---- Dissertation Precis for Glushko Prize in Cognitive Science (Dr Victoria Leong) ----

INTRODUCTION

Speech Rhythm Cognition : A Multi-Disciplinary Account

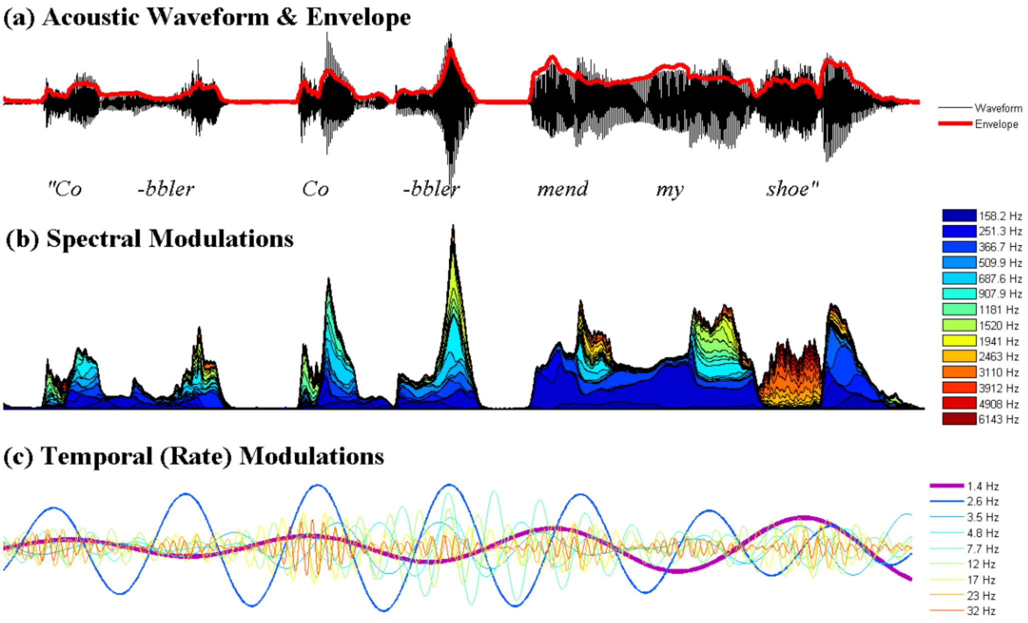
Rhythm is our perceptual experience of the slow-varying temporal structure of the acoustic signal. The rhythm (or prosody) of speech is studied under many guises across the different disciplines of cognitive science. In my PhD thesis, I attempt to provide a contemporary, cross-disciplinary account of speech rhythm cognition, synthesising perspectives from psychology, linguistics, neuroscience, and acoustics. The result is a new class of Amplitude Modulation Phase Hierarchy (AMPH) computational models that make use of the emergent temporal statistics of the speech signal to detect syllables and decode rhythm patterns, as infants might, providing a simulation of how infants' language acquisition is 'boot-strapped' from the speech signal. These cognitive models are neuro-plausible, psychologically-validated, and computationally-efficient, and may even be considered as artificially-intelligent. They serve dual functions as theoretical frameworks for understanding language development, and practical tools for the experimental analysis of speech data. In this precis, I first provide a general introduction to speech rhythm as it is studied in each cognitive discipline, before giving a chapter-by-chapter summary of my thesis.

Psychology & Education. In spoken English, rhythm manifests as the alternation of strong (S) and weak (w) syllables, as in the phrase "HA-ppy BIRTH-day" (S-w S-w). These Strong-weak rhythm patterns play a crucial role in 'boot-strapping' early language acquisition for English infants [1,2]. By the age of 7.5 months, infants already begin to use the strong-weak (S-w) rhythm pattern as a physical template for segmenting words from continuous speech [3]. Thus, rhythm patterns form an integral part of infants' developing mental representations of speech sounds, or, 'phonology' [4]. To support language acquisition in young learners, adults spontaneously exaggerate the rhythm and prosody of their speech when addressing infants or children [5,6], thereby highlighting word and phrase boundaries to the listener. Developmentally, children with good rhythm and prosodic awareness typically go on to develop good reading skills [7], while poor rhythm and prosodic skills are often found in children with dyslexia, who struggle to learn to read due to their impoverished phonological representations [8,9]. Thus, the study of speech rhythm has strong psychological and educational importance.

Linguistics. Historically, the study of prosodic rhythm has been the domain of linguists [10-16] and phoneticians [17-19]. In the early years, prominent linguists [10-11] established the notion of language 'rhythm classes', which claimed that languages in the world differ rhythmically according to the particular phonological level at which durational isochrony manifests. Languages like Spanish and

Italian are rhythmically 'syllable-timed' because syllables are thought to be uttered at regular intervals, giving these languages a 'machine-gun'-like feel. Conversely, English and Dutch are 'stress-timed' because prosodic *stress* is thought to occur at regular intervals in these languages, while syllable intervals may compress or stretch in accommodation. Despite the intuitive appeal of this notion, researchers have repeatedly failed to find convincing evidence for durational isochrony at either the stress or the syllable level [20-22], spawning a reactionary generation of 'rhythm-metrics' predicated on the notion of *segmental* (phoneme) durational variation [17-19]. Rhythm-metric methods primarily capture distributional differences in the time intervals between successive consonants or vowels in different languages. These methods impose a formal phonetic framework on the acoustic signal which the signal does not actually possess [23-24]. Thus, the psychological model underlying rhythm-metrics is relevant to expert listeners who already possess segmental knowledge, but cannot explain how even newborn infants are able to discern rhythm in the speech signal [25]. Infants' 'innate' sensitivity to rhythm must therefore depend on some emergent temporal property of the acoustic signal. Other articulatory [26], cognitive [27-28], and physiological models [29-30] have attempted to pinpoint what this temporal property (or properties) might be, but the question of what underlies speech rhythm remains unanswered.

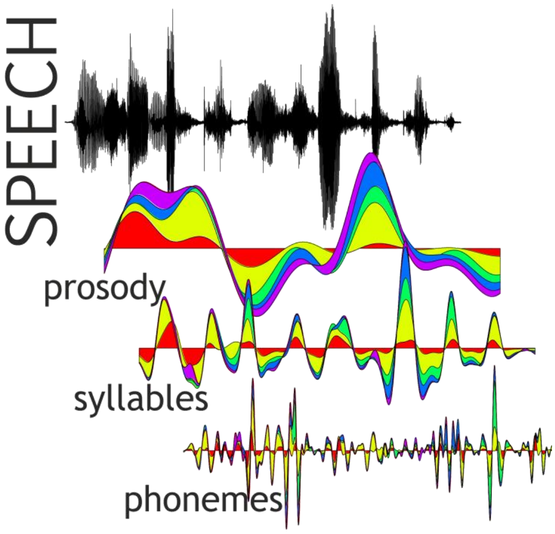
Figure 1(a). Example of the acoustic waveform of the speech signal, with its amplitude envelope overlaid as a bold red line. 1(b). Example of the different spectral envelopes for different acoustic frequencies in the same sentence, where each spectral envelope is shown as a different color. The envelopes are stacked vertically, with higher energy corresponding to a greater area underneath each curve. 1(c) Example of the different modulation rates present within a single envelope, notice that the dominant modulation rate in this sample is ~2.6 Hz, corresponding to the syllable rate of the utterance.



Acoustics & Engineering. Meanwhile, in the field of speech acoustics and engineering, where a deep understanding of the temporal structure of speech exists, researchers have largely ignored the study of rhythm, except where it pertains to speech *intelligibility* or *recognition* (although see [31-35]). Nonetheless, a crucial and surprising finding by this field is that speech intelligibility depends heavily on the slow-varying *amplitude envelope* of the signal [36], and in particular on amplitude modulation (AM) rates as slow as 4-16 Hz [37-38], which relate primarily to the syllable patterning in speech [34-35]. At the simplest level, the amplitude envelope can be thought of as the 'outline' of the acoustic waveform, reflecting slow-varying fluctuations in the intensity of the signal over time (see Figure 1a). Real speech, however, is more complex because different acoustic frequencies can each have different envelope patterns (Figure 1b). Furthermore, each envelope contains modulations that span a range of different rates, forming a 'modulation spectrum'. Within this spectrum, the modulation strength of some rates (e.g. the syllable rate) is more dominant than other rates (Figure 1c). Thus, if one were to divide the speech signal into '*a'* acoustic frequency bands, there would be *'a'* different spectral envelopes for the same sentence. Moreover, each of these *'a'* spectral envelopes will contain amplitude modulation at '*b*' different *rates*, resulting in an *a x b* spectro-temporal representation of the envelope. Embedded within this spectro-temporal envelope are important cues to speech intelligibility, and also to prosodic rhythm and stress.

Neuroscience. The importance of envelope modulation patterns for speech perception has been captured in recent *neural* models that propose a relationship between speech rhythms and *brain rhythms* [39-42]. For example, Poeppel and colleagues [40] argue that neuronal oscillations in the 'theta' (3-7 Hz) and 'gamma' (25-40 Hz) range track syllable and phoneme patterns in speech respectively, by 'phase-locking' or 'entraining' to the envelope modulation patterns at these two rates, concurrently sampling the speech signal at these two timescales. If human listeners are relying primarily on acoustic modulation patterns from the envelope to understand speech, then the envelope is also likely to be an important source of *prosodic* cues to rhythm and stress.

Figure 2. Stylistic representation of the AM hierarchy with 3 dominant modulation rates corresponding to the timescales for prosodic stress, syllables and phonemes respectively. Please refer to Chapter 4 of the thesis for full details.



A Cross-Disciplinary Synthesis. In this thesis, two 'Amplitude Modulation Phase Hierarchy (AMPH)' models are developed to 'mine' the envelope's rich spectro-temporal structure for cues to speech rhythm and stress. Inspired by Poeppel's multi-timescale model of speech processing, the

AMPH models are based on major AM rates in the envelope that are nested together as an 'AM hierarchy'. In a statistical analysis of the modulation structure of speech (Chapter 4), the 3-tier AM hierarchy that emerges (statistically) bears an astounding symmetry to hierarchically-nested neuronal oscillations in the brain [40-41], as well as to hierarchically-nested linguistic prosodic structure [12-16], with timescales corresponding to prosodic stress, syllables and phonemes respectively. Thus, the AMPH models provide a computational account of how hierarchical ***brain*** rhythms might extract hierarchical ***linguistic*** structure from the emergent hierarchical ***acoustic*** modulation structure of the speech signal. Furthermore, the AMPH models also function as *psychological* models of how infants bootstrap their language learning from the statistics in the speech signal (see [1-2]). Like the newborn infant, the AMPH models are capable of detecting prosodic rhythm patterns solely from the acoustic information in the speech signal, without the need for any prior manual speech labelling or phonetic segmentation. Finally, I use the AMPH models to address questions of educational importance, such as the assessment of speech rhythm perception and production in *dyslexia*, and the characterisation of rhythm in *child-directed speech*. As a testament to the highly multi-disciplinary nature of my work, portions of this thesis have been published or are currently under consideration by journals in psychology, linguistics, audiology and neuroscience (see References for details). What follows is a chapter-by-chapter summary of my thesis, structured in 4 parts :

Part I : Introduction & Literature Review (*Chapter 1*)
Part II : The Amplitude Modulation Phase Hierarchy (AMPH) Model *(Chapters 2-3)*
Part III : The New Spectral AMPH Model *(Chapters 4-6)*
Part IV : Using the S-AMPH Model in Data Analysis *(Chapters 7-8)*

PART I

Chapter 1

In this Introduction, I provide a panoramic survey of disciplines that have had historical, conceptual or methodological significance in the study of speech rhythm. I take, by turn, the perspective of the developmental psychologist (1.1), the linguist phonetician (1.2-1.3), the speech engineer (1.4-1.8), the cognitive neuroscientist (1.9), and the educator (1.11-1.12). In each personification, I explain the unique epistemology and concerns that have motivated the study of rhythm, and the attendant achievements and limitations of each field. It quickly becomes apparent to the reader that rhythm has had a long tradition of study within many of these fields. Yet, very little cross-disciplinary dialogue has occurred, perhaps because speech rhythm is studied under so many different guises and labels: prosody, language classes, perceptual-centres, the amplitude envelope, syllable detection, neuronal oscillations, nursery rhymes, etc. I argue for a unifying, cross-disciplinary account of speech rhythm that leverages on the insights and advances of all these fields. What follows is my attempt to deliver just such an account of speech rhythm.

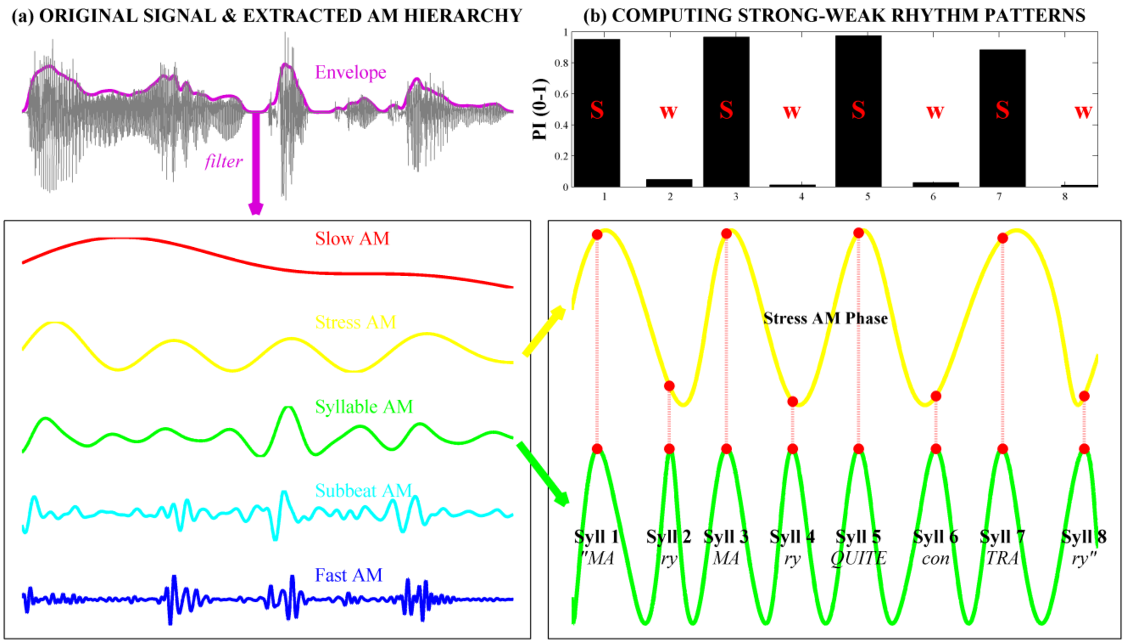
PART II

Here, the first Amplitude Modulation Phase Hierarchy (AMPH) model for speech rhythm is introduced (Chapter 2). This original AMPH model is derived theoretically, on the basis of previous literature. The Stress Phase Code is introduced, which is an algorithm for computing 'Strong-weak' syllable stress patterns using AM statistics in the model. Finally, in Chapter 3, the core assumptions of the AMPH model are tested in a tone-vocoding experiment with human listeners.

Chapter 2

The AMPH model represents speech AMs at different modulation rates as an ***AM hierarchy***. The model assumes that speech contains amplitude modulation on certain key timescales, corresponding to the typical duration of major phonological units such as prosodic stress 'feet' (motifs of strong and weak syllables), syllables, phonemes, etc. Each of these phonological tiers is assumed to occupy a separate AM tier, and individual AM cycles within each tier can be taken to represent individual phonological units. For example, Figure 3 shows the AM hierarchy for the nursery rhyme "Mary Mary quite contrary", which has a hierarchical prosodic structure of 8 syllables nested within 4 stress feet, each of which have a 'S-w', or trochaic motif.

Figure 3. AM hierarchy for the rhyme 'Mary Mary quite contrary'. AM cycles represent phonological units: Notice that the 8 Syllable AM cycles correspond to the 8 uttered syllables within the sentence. Oscillatory phase relationships determine prosodic prominence : Notice that the Stress Phase Code correctly predicts the trochaic pattern of alternating 'S' (strong) and 'w' (weak) syllables.

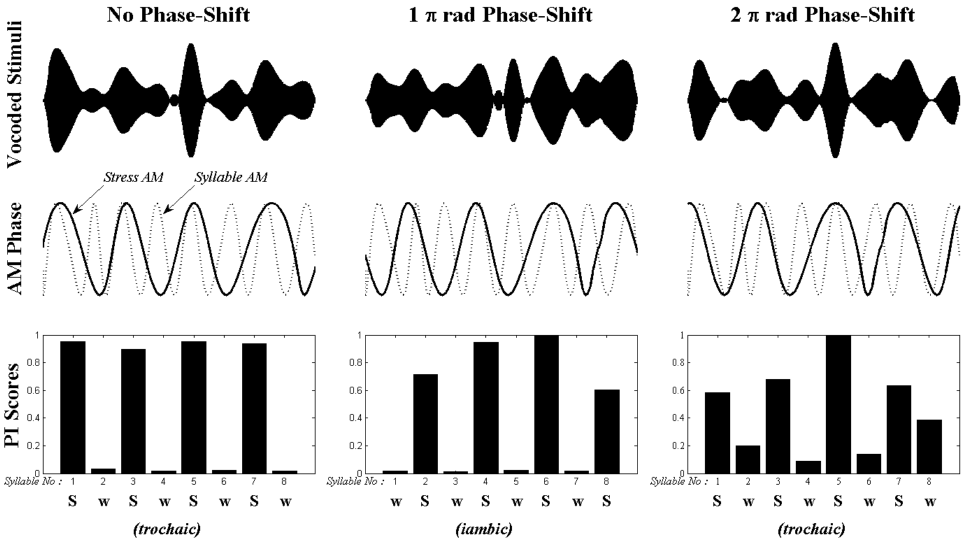


Within this AM hierarchy representation, 'Strong (S)' - 'weak' (w) prosodic patterns of relative prominence are captured as oscillatory phase relationships between adjacent (nested) AM tiers. Thus, the pattern of ***syllable stress*** and rhythm (the key variable of interest) is specified by the phase relationship between the 'Syllable' AM (~5 Hz) and 'Stress' AM (~2 Hz) tiers, as shown in Figure 3b. In this example, the Syllable AM peak corresponding to the first syllable "Ma" occurs near the oscillatory peak of the parent Stress AM tier (vertical dotted line) whereas the second syllable "-ry" occurs near the Stress AM oscillatory trough. This phase pattern corresponds to a 'S-w' motif for the word "MA-ry". The Stress Phase Code is a computational algorithm which transforms these circular oscillatory phase relationships into a linear metric of prosodic prominence (Section 2.5.2, p.72). I also explain how ***poetic meter*** might relate to specific *n:m* phase-locking ratios between the Stress AM and the Syllable AM (Section 2.5.1, p. 68), and explore potential segmentation schemes that could emerge 'for free' from this basic AMPH model (Section 2.6, p. 77).

Chapter 3

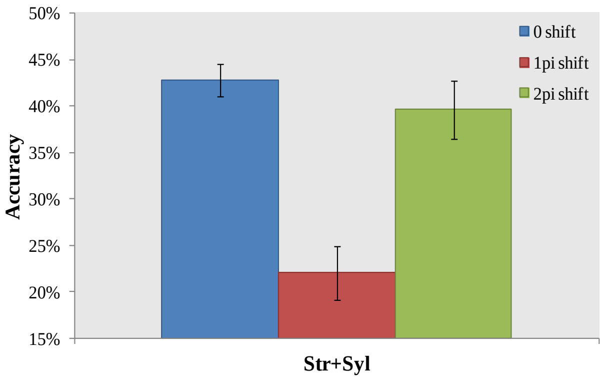
Next, a rhythm perception experiment was conducted to assess the *psychological* validity of the AMPH model. A primary assumption is that Strong-weak syllable stress patterns arise from the phase-relationship between 'Syllable'-rate and 'Stress'-rate AMs in the envelope. If so, incremental phase displacements of the Stress-Syllable AM relationship should cause *circular* perturbations in the perceived syllable stress pattern. That is, phase-shifts in the Stress-Syllable AM relationship of up to 1π radians (half a cycle) should move participants' perception of a given syllable toward the opposite prominence (e.g. from strong to weak), but larger shifts of up to 2 π radians (a full cycle) should bring perception back to the original value (e.g. strong). Thus, when phase-shifted by 1π radians, an originally trochaic (S-w) sentence should now be perceived to have an iambic (w-S) rhythm, and vice versa. By contrast, when phase-shifted by 2π radians, sentences should maintain their original rhythm pattern. This phase-shift prediction was tested in a rhythm perception experiment using tone-vocoded nursery rhyme sentences that had either a trochaic or an iambic rhythm pattern. In the vocoding process, AMs within the amplitude envelope were extracted and used to modulate a pure sine tone, while the original fine structure of the sentence was discarded. This process made the AM patterns audible, and effectively isolated rhythm while the sentence itself remained unintelligible (i.e. sounding like a sequence of rhythmic pulses). To control for methodological artifacts, I used two fundamentally different methods to extract AMs from the envelope, generating two parallel sets of vocoded stimuli. The first method was a traditional modulation filterbank [MFB] (Section 3.1.4.1, p.84), while the second method utilised Bayesian Probabilistic Amplitude Demodulation (PAD, [43], see Section 1.7.2, p. 33). Figure 4 provides an example of the normal and phase-shifted MFB stimuli used in the experiment. As shown in Figure 5, the results of the experiment indicated that participants did indeed base their rhythm perception upon the phase relationship between the Stress AM and the Syllable AM, producing the predicted circular pattern of responding with increasing phase-shifts.

Figure 4. Illustration of the effect of phase-shifting on the rhythm pattern of 'Mary Mary'.(Top row): Tone-vocoded stimuli used in the experiment. (Middle row): Corresponding Stress (bold) and Syllable (dotted) AM phase patterns. Phase values are projected onto a cosine function for visualisation purposes. Only Stress AMs were phase-shifted while Syllable AMs were held constant. (Bottom row) : Stress Phase Code prominence index (PI) scores of syllables. Strong syllables ('S') have a prominence value of >0.5, weak syllables ('w') have a prominence value of <0.5.



The first AMPH model was a simple theory-driven model of how the speech rhythm percept might arise from AM patterns in the envelope. It made psychologically-accurate predictions as to how listeners would respond to perturbations in the Stress-Syllable AM phase-relationship. However, the AMPH model (1) did not take into account spectral differences in the envelope across acoustic frequencies and (2) was developed and tested exclusively using metronome-timed (durationally-regular) speech. These short-comings are addressed in Part III with an improved model.

Figure 5. Performance of participants in the tone-vocoder task for non-phase-shifted and phase-shifted Stress+Syllable AMs.



PART III

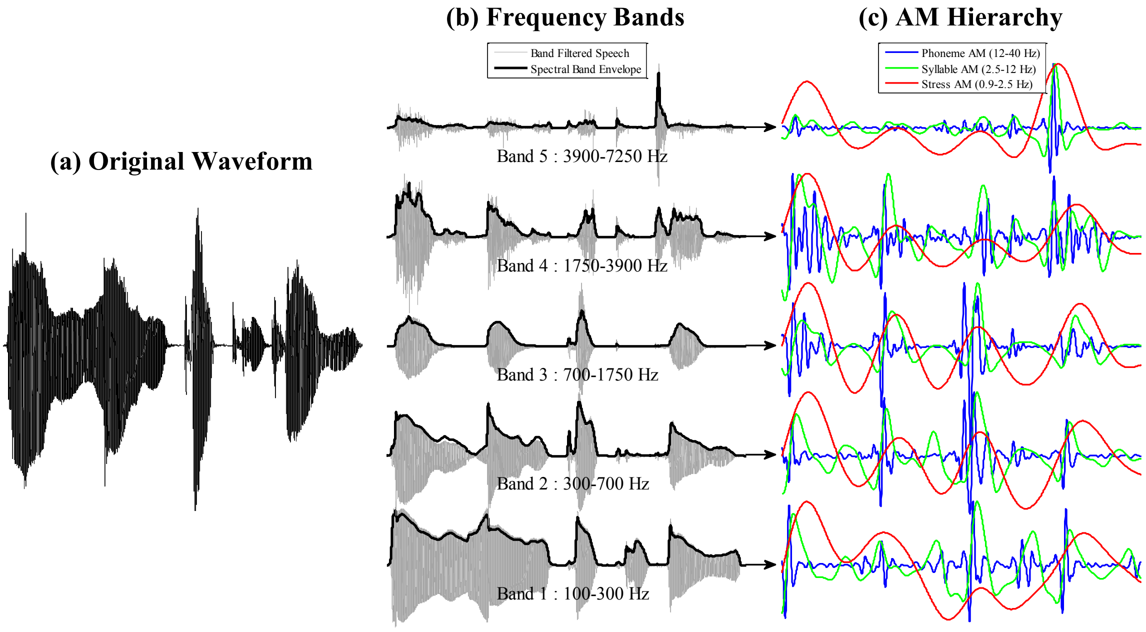
Here, I introduce a new Spectral Amplitude Modulation Phase Hierarchy (S-AMPH) model, which is based on a revised AM hierarchy that is designed 'ground-up' from the actual modulation statistics of the speech signal (rather than relying on theoretical assumptions). The new model also incorporates a ***spectral*** dimension, to take into account speech modulation patterns at different acoustic frequencies (i.e. spectral envelopes), as well as at different temporal rates (i.e. the AM hierarchy). To generate this new spectro-temporal representation of the envelope, a principal component analysis (PCA) procedure is employed in Chapter 4. In line with this new spectro-temporal representation, new prosodic indices for computing 'Strong-weak' stress patterns are developed in Chapter 5. Finally, the original AMPH and new S-AMPH models are functionally compared in automatic syllable detection and prosodic stress transcription exercises (Chapter 6).

Chapter 4

The aim of the PCA process is to derive a ***low-dimensional***, ***data-driven*** representation of the dominant spectral and temporal modulation structure within the speech envelope. This new spectro-temporal representation then goes on to form the basis for the new S-AMPH model. The dataset for the PCA analysis is a new and larger corpus of naturally-produced child-directed speech. In the first step, highly-detailed spectral and modulation rate representations of the data are extracted, simulating the fine frequency decomposition that occurs at the human cochlear. PCA is then applied to these 'high-dimensional' representations, with the aim of identifying major patterns of covariation within these detailed representations that signify the presence of dominant 'bands' of modulation in the spectral or rate domains. The PCA results suggest that the optimal spectro-temporal architecture for the speech signal is comprised of 5 major acoustic frequency bands, whose spectral envelopes are further decomposed into 3 major modulation rates (i.e. a 3-tier AM hierarchy). This 5 (spectral) x 3 (modulation rate) representation is shown in Figure 6. Compared to the AMPH model, this new representation of the envelope is more complex in the spectral domain (5 spectral bands instead of 1), and less complex in the AM rate domain (3 modulation rates or AM tiers instead of 5). However, in both representations, the Syllable rate (~5 Hz) emerged as a dominant timescale, as did the Stress rate (~2 Hz). The endorsement of both Stress-rate and Syllable-rate modulations as major components of speech modulation structure lends confidence to later computations of prosodic stress, which depend critically on these two AM rates.

It is worth noting that the 3 dominant modulation rates (or timescales) that emerge from the PCA analysis bear a striking and biologically-fortuitous correspondence to 3 important timescales of neuronal oscillatory activity in the brain : 'delta', 'theta' and 'gamma' oscillations. According to multi-timescale models of speech processing [40-41], these 3 bands of neural oscillatory activity are implicated in the temporal sampling of phonological information on equivalent timescales: i.e. delta (stress patterns, ~2 Hz), theta (syllables, ~5 Hz) and gamma (phonemes, ~25 Hz). It is in reference these classic phonological timescales that the 3 AM tiers in the S-AMPH model are named the 'Stress AM', 'Syllable AM' and 'Phoneme AM' respectively.

Figure 6. Signal-processing stages in the S-AMPH model. (a) Original acoustic waveform of the spoken sentence "Mary Mary quite contrary" . (b) In the S-AMPH model, the original speech signal is first filtered into 5 frequency bands, and the Hilbert envelope is computed for each frequency band. (c) A 3-tier AM hierarchy is then extracted from the envelopes of each frequency band. The resulting 'Stress' (0.9-2.5 Hz), 'Syllable' (2.5-12 Hz) and 'Phoneme' (12-40 Hz) AMs are shown overlaid in different colours. These correspond to prosodic stress patterns, syllable patterns and phoneme patterns respectively. This results in a 5 (frequency band) x 3 (AM hierarchy) spectro-temporal representation of the speech amplitude envelope.



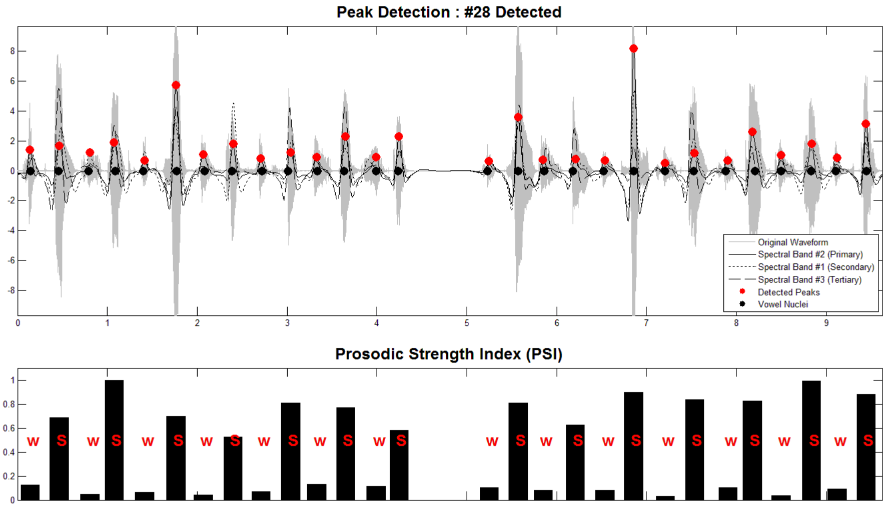
Chapter 5

As proposed in the AMPH model, the prosodic strength of a syllable ('Strong' or 'weak') is related to the phase of the Stress AM at which it occurs. Chapter 5 outlines new algorithms for syllable detection, which take advantage of the enhanced spectral complexity of the S-AMPH model. The chapter also discusses a modified Prosodic Strength Index (PSI), which refines the phase computation of the original Stress Phase Code (Chapter 2), in order to reflect the *actual* Stress-phase distribution of syllables in naturally-produced speech.

Chapter 6

If the modulation patterns contained within the 'Syllable' and 'Stress' tiers of the AM hierarchy reflect real syllables and prosodic stress in an utterance, it follows that one should be able to use these AMs to automatically locate syllables and 'decode' stress patterns in speech. Accordingly, the S-AMPH and AMPH models were functionally evaluated to see if their 'Syllable' and 'Stress' AMs could successfully be used for (1) automatic syllable detection, and (2) automatic prosodic stress transcription. For this exercise, two manually-labelled speech corpora were used. In one corpus, the speakers produced metronome-timed (rhythmically-regular) speech, whereas in the other corpus, speech was freely-produced. Figure 7 shows an example of the performance of the S-AMPH model in both automatic syllable detection (black dot = actual syllable location, red dot = automatically-detected syllable) as well as prosodic strength assessment. Syllable detection accuracy was assessed by using peaks in the Syllable AM to identify syllable vowel nuclei in the utterance. For this measure, both AMPH and S-AMPH models performed very well for metronome-timed speech, with the AMPH model registering ~94% accuracy and the S-AMPH model registering ~97% accuracy. For freely-produced speech, the performance of both models decreased. However, the S-AMPH model still showed a distinctly superior performance, registering >80% accuracy as compared to ~60% accuracy achieved by the AMPH model. Therefore, the multi-band spectral complexity of the S-AMPH model made it better able to handle the challenges of syllable detection in natural speech.

Figure 7. Example of an Iambic (w-S) patterned sentence, syllable peaks detected (red dots) and actual location of vowel nuclei (black dots). (bottom) Assignment of syllable prosodic strength using the PSI. Individual bars correspond to syllables, and the height of each bar shows the PSI value. Syllables with a PSI value of ≥0.4 were considered 'Strong (S)', syllables with a PSI value of <0.4 were considered 'weak (w)'. In this example, all 28 syllables were correctly assigned as 'Strong (S)' or 'weak (w)'.



In the prosodic stress transcription exercise, both models were again highly accurate for metronome-timed speech, yielding accuracies of ~93% (AMPH) and ~94% (S-AMPH). However, for freely-produced speech, performance for both models dropped to ~65% (AMPH) and ~70% (S-AMPH), with no statistical difference in the performance of the two models. Although the performance accuracy for both models might appear low, it is not dissimilar to the accuracy achieved by other models that are specifically designed for the purpose of automatic stress transcription [44]. Thus, the evaluation exercise provides empirical support for the assumption that the

Stress and Syllable AM patterns captured within the AMPH models correspond reasonably well to actual syllable and stress patterns in speech. If so, these models could be used to assess rhythmic differences in speech data, as discussed next.

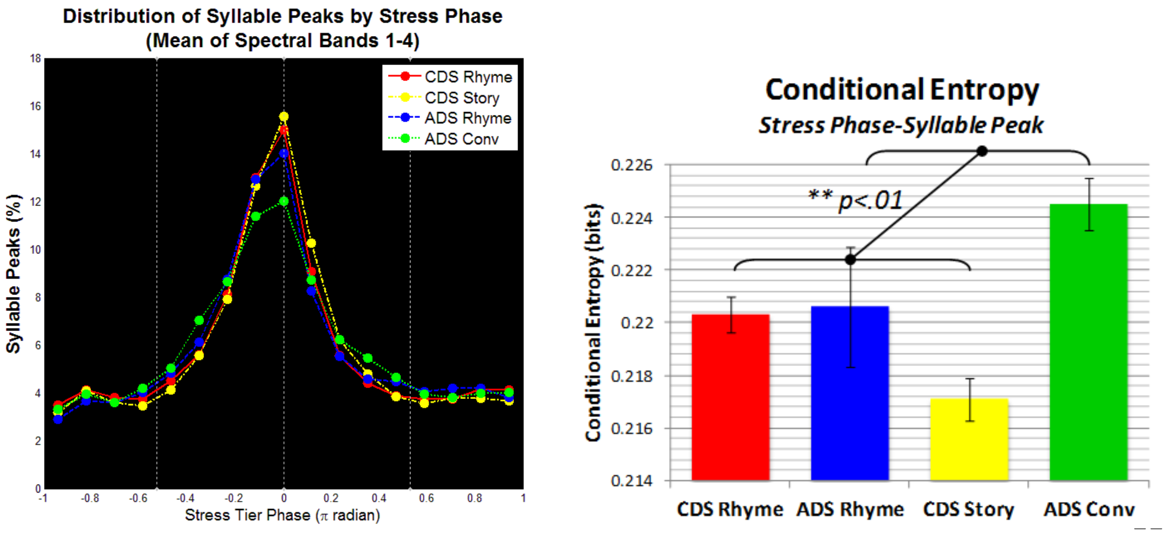
PART IV

In Part IV of this thesis, the S-AMPH model is used as a speech rhythm analysis tool in two real-world psychological studies on child-directed speech and dyslexia.

Chapter 7

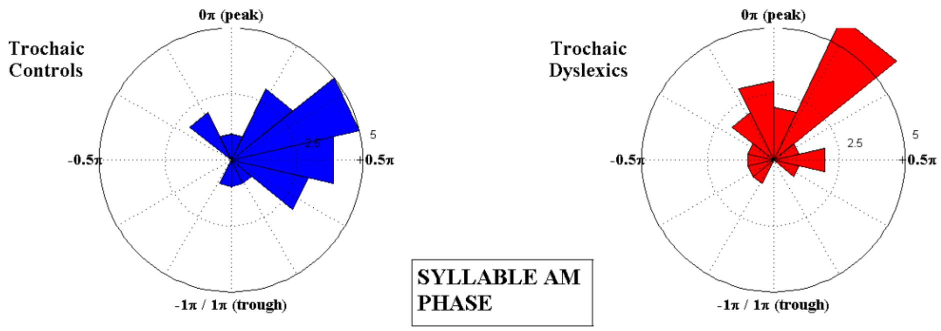
Child-directed speech (CDS) is prosodically-enhanced to accommodate the needs of the child listener. Here, the S-AMPH model is used to assess the rhythmic changes accompanying this prosodic enhancement. The key finding is that CDS is more rhythmically-regular than ADS across multiple timescales (stress, syllable and phoneme). CDS also shows a more tightly phase-locked AM hierarchical structure that indicates stronger prosodic patterning, and is associated with lowered entropy in the signal (see Figure 8). Surprisingly, the rhythmic patterning found in CDS storybook readings (e.g. 'Goldilocks') is as strong as that of nursery rhymes which have a regular poetic meter. This suggests that adults spontaneously enhance the rhythmicity of their speech when they are reading to children, even if the material itself does not have a clear poetic meter. The rhythmic enhancements in CDS are consistent with the exaggeration of word and phrase boundaries in the acoustic signal, which could help the child to segment words from the speech stream more easily.

Figure 8. (Left) Hierarchical distribution of peaks for each modulator tier with respect to the phase of the upper tier. The left plot shows the distribution of Syllable peaks with respect to Stress phase. The right plot shows the distribution of Phoneme peaks with respect to Syllable phase. The distributions shown are the mean distributions across of spectral bands 1-4. (Right) Corresponding conditional entropy (CE) scores for the distribution pattern of each speech corpus. Distributions with higher kurtosis have a lower entropy while distributions with lower kurtosis have a higher entropy. Errorbars show the standard error across 6 speakers.



Previous work by Goswami and colleagues [45, see Appendix 1of thesis] found that adults with dyslexia have poorer prosodic sensitivity to syllable stress patterns in words (e.g. differentiating between "MI-li-ta-ry" [S-w-w-w] and "mi-LI-ta-ry" [w-S-w-w]). This syllable stress deficit was related to poorer psycho-acoustic sensitivity to amplitude changes in non-speech sounds (sine tones). Since syllable stress patterns are transmitted by slow-varying AMs in the speech signal, the logical next step was to test whether dyslexics also showed impaired perception and production of *speech* AM patterns. In Chapter 8, I describe 3 speech rhythm tasks that are performed by dyslexic and non-dyslexic adults, assessing speech AM *perception*, speech AM *entrainment* (tapping to speech AMs) and speech AM *production* respectively. In all 3 tasks, dyslexics consistently showed disruptions to syllable-level timing. For example, Figure 9 shows that in the tapping task (Section 8.3.3, p. 225), dyslexics entrained to an earlier phase of the Syllable AM cycle as compared to controls. Moreover, individual differences in syllable-timing (both in perception and production) were strongly related to participants' phonological and reading skills. The S-AMPH indices uncovered differences between dyslexics and controls that were not evident from conventional analysis. These deficits in syllable timing and temporal organisation had been predicted in theory [46], but had been difficult to uncover using conventional methods of speech analysis. Therefore, the S-AMPH proved to be a useful analytical tool to complement traditional methods of speech analysis.

Figure 9. Please see Section 8.3.3, p. 225 for the task description. Compass phase plots of the distribution of Syllable AM taps for controls (left) and dyslexics (right). The top of the plot corresponds to the oscillatory peak, the bottom corresponds to the trough. Phase values increase in a clockwise direction. The length of radial spokes indicates the number of observations within each phase bin (with concentric circles indicating 2.5 and 5 observations).The plots show that dyslexics tap at an earlier Syllable AM phase as compared to controls.



SUMMARY & CONCLUSION

In this thesis, I investigate speech rhythm from a multi-disciplinary perspective, aiming for a unifying account that transcends traditional boundaries. The resulting AMPH and S-AMPH models are signal-grounded, neuro-plausible, psychologically-validated, computationally-efficient, and ecologically-relevant to language development in the real world. From the perspective of a developmental psychologist, the AMPH models act as sensory-acoustic, computational accounts of the speech modulation statistics that infants might use to 'boot-strap' their early language acquisition. From the linguistic perspective, these amplitude-based models provide a useful tool for amplitude-based measurement of speech rhythm, complementing the previous emphasis on durational metrics. As neural-grounded computational accounts, these models provide a deep description of how oscillatory mechanisms in the brain could engage with speech spectro-temporal structure to extract prosodic structure. Finally, from the educational perspective, the AMPH models are useful speech analysis tools that can be readily applied to address issues of interest to parents and teachers. Therefore, this thesis truly unites multiple cognitive disciplines, and is of relevance to a range of audiences.

(3830 words excluding figures)

References

[1] Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the art. In E. Wanner, & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). Cambridge, UK: Cambridge University Press.

[2] Morgan, J. & K. Demuth. 1996. Signal to syntax: An overview. In J. Morgan and K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates. pp. 1-22.

[3] Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.

[4] Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, 105, 376-385.

[5] Fernald, A. (1989). Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, 60, 1497-1510.

[6] Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104-113.

[7] Whalley, K, & Hansen, J. (2006). The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading*, 29, 288-303.

[8] Wood, C., & Terrell, C. (1998). Poor readers' ability to detect speech rhythm and perceive rapid speech. *The British Journal of Developmental Psychology*, 16(3), 397-408.

[9] Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S.K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99, 10911–10916.

[10] Abercrombie, D. (1967): *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

[11] Pike, P. (1945). *The intonation of American English*. Ann Arbor: University of Michigan.

[12] Selkirk, E.O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563-605.

[13] Selkirk, E.O. (1984). *Phonology and syntax. the relation between sound and structure.* Cambridge, Ma.: MIT Press.

[14] Selkirk, E.O. (1986). On derived domains in sentence phonology. *Phonology Yearbook* 3:371–405.

[15] Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249-336.

[16] Hayes, B. (1995). *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.

[17] Ramus, F., Nespor, I., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

[18] Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (eds.). *Laboratory phonology* (Vol. 7, pp. 515– 546). Berlin: Mouton de Gruyter.

[19] Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. *Proceedings of the International Congress of Phonetics Science*. (pp.471-474). Barcelona.

[20] Dauer, R. (1983). Stress-timing and syllable timing revisited. *Journal of Phonetics*, 11, 51-62.

[21] Roach, P.J. (1982). On the distinction between "stress-timed" and "syllable-timed" languages, in D.Crystal (Ed.) *Linguistic Controversies*, pp. 73-79. London, Edward Arnold.

[22] Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46-63.

[23] Port, R. F., & Leary, A. (2005). Against formal phonology. *Language*, 85, 927-964

[24] Huckvale, M. (1997) . 10 things engineers have discovered about speech recognition. *NATO ASI Workshop on Speech Pattern Processing*. pp. 1-5.

[25] Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24 , 756–766.

[26] Barbosa, P.A. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. *In Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, pages 163-166.

[27] Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26, 145–171.

[28] Port, R. (2003) Meter and speech. *Journal of Phonetics*, 31, 599-61.

[29] Todd, N.P.M. (1994). The auditory “primal sketch”: a multiscale model of rhythmic grouping. *Journal of New Music Research*, 23, 25–70.

[30] Lee, C., & Todd, N. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory ‘primal sketch’ to two multi-language corpora. *Cognition*, 93, 225-254.

[31] Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 26, 138.

[32] Fry, D. B. (1958). Experiments in the perception of stress, *Language and Speech*, 1, 126–152.

[33] Bolinger, D. (1958). A theory of the pitch accent in English, Word: Journal of the International Linguisic Association 7, pp. 199–210, reprinted in D. Bolinger, *Forms of English: accent, morpheme, order*. Harvard University Press, Cambridge, MA.

[34] Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 29, 159–176.

[35] Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech - a syllable-centric perspective. *Journal of Phonetics*, 31, 465-485.

[36] Shannon R.V., Zeng, F-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.

[37] Drullman, R., Festen, J.M., & Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*, 95, 1053-1064.

[38] Drullman, R., Festen, J.M., & Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America*, 95, 2670-2680.

[39] Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41, 245-255.

[40] Giraud, A.L. & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511-517.

[41] Ghitza, O. (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

[42] Peelle, J.E., Davis, M.H. (2012) Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Language Sciences*, 3, 320.

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

[44] Silipo, R., & Greenberg, S. (1999). Automatic transcription of prosodic stress for spontaneous English discourse. "The Phonetics of Spontaneous Speech," *ICPhS-99*, San Francisco, CA, August.

[45] Leong, V., Hamalainen, J., Soltesz, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, 64, 59-73.

[46] Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15, 1 3-10.

Publications arising from PhD thesis :

Goswami, U, & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4 (1), 67-92.

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, 308, 141-161.

Leong, V., & Goswami, U. (revise-resubmit). Infant NEuro-Acoustic emergent Phonology (iNEAP) : 'Booting-up' phonology in the brain from acoustic temporal structure. *Psychological Review.*

Leong, V., & Goswami, U. (under review). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*

Leong, V., & Goswami, U. (in preparation). The modulation statistics of child-directed speech.

Leong, V., Stone, M., Turner, R. & Goswami, U. (in revision). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America*

Leong, V., Turner, R., Stone, M., & Goswami, U. (in revision). A new amplitude-based metric for nursery rhyme rhythm : The Amplitude Modulation Phase Hierarchy (AMPH).