently the phase relations between AM patterns with different rise times were extracted. Empirical data are currently insufficient to determine the precise mechanisms by which rise time perception, amplitude modulation, rhythm and phonology are related. Nevertheless, we remain

optimistic that a combination of modelling and behavioural experiments based on an AMPH perspective, coupled with hypothesis-driven oscillatory studies (see Sections 7 and 8), will yield important data concerning temporal structure and speech rhythm. One reason for such optimism is the theoretical convergence visible in the other papers in this symposium.

## **6. Theoretical convergence with Krivokapić's foot oscillators and Arvaniti and Rodriquez's supra-syllabic periodicities**

For example, the AM phase hierarchy data described in Sections 3 and 4 support Arvaniti and Rodriquez's (this volume) view that rhythm patterns in speech are not found in durational isochronies. Arvaniti and Rodriquez compared perceptual discrimination in an AAX paradigm using stimuli based on words in English, Polish, Danish, Spanish, Greek and Korean. These six languages span a range of rhythm classes. Nevertheless, Arvaniti and Rodriquez reported that successful discrimination did not depend on differences between rhythm classes. While rhythmic timing did play a role in successful discrimination in their paradigm, so did other acoustic factors, particularly F0 and tempo. Arvaniti and Rodriquez noted that speech rhythm may be cross-linguistically based on the grouping of prosodic constituents such as syllables and their relative prominence. This view converges with the AMPH perspective. In AMPH models (Leong 2012), the basic statistical regularities that create prominence-based groupings would be the *phase relations* between the Stress and Syllable AM tiers. Since the phase relationship between AM tiers can remain constant even while foot duration varies, our data suggest that the experience of speech rhythm may arise from regularities *in phase relationships* rather than invariance in foot or syllable duration. Hence, as argued by Arvaniti and Rodriquez, specific temporal intervals based on the duration of linguistic entities such as vocalic versus consonantal elements is the wrong place to look for the temporal structure that yields speech rhythm.

Meanwhile, the AMPH perspective also supports the conclusions drawn by Krivokapic (this volume). In Krivokapic's study, speakers of English from India and American English speakers either read aloud solo, or read aloud together. Both populations showed mixed properties when reading solo; nevertheless each population tended towards their classical rhythm type (e.g., the American English speakers showed characteristics of stress timing, while the Indian English speakers showed characteristics of syllable timing). For example, while all speakers showed a tendency to increase the duration of the foot as the number of syllables in the foot increased, traditionally considered a property of rhythmic syllable timing, qualitatively this tendency was stronger in the Indian English speakers. When reading together, a degree of rhythmic convergence was observed. One Indian English speaker in particular became rhythmically more like the American English speakers. This Indian English speaker appeared to temporally "squeeze" extra syllables in the foot, so that the duration of the foot became less affected by the number of syllables in it. At the same time, this speaker produced shorter stressed syllables than when speaking solo. At the group level, some

rhythmic convergence could also be observed in the American English participants, who became more syllable-timed. Krivokapić (this volume) suggests tentatively that temporal coupling between a “foot oscillator” and a “syllable oscillator” could underpin the observed rhythmic convergence. Drawing on oscillatory models for articulatory timing proposed by Saltzman and his colleagues (Goldstein, Byrd, and Saltzman 2006; Saltzman et al. 2008; Nam, Goldstein, and Saltzman 2009), Krivokapić suggests that convergence between the coupling relations among the different oscillators (foot, syllable etc.) could be the key variable to study when relating temporal structure of speech rhythm.

Again, we see exciting convergence here with our AMPH perspective. Since speech is produced by motor articulators like the jaw, lips and tongue, the oscillatory AM tiers and patterns that we describe could well correspond to these discrete articulators and their couplings (see also Tilsen 2009). Therefore, the rhythmically-important Stress-Syllable phase relationship documented by Leong, Hämäläinen, Soltész, and Goswami (2011) could be an acoustic correlate of the 'coupling function' proposed by Saltzman and colleagues (2008) that phase-locks the articulatory foot and syllable oscillators together. In Krivokapić's (this volume) data, it is this 'coupling function' that becomes mutually tuned in speaker convergence, leading to a convergence in rhythm. On an AMPH perspective, rhythmic convergence could be marked by a change in the ratio of Stress:Syllable AM power in the speech envelope (Leong 2012). The data of Leong et al. (2011) would further suggest that rhythmic convergence could also be marked by a change in the strength of phase-locking observed between AMs in the speech envelope (in addition to, or independent of, power changes). These predictions from an AMPH perspective are open to empirical testing, and may offer theoretical convergence concerning the different perspectives on speech rhythm discussed in this volume.

## **7. The brain: Oscillatory neuronal entrainment and speech encoding**

As noted earlier, the amplitude envelope of speech contains a range of modulations at different temporal rates, with the 'modulation spectrum' within the envelope typically showing the highest power between 2 – 12 Hz (and peaking around 3 - 5 Hz irrespective of differences in language or speech rate, see Greenberg et al. 2003; Greenberg 2006; Houtgast and Steeneken 1985). These observations have been related to the neural encoding of speech by a family of “multi-time resolution models” of speech processing developed in the field of auditory neuroscience (e.g., Poeppel 2003; Hickok and Poeppel 2007; Ghitza and Greenberg 2009; Ghitza 2011). Multi-time resolution models of speech processing suggest that different rates of amplitude modulation in the envelope are encoded by neuronal oscillations at corresponding temporal rates. Therefore, different levels of the prosodic hierarchy may be encoded by neuronal oscillatory networks that align their endogenous activity with amplitude modulation at different temporal rates in the speech signal (Giraud and Poeppel 2012).

These models have been supported by studies in auditory neuroscience that have demonstrated that neuronal oscillations indeed entrain to the syllable structure of speech (e.g., Luo and Poeppel 2007). The multi-modal nature of speech has been addressed by studies showing that visual and auditory *low frequency phase alignment* plays a critical role in efficient speech perception (e.g., Luo, Liu, and Poeppel 2010). The temporal aspects of syllable production are constrained within certain physiological norms which are reflected in both auditory and visual temporal information. Configuring the vocal tract to shape the acoustic signal simultaneously deforms the face, and speech acoustics can be estimated reliably from face motion or “visual prosody” (e.g., Munhall et al. 2004; Yehia, Kuratate, & Vatikiotis-Bateson 2002). Efficient resetting of the *phase* of oscillations within audiovisual neural networks has been proposed to facilitate speech processing (Schroeder et al. 2008). The mechanisms by which oscillatory networks in auditory cortex entrain to acoustic input have been investigated in detail using direct cell recordings (electrophysiology) in animal models (e.g., Lakatos et al. 2008). Electrophysiology has revealed that the natural rhythmic fluctuations in excitability of neuronal networks in auditory cortex can entrain to the temporal structure of rhythmic input (e.g., Lakatos et al. 2008; Schroeder and Lakatos 2009). When rhythmic predictability is established, neural networks realign the phase of their oscillations so that they are in a high excitability phase when a new event occurs, which enhances (amplifies) the neuronal processing of that event. Auditory events that are “out of phase” with the rhythmic stimulus stream are suppressed, as they arrive during low excitability phases. Hence these oscillatory networks align their firing patterns so that they are firing in-phase with the rhythmic structure of acoustic signals, thereby supporting signal parsing and efficient signal encoding. As the speech signal does not appear to be encoded efficiently in dyslexia, with evidence across languages for impaired phonological representations, such oscillatory mechanisms might be expected to be atypical in children with reading difficulties.

## **8. Neuronal rhythmic entrainment and reading development**

We have recently begun to gather neural data relevant to the proposal that impaired phonological representation in developmental dyslexia is related to atypical auditory oscillatory entrainment. As individual differences in phonological development are the strongest behavioural predictor that we have of individual differences in reading development (Goswami and Bryant 1990), we have also studied neuronal rhythmic entrainment in typically-developing children without dyslexia and its relations to reading. In our study of typically-developing children (Power et al. 2012), we used a rhythmic speech paradigm (repetition of the syllable “ba” at a 2 Hz rate) to study both auditory and visual entrainment. Children either watched a video of a “talking head” repeating the syllable (enabling auditory and visual entrainment), heard the auditory soundtrack alone (enabling auditory entrainment), or saw the talking head without hearing the speech (enabling visual entrainment). We demonstrated significant auditory and visual entrainment to rhythmic speech, with entrainment both at



the stimulation rate (2 Hz, delta) and also (for auditory entrainment) at the theta rate studied by Poeppel and his colleagues (the “syllable rate”, between 4 – 7 Hz, see Poeppel, 2003). In addition, we found that preferred phase in the theta band was altered by predictive visual speech information. In Power et al. (2012), we argued that the data supported Schroeder et al.’s (2008) suggestion that visual rhythmic information modulates auditory oscillations to the optimal phase for auditory processing and audio-visual integration. Further, we found that individual differences in auditory theta entrainment were related to individual differences in reading development in our sample of children.

In a second oscillatory study using MEG, we compared neuronal entrainment to amplitude-modulated noise at 4 temporal rates (2 Hz, 4 Hz, 10 Hz, 20 Hz) in adults with and without a childhood history of developmental dyslexia (Hämäläinen et al. 2012). As the slower temporal rates (2 Hz and 4 Hz) would carry prosodic and syllable-level information in speech, we predicted reduced phase locking at these two rates in the dyslexic participants. We expected this difference to be right-lateralised, as entrainment to slower temporal rates is greater in the right hemisphere in neurotypical adults (see Hickok and Poeppel 2007). The data showed significantly reduced entrainment to the 2 Hz rate in a superior source in the right hemisphere in the dyslexic participants, and also significantly reduced right hemisphere entrainment overall for the dyslexics, although the differences at 4 Hz did not reach significance. There was also significantly *greater* entrainment at the 10 Hz rate in the *left hemisphere* in dyslexia, a result which we had not predicted, and which may indicate compensatory mechanisms.

Finally, in a recent study (Soltész et al. 2013), we investigated possible links between oscillatory phase and behaviour in dyslexia in a rhythmic entrainment paradigm. Most studies of oscillatory entrainment have relied on rhythmic paradigms, typically on the presentation of isochronous streams of simple tones or light flashes (e.g., Lakatos et al. 2008; Stefanics et al. 2010; Gomez-Ramirez et al. 2011). Soltész et al. (2013) investigated event-related oscillatory EEG activity and contingent negative variation (CNV) to an auditory rhythmic tone stream in dyslexia. Tones were delivered at a rate of 2 Hz, and participants had to press a button when a tone was replaced in the rhythmic stream by white noise. Adults with dyslexia were as fast and accurate as control adults in making the button press response. Nevertheless, the CNV (an ERP response related to the efficiency of predictive timing) and inter-trial coherence (ITC, a measure of neuronal entrainment) were both significantly reduced in the participants with dyslexia. Whereas behaviour (response time) was related to the instantaneous phase of the delta oscillation in the controls, as also reported by Stefanics et al. (2010: faster responses are found during the rising phase of the oscillation), this relationship was absent in the dyslexics. Individual differences in both the amplitude of the CNV and the ITC were significantly related to individual differences in phonological processing (detection of syllable stress and phoneme awareness in a Spoonerism task), and to reading and spelling. These data are suggestive of atypical functional neuronal rhythmic entrainment in developmental dyslexia to low-frequency rhythmic auditory stimuli. According to an AMPH perspective, atypical rhythmic entrainment at 2 Hz

would be expected to affect the ability to predict when stressed syllables are likely to occur in the continuous acoustic stream, and therefore would affect the representation of the temporal structure of spoken words – phonological representation.

This oscillatory account should be regarded as speculative, and more research is clearly required. Nevertheless, taken together with the developmental data reported by Power et al. (2012) and the AM entrainment data reported by Hämäläinen et al. (2012), the oscillatory study reported by Soltész et al. (2013) suggests that the efficiency of neuronal entrainment at lower temporal frequencies (delta and theta) is related to reading development and dyslexia. The AMPH approach to speech rhythm would predict that atypical entrainment at slower rates would have consequences throughout the phonological system. Indeed, a study of adult French-speaking dyslexics (Lehongre et al. 2011) has suggested atypical oscillatory entrainment to faster gamma rate (~30 Hz) amplitude modulations (these authors measured entrainment to a complex white noise stimulus linearly increasing in modulation rate between 10-80 Hz). Using the auditory steady-state response (ASSR) as a measure of oscillatory entrainment (a power increase at the rate of stimulation), the French study found that the ASSR at 30 Hz (low gamma rate) in auditory cortex was not left-dominant in dyslexic adults, in contrast to control participants. The degree of reduced leftward bias was correlated with measures of phonemic processing. Lehongre et al. also observed increased entrainment at high frequencies (>50 Hz) in dyslexics, and this was argued to reflect “over-sampling” at the phonemic rate. Therefore, gamma entrainment may also be atypical in developmental dyslexia.

“Over-sampling” may be expected theoretically to lead to phonological representations in the mental lexicon that are differently-specified at the phoneme level. This atypical representation would be additional to the syllable-level differences revealed by our own work. For example, allophones may be perceived as different speech sounds. Although beyond the scope of the current paper, there is some behavioural evidence that is consistent with this viewpoint. For example, children with dyslexia indeed continue to perceive allophones that control children no longer distinguish (e.g., Serniclaes et al. 2004). Children with dyslexia are also significantly more sensitive to a Ba/Wa discrimination based on frequency rise time (formant transition duration) than good readers (Goswami et al. 2011b), possibly suggestive of over-specified representations. Finally, children who are poor readers can show an enhanced brainstem EEG response to allophones in comparison to good readers (Chandrasekaran et al. 2009).

On an AMPH perspective, atypical neuronal entrainment to the low frequency temporal information supporting prosodic structure would have effects throughout the phonological system, so that dyslexic children’s phonological representations for spoken words would be *different* from those of other children in terms of the global perceptual representation (rather than “under-specified” or “imprecise”, see Snowling 2000). Each perceptual experience of speech from infancy onwards would be atypical, so that a lexicon of spoken word forms would develop that was subtly different at all linguistic levels, including the phonetic level. In accord with the theoretical view that phonological

representations can be conceived of as episodic, as suggested for example by Goldinger (1998), all perceptual details of experienced spoken words, including apparently idiosyncratic aspects such as speaker voice, would be stored in memory and would form the basis of spoken word representation. By hypothesis, for dyslexic individuals many of these perceptual details would be *different* from the details stored by neurotypical individuals, for example because of atypical amplitude modulation phase alignment and imprecise discrimination of amplitude rise times. This auditory hypothesis would predict that atypical phase alignment and impaired rise time discrimination should be found in affected individuals across languages, at least prior to acquiring reading. However, the effects of atypical auditory processing on reading development would be expected to differ with orthographic learning in different languages, depending on factors like orthographic transparency and whether stress is marked in the orthography.

### **9. Conclusion: Rhythm and temporal structure—into the future?**

The behavioural and neural data summarised here from children and adults with and without dyslexia support the proposals made by Poeppel, Giraud, Greenberg and their colleagues that the brain detects and represents (entrains to) temporal structure in the acoustic environment (multi-time resolution models of speech processing, e.g., Hickok and Poeppel 2007; Ghitza and Greenberg 2009; Giraud and Poeppel 2012), and the view that the efficiency of this process is related to phonological development (Goswami 2011). Further, our data suggest that the temporal structure of amplitude modulation yields phase regularities and statistics that influence the perceptual experience of rhythm (Leong, Turner, Stone, and Goswami 2011; Leong 2012). In particular, the phase relationship between the ‘Stress’ and ‘Syllable’ AM tiers in the amplitude envelope appears to yield a temporal statistic that is important for English speech rhythm. Clearly, the critical AM tiers providing these statistics may differ across languages with different rhythm “types. There may also be individual differences in the precision with which these phase statistics are computed as languages are learned. Nevertheless, this temporal statistic may yield an approximate periodic structure to which the infant brain may entrain. Even small individual differences in the efficiency of this early entrainment would have developmental consequences for the phonological lexicon. Developmentally, children’s language games and nursery rhyme play may be important for improving the efficiency of such neuronal entrainment mechanisms.

By this developmental proposal, infants would use neural entrainment to syllable beats as one foundation for language acquisition, perhaps beginning inside the womb. Successful entrainment would enable accurate predictions about the timing of important future speech events (e.g., stressed syllables, which occur on average at a 2 Hz rate across languages). We propose that infants may form temporal expectancies about when the next syllable should occur on the basis of statistical regularities that are cued acoustically by features like AM rise time. This “quasi-periodic skeleton” may form a basic temporal structure onto which the different grouping and phonetic factors that characterise

different languages may then be scaffolded. This would enable early perceptual learning to reflect the different combinations of phonetic inventories and critical features that are unique in a language, while simultaneously using universal neural oscillatory processes that track temporal regularities to encode the beat distribution patterns in speech that are cued by stressed syllables and P-centres (see Kotz and Schwartz 2010). With perceptual experience, infants would then acquire the non-isochronous rhythmic and prosodic characteristics of their native language, scaffolding these temporal features onto this underlying quasi-periodic skeleton. Focusing future research on AM hierarchies in natural speech is methodologically challenging, but it may yield rich rewards for our understanding of speech rhythm.

### **Acknowledgments**

We thank the Medical Research Council (grants G0400574 and G0902375), and the Economic and Social Research Council (grant RES-000-23-0475) for funding the published work summarised in this paper. We also thank the Harold Hyam Wingate Foundation for the provision of a Research Scholarship to VL, which supported the AMPH work.

Correspondence e-mail address: [ucg10@cam.ac.uk](mailto:ucg10@cam.ac.uk)

## References

- Arvaniti, Amalia, & Tara Rodriguez. This volume. The role of rhythm class, tempo, and  $F_0$  in language discrimination: implications for the nature of speech rhythm
- Chandrasekaran, Bharath, Jane Hornickel, Erika Skoe, Trent Nicol, & Nina Kraus. 2009. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64. 311–319.
- Corriveau, Kathleen, & Usha Goswami. 2009. Rhythmic motor entrainment in children with speech and language impairment: Tapping to the beat. *Cortex* 45. 119–130.
- Cummins, Fred. 2012. Oscillators and syllables: a cautionary note. *Frontiers in Psychology* 3. 364.
- Frith, Uta. 1998. Editorial: Literally changing the brain. *Brain* 121. 1051–1052.
- Georgiou, George K., Athanassios Protopapas, Timothy Papadopoulos, Christos Skaloumbakas, & Rauno Parrila. 2010. Auditory temporal processing and dyslexia in an orthographically consistent language. *Cortex* 46. 1330–1344.
- Ghitza, Oded. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology* 2. 130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, Oded., & Steven Greenberg. 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.
- Giraud, Anne-Lise, & David Poeppel. 2012. Speech perception from a neurophysiological perspective. In David Poeppel, Tobias Overath, Arthur N. Popper, & Richard R. Fay (eds.), *Human Auditory Cortex*, 225–260. New York, NY: Springer.
- Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105. 251–279.
- Goldstein, Louis, Dani Byrd, & Elliot Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (ed.), *Action to Language via the Mirror Neuron System*, 215–249. New York: Cambridge University Press.
- Gomez-Ramirez, Manuel, Simon P. Kelly, Sophie Molholm, Pejman Sehatpour, Theodore H. Schwartz, & John J. Foxe. 2011. Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *Journal of Neuroscience* 31. 18556–18567.
- Goswami, Usha, Jennifer Thomson, Ulla Richardson, Rhona Stainthorp, Diana Hughes, Stuart Rosen, & Sophie K. Scott. 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.

- Goswami, Usha. 2011. A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences* 15(1). 3–10.
- Goswami, Usha, & Peter Bryant. 1990. *Phonological Skills and Learning to Read*. East Sussex: Erlbaum.
- Goswami Usha, Danielle Gerson, & Luisa Astruc. 2010. Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing* 23. 995–1019.
- Goswami, Usha, Tim Fosker, Martina Huss, Natasha Mead, & Dénes Szücs. 2011b. Rise time and formant transition duration in the discrimination of speech sounds: The ba-wa distinction in developmental dyslexia. *Developmental Science* 14. 34–43.
- Goswami, Usha, Martina Huss, Natasha Mead, Tim Fosker, & John P. Verney. 2012. Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*.
- Goswami, Usha, Natasha Mead, Tim Fosker, Martina Huss, Barnes, L., & Leong, V. Submitted. Impaired perception of syllable stress in children with dyslexia: A longitudinal study.
- Goswami, Usha, H.-L. Sharon Wang, Alicia Cruz, Tim Fosker, Natasha Mead, & Martina Huss. 2011a. Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience* 23. 325–337.
- Greenberg, Steven. 2006. A multi-tier framework for understanding spoken language. In Steven Greenberg & William Ainsworth (eds.), *Listening to Speech: An Auditory Perspective*, 411–434. Mahwah, NJ: LEA.
- Greenberg, Steven, Hannah Carvey, Leah Hitchcock, & Shawn Chang. 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics* 31. 465–485.
- Hämäläinen, Jarmo A., P. H. T. Leppänen, K. Eklund, J. Thomson, U. Richardson, T. K. Guttorm, C. Witton, A-M. Poikkeus, Usha Goswami, & H. Lyytinen. 2009. Common variance in amplitude envelope perception tasks and their impact on phoneme duration perception and reading and spelling in Finnish children with reading disabilities. *Applied Psycholinguistics* 30. 511–530.
- Hämäläinen, Jarmo A., André A. Rupp, Fruzsina Soltész, Denés Szücs, & Usha Goswami. 2012. Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *Neuroimage* 59(3). 2952–2961.
- Hickok Gregory, & David Poeppel. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience* 8. 393–402.
- Houtgast, T., & H. Steeneken. 1985. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America* 77. 1069–1077.

- Huss, Martina, John Verney, Tim Fosker, Natasha Mead, & Usha Goswami. 2011. Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689.
- Kotz, Sonja A., & Michael Schwartz. 2010. Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences* 14. 392–399.
- Krivokapić, J. this volume. Rhythm and convergence between speakers of American and Indian English
- Lakatos, Peter, Gergo Karmos, Ashesh D. Mehta, Istvan Ulbert, & Charles E. Schroeder. 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320. 110–113.
- Lehongre, Katia, Franck Ramus, Nadège Villiermet, Denis Schwartz, & Anne-Lise Giraud. 2011. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* 72. 1080–1090.
- Leong, Victoria 2012. Prosodic rhythm in the speech amplitude envelope : Amplitude modulation phase hierarchies (AMPHs) and AMPH models. Ph.D. dissertation, University of Cambridge.
- Leong Victoria, Jarmo Hämäläinen, Fruzsina Soltész, & Usha Goswami. 2011. Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language* 64. 59–73.
- Leong, Victoria, R. Turner, M. Stone, & U. Goswami. 2011. An amplitude modulation (AM) cascade model for nursery rhyme rhythm. Poster presented at the 13th Rhythm Perception & Production Workshop, Leipzig, Germany.
- Luo, Huan, & David Poeppel. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54. 1001–1010.
- Luo, Huan, Zuxiang Liu, & David Poeppel. 2010. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology* 8. e1000445.
- Munhall, K.G, J. A. Jones, D. E. Callan, T. Kuratate, & E. Vatikiotis-Bateson. 2004. Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science* 15. 133–137.
- Nam Hosung, Louis Goldstein, & Elliot Saltzman. 2009. Self-organization of syllable structure: a coupled oscillator model. In Francois Pellegrino, Egidio Marisco, & Ioana Chitoran (eds.), *Approaches to Phonological Complexity*. Berlin/New York: Mouton de Gruyter.
- Obleser, Jonas, Björn Herrmann, & Molly J. Henry. 2012. Neural oscillations in speech: Don't be enslaved by the envelope. *Frontiers in Human Neuroscience* 6. 250.
- Poelmans, Hanne, Heleen Luts, Maaïke Vandermosten, Bart Boets, Pol Ghesquière, & Jan Wouters. 2011. Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities* 32. 2810–2819.

- Poeppel, David. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' *Speech Communication* 41. 245–255.
- Power, Alan James, Natasha Mead, Lisa Barnes, & Usha Goswami. 2012. Neural entrainment to rhythmically-presented auditory, visual and audio-visual speech in children. *Frontiers in Psychology* 3. 216; doi 10.3389/fpsyg.2012.00216
- Richardson Ulla, Jennifer M. Thomson, Sophie K. Scott, & Usha Goswami. 2004. Auditory processing skills and phonological representation in dyslexic children. *Dyslexia* 10. 215–233.
- Rothermich, Kathrin, Maren Schmidt-Kassow, & Sonja A. Kotz. 2012. Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia* 50(2). 232–244.
- Saltzman, Elliot L., Hosung Nam, Jelena Krivokapic, & Louis Goldstein. 2008. A task-dynamic toolkit for modelling the effects of prosodic structure on articulation. In Plinio A. Barbosa, Sandra Madureira, & César Reis (eds.), *Proceedings of the 4<sup>th</sup> International Conference on Speech Prosody*, 175–184. Campinas, Brazil,.
- Schroeder, Charles E., & Peter Lakatos. 2009. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neuroscience* 32. 9–18.
- Schroeder, Charles E., Peter Lakatos, Yoshinao Kajikawa, Sarah Partan, & Aina Puce. 2008. Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences* 12(3). 106–113
- Scott, Sophie, & Carolyn McGettigan. 2012. Amplitude onsets and spectral energy in perceptual experience. *Frontiers in Psychology* 3. 80.
- Serniclaes, Willy, Sandra Van Heghe, Philippe Mousty, René Carré, & Liliane Sprenger-Charolles. 2004. Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology* 87. 336–361.
- Snowling, Margaret J. 2000. *Dyslexia* (2nd ed.). Oxford, UK: Blackwell Publishers.
- Soltész, F., Szűcs, D., White, S., Leong, V., & Goswami, U. (2013). Differential phase entrainment of delta oscillations in developmental dyslexia in a rhythmic auditory target detection task. *Cerebral Cortex*.
- Stefanics, Gábor, Balázs Hangya, István Hernádi, István Winkler, Peter Lakatos, & István Ulbert. 2010. Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience* 30. 13578–13585.
- Stefanics Gábor, Tim Fosker, Martina Huss, Natasha Mead, Dénes Szűcs, Usha Goswami. 2011. Auditory sensory deficits in developmental dyslexia: a longitudinal ERP study. *Neuroimage* 57(3). 723–732.
- Surányi, Zsuzsanna, Valéria Csépe, Ulla Richardson, Jennifer M. Thomson, Ferenc Honbolygó, & Usha Goswami. 2009. Sensitivity to rhythmic parameters in dyslexic children: a comparison of Hungarian and English. *Reading and Writing* 22. 41–56.



- Thomson Jennifer M., Ben Fryer, James Maltby, & Usha Goswami. 2006. Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading* 29. 334–348.
- Thomson Jennifer M., & Usha Goswami. 2008. Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris* 102 120–129.
- Tilsen, Sam. 2009. Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive Science* 33. 839–879.
- Tilsen, Sam, & Amalia Arvaniti. Submitted. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages.
- Whalley, Karen, & Julie Hansen. 2006. The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading* 29. 288–303.
- Wimmer, Heinz. 1993. Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics* 14. 1–33.
- Yehia, Hani, Takaaki Kuratate, & Eric Vatikiotis-Bateson. 2002. Linking facial animation, head motion and speech acoustics. *Journal of Phonetics* 30. 555–568.
- Ziegler, Johannes C., & Usha Goswami. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin* 131. 3–29.