



# OPEN MIND

Discoveries in  
Cognitive Science

an open access  journal

## The Temporal Modulation Structure of Infant-Directed Speech

Victoria Leong<sup>1</sup>, Marina Kalashnikova<sup>2</sup>, Denis Burnham<sup>2</sup>, and Usha Goswami<sup>1</sup>

<sup>1</sup>Centre for Neuroscience in Education, Department of Psychology, University of Cambridge

<sup>2</sup>The MARCS Institute for Brain, Behaviour and Development, Western Sydney University

**Keywords:** infant-directed speech, phase, oscillations, amplitude modulation

### ABSTRACT

The temporal modulation structure of adult-directed speech (ADS) is thought to be encoded by neuronal oscillations in the auditory cortex that fluctuate at different temporal rates. Oscillatory activity is thought to phase-align to amplitude modulations in speech at corresponding rates, thereby supporting parsing of the signal into linguistically relevant units. The temporal modulation structure of infant-directed speech (IDS) is unexplored. Here we compare the amplitude modulation (AM) structure of IDS recorded from mothers speaking, over three occasions, to their 7-, 9-, and 11-month-old infants, and the same mothers speaking ADS. Analysis of the modulation spectrum in each case revealed that modulation energy in the theta band was significantly greater in ADS than in IDS, whereas in the delta band, modulation energy was significantly greater for IDS than ADS. Furthermore, phase alignment between delta- and theta-band AMs was stronger in IDS compared to ADS. This remained the case when IDS and ADS were rate-normalized to control for differences in speech rate. These data indicate stronger rhythmic synchronization and acoustic temporal regularity in IDS compared to ADS, structural acoustic differences that may be important for early language learning.

### INTRODUCTION

Human speech perception relies in part on neural tracking of the temporal modulation patterns in speech at different timescales simultaneously: “multi-time resolution processing” (Chait, Greenberg, Arai, Simon, & Poeppel, 2015; Ghitza & Greenberg, 2009; Greenberg, 2006; Luo & Poeppel, 2007; Poeppel, 2003). According to multi-time resolution models, cortical oscillations entrain or phase-align their activity to modulations at corresponding timescales in the signal, thereby encoding the different energy patterns, and binding the information together in the final speech percept (Ghitza, 2011; Ghitza, Giraud, & Poeppel, 2012; Giraud & Poeppel, 2012; Poeppel, 2014). Exploration of the temporal characteristics of adult-directed speech (ADS) has revealed that accurate oscillatory phase alignment is mediated in part by amplitude “rise times,” auditory “edges” associated with amplitude (energy) modulations that help to specify temporal modulation rates (Gross et al., 2013). Rise times in the theta band appear to be particularly important for successful speech encoding, as acoustic “landmarks” (amplitude rise times) in critical band envelopes in the theta range of ADS provide perceptual markers that are critical for speech intelligibility (Doelling, Arnal, Ghitza, & Poeppel, 2014). When acoustic landmarks in the theta range (in this study, 2–9 Hz) were removed, speech became unintelligible, and when artificial landmarks of uniform height were inserted instead, the speech became intelligible again. Accordingly, Doelling et al. argued that auditory “edges”

Citation: Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind: Discoveries in Cognitive Science*. Advance Publication. doi:10.1162/opmi\_a\_00008

DOI:  
[http://doi.org/10.1162/opmi\\_a\\_00008](http://doi.org/10.1162/opmi_a_00008)

Supplemental Materials:

Received: 27 April 2016  
Accepted: 27 January 2017

Competing Interests: The authors declare no competing interests exist.

Corresponding Author:  
Usha Goswami  
[ucg10@cam.ac.uk](mailto:ucg10@cam.ac.uk)

Copyright: © 2017  
Massachusetts Institute of Technology  
Published under a Creative Commons  
Attribution 4.0 International  
(CC BY 4.0) license



The MIT Press

(amplitude rise times) in the theta band drive oscillatory activity to track and entrain to speech at its syllabic rate. Notably, all experiments to date have investigated literate adult participants.

By contrast, in child populations, delta band information appears to be critical for successful oscillatory entrainment. Studies of children with poor literacy (developmental dyslexia) reveal impaired encoding of speech envelopes between 0–2 Hz (Molinaro et al., 2016; Power, Colling, Mead, Barnes, & Goswami, 2016). Children with dyslexia also show atypical phase entrainment to rhythmic speech in the delta band, but equivalent entrainment to control children in the theta band (Power, Mead, Barnes, & Goswami, 2013). Further, studies of oscillatory entrainment to speech by typically developing children show a significant relationship between theta entrainment and learning to read (Power, Mead, Barnes, & Goswami, 2012). For child populations, therefore, it is possible that acoustic information in frequency bands other than theta may be of equal or even greater importance for speech processing. To explore this possibility, information about the temporal modulation structure of child-directed speech (CDS) and infant-directed speech (IDS) is required.

To explore CDS, we recently applied a probabilistic amplitude demodulation approach to modeling the rhythm patterning in children’s nursery rhymes (e.g., capturing whether they are iambic or trochaic). We compared its output to that of an engineered modulation filterbank approach (Leong, 2012; Leong, Stone, Turner, & Goswami, 2014; Turner, 2010), ultimately deriving a Spectral-Amplitude Modulation Phase Hierarchy (S-AMPH) model (Leong, 2012; Leong & Goswami, 2015). The S-AMPH modeling demonstrated that English nursery rhymes contained amplitude modulations in three critical temporal rate bands (corresponding to delta, theta, and beta/low gamma, for this speech corpus 0.9–2.5 Hz, 2.5–12 Hz, and 12–40 Hz; see (Leong, 2012; Leong & Goswami, 2015). Amplitude modulations in these bands were hierarchically nested in the nursery rhymes, and *phase alignment* between the slower amplitude modulation (AM) bands (delta and theta) played a key role in the perception of rhythmic patterning (judging whether the nursery rhymes were trochaic or iambic). A strong syllable was perceived when delta and theta modulation peaks were in alignment. Accordingly, the strong rhythmic character and acoustic temporal regularity of spoken nursery rhymes appears to be influenced by amplitude modulations in the delta band.

Like nursery rhymes, IDS is a highly rhythmic and temporally regular signal (Fernald et al., 1989; Jusczyk, Cutler, & Redanz, 1993). Accordingly, it is plausible that delta band information and delta-theta phase alignment may also be important characteristics of IDS. Infant-directed speech is classically described in terms of its higher pitch and greater pitch range, slower rate and higher amplitude than ADS, with hyperarticulated vowels (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997). Vowel hyperarticulation in IDS has been proposed to support phoneme discrimination (Liu, Kuhl, & Tsao, 2003) and word recognition (Song, Demuth, & Morgan, 2010), while high pitch and greater pitch range capture infant attention (Fernald & Kuhl, 1987) and facilitate speech segmentation (Thiessen, Hill, & Saffran, 2005). Neural multi-time resolution models of speech encoding (Giraud & Poeppel, 2012) raise the possibility that the temporal modulation hierarchy may also provide acoustic landmarks that support infant language learning, helping infants to locate word and phrase boundaries. Currently, the temporal modulation structure of IDS is unexplored. Accordingly, here we applied the S-AMPH modelling approach (Leong & Goswami, 2015) to IDS.

The S-AMPH model is a low-dimensional representation of the speech signal (see also Greenberg & Arai, 2001). The model generates a hierarchical representation of the dominant spectral (acoustic *frequency*) and temporal (oscillatory *rate*) modulation patterns in the

speech envelope. Amplitude modulation patterns in three temporal rate bands (~delta, theta, and beta/low gamma) form a three-tier nested hierarchy, thereby preserving AM patterning across different speaker rates. Annotation of the nursery rhyme (CDS) corpus revealed that for these nursery rhymes, this AM hierarchy (centered on ~2 Hz, ~5 Hz, and ~20 Hz AMs) mirrored the linguistic phonological hierarchy of stressed syllables, syllables, and onset-rime units in the spoken material (Leong & Goswami, 2015). Oscillatory cycles in each AM band thus supported the identification of phonological units of different sizes. The number of bands and their respective bandwidths used by the S-AMPH were originally determined using principal component analysis (PCA) dimensionality reduction of original high-dimensional spectral (29 equivalent rectangular bandwidth [ERB]-spaced frequency channels spanning 100–7,250 Hz) and temporal (24 modulation channels spanning 0.9–40 Hz) envelope representations (Leong, 2012; Leong & Goswami, 2015).

Here we applied the S-AMPH to IDS collected from a sample of 24 mother–infant dyads at 7, 9, and 11 months of age, and we also analyzed ADS collected from the same mothers. While application of the S-AMPH formulae to novel corpora cannot be assumed to identify the same linguistic units, the methods nevertheless enable comparison of relative speech energy in delta, theta, and beta/low-gamma bands and AM phase synchronization. We were interested to see whether natural IDS has a different temporal modulation structure to ADS, and whether the strength of phase alignment between the different AM bands (delta, theta, beta/low gamma) might show a different patterning in IDS as compared to ADS.

## **METHODS**

### ***Participants***

Twenty-four mother-infant dyads (9 male, 15 female infants) participated in the study.<sup>1</sup> None of the families had any history of language or cognitive disorders. All mothers spoke Australian English as their native language, and had normal (not > 0.5 *SDs* below the mean) nonverbal IQ on a combined subtest score averaging block design and matrix reasoning subscales of the Wechsler Adult Intelligence Scale IV (WAIS-IV; Wechsler, 2008) ( $M = 12.2$ ,  $SD = 2.0$ , compared with the population mean, 10.0, and  $SD$ , 3.0).

### ***Design and Experimental Setup***

There were four speaking conditions, three IDS, and one ADS. Infant-directed speech samples were collected longitudinally at three time-points, when the infants were 7 (mean age = 31 weeks,  $SD = 1.1$ ), 9 (mean age = 40 weeks,  $SD = 1.2$ ), and 11 (mean age = 48 weeks,  $SD = 1.2$ ) months of age. Mothers were told that the purpose of the experiment was to capture their natural interactions with their infants during a brief play session and soft toys and pictures were provided. Mothers were asked to talk to their infants for as long as they felt appropriate. All mothers and infants were observed to be motivated and engaged in the task. Mother and infant interacted alone in a laboratory room. The mother sat facing the infant, who was sitting in a highchair. A video camera was mounted in each corner of the room to allow for monitoring and video recording of the session, but only the audio data are analyzed here. The mother wore a head-mounted microphone (AudioTechnica AT892), connected to Adobe Audition CS6 software via an audio input/output device (MOTU Ultralite MK3). Speech samples were digitally recorded at 16 kHz (or at 44.1 kHz and resampled to 16 kHz). For the

---

<sup>1</sup> A subset of nine mother-infant dyads also contributed data to an earlier study piloting these analyses (Leong, Kalashnikova, Burnham, & Goswami, 2014).

ADS recordings, each mother interacted with a female experimenter in the laboratory room using the same recording apparatus, and the infant was not present. Mothers were interviewed about the experimental play sessions. Adult-directed speech samples were collected when the infant was approximately 12 months old (mean age: 53 weeks,  $SD = 8.2$ ). Both IDS and ADS recording sessions lasted between 5 and 10 minutes.

### Speech Data

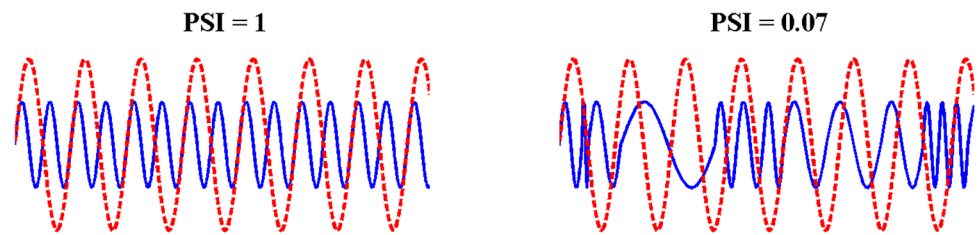
The raw audio recordings were manually divided into short segments for analysis using Praat software (Boersma & Weenink, 2013). Each segment contained a complete phrase from the original utterance. Portions of speech that were interrupted by the infant or adult addressee, or that contained excessive background noise, were not used. In total, across the four speaking conditions, mothers contributed 679 speech segments for analysis. On average, segments were 9.13s in length (range 5.02s–16.92s), and mothers contributed an average of 7.8, 7.9, and 7.1 segments at the infant ages of 7, 9, and 11 months, respectively. Mothers contributed on average 6.1 segments in ADS. The mean length of speech segments for the different speaking conditions was 9.2s (7 months IDS), 9.5s (9 months IDS), 9.1s (11 months IDS), and 8.7s (ADS).

### Speech Analysis Method

**Step 1: The S-AMPH Representation of the Amplitude Envelope** Each speech segment was z-scored (to standardize its mean and standard deviation), and a low-dimensional spectro-temporal model (i.e., S-AMPH) of its amplitude envelope was extracted using a two-stage filtering process, as described in Leong (2012) and Leong and Goswami 2015. First, the z-scored acoustic signal was band-pass filtered into five frequency bands (channel edge frequencies: 100, 300, 700, 1,750, 3,900, and 7,250 Hz) using a series of adjacent finite impulse response (FIR) filters. Next, the Hilbert envelope was extracted from each band-filtered signal and these five Hilbert envelopes were down-sampled to 1,050 Hz and passed through a second series of band-pass filters to isolate the three different AM bands within the envelope modulation spectrum. These three AM bands corresponded to delta rate modulations (0.9–2.5 Hz), theta rate modulations (2.5–12 Hz) and beta/low-gamma rate modulations, respectively (12–40 Hz). The result of this two-step filtering process was a 5 (frequency)  $\times$  3 (rate) spectro-temporal representation of the speech envelope, described as comprising delta-, theta-, and beta/low gamma-rate AM bands for each of the five spectral bands (Leong & Goswami, 2015).

**Step 2: Multi-timescale Synchronization Measure: Phase Synchronization Index (PSI)** Next, a measure of multi-timescale phase synchronization, the Phase Synchronization Index (PSI), was computed between adjacent modulation rate bands in the S-AMPH representation of each speech segment (i.e., delta-theta phase synchronization and theta-beta/low-gamma phase synchronization). This PSI computation was performed separately for each of the five spectral bands in the S-AMPH representation. The  $n : m$  PSI was originally conceptualized to quantify phase synchronization between two oscillators of different frequencies (e.g., muscle activity; Tass et al., 1998), and was subsequently adapted for use in neural analyses of oscillatory phase-locking (e.g., Schack & Weiss, 2005). This adaptation is applied to speech AMs in our model. The PSI was computed as:

$$PSI = |\langle e^{1(n\theta_1 - m\theta_2)} \rangle| \quad (1)$$



**Figure 1.** Hypothetical amplitude modulation (AM) pairs (sinusoids) that yield Phase Synchronization Index (PSI) scores of 1 (left) and 0.07 (right), respectively. The red and blue curves represent AMs with a frequency ratio of 1:2.

In Equation 1,  $n$  and  $m$  are integers describing the frequency relationship between the two AMs being compared. Following previous studies (Leong & Goswami, 2014, 2016; Leong, Stone, et al., 2014), for the delta-theta AM band analysis, an  $n : m$  ratio of 2:1 was used, while for the theta-beta/low-gamma AM band analysis, an  $n : m$  ratio of 3:1 was used. The values  $\theta_1$  and  $\theta_2$  refer to the instantaneous phase of the two AMs at each point in time. Therefore,  $(n\theta_1 - m\theta_2)$  is the generalized phase difference between the two AMs, which was computed by taking the circular distance (modulus  $2\pi$ ) between the two instantaneous phase angles. The angled brackets denote averaging of this phase difference over all time-points. The PSI is the absolute value of this average, and can take values between 0 (no synchronization) and 1 (perfect synchronization), as illustrated in Figure 1. A sound with a PSI of 1 is perceived as being perfectly rhythmically regular (a repeating pattern of strong and weak beats), whereas a sound with a PSI of 0 is perceived as being random in rhythm.

**Step 3: Modulation Spectrum** To assess the modulation spectrum of the speech samples, the sub-band Hilbert envelopes of the stimuli (resulting from stage 1 of the previously-described S-AMPH decomposition procedure) were individually passed through a modulation FIR filterbank with 24 channels logarithmically spaced between 0.9–40 Hz. For each speech sample, and each frequency sub-band, the mean power across all modulation channels was computed, and the relative power difference from this mean was computed for each modulation channel. Finally, the differenced modulation power spectrum was averaged across the five frequency sub-bands for each speech sample, and this grand average was used for statistical analysis.

#### **Control Analysis With Rate-Normalized Data**

The speaking rate of IDS was slower than that of ADS. As this could introduce potential confounds, a rate-normalization procedure was also performed, rescaling the IDS data samples to the same temporal rate as the ADS samples. First, each IDS and ADS speech sample was manually annotated (by native English speakers) to ascertain the number of syllables and stressed syllables that it contained. This provided a mean syllable rate and stressed syllable rate per second for each participant in each speaking condition. For each IDS condition, and for each participant, a scaling factor was then computed based on the ratio between their stressed syllable rate in the ADS condition, and their stressed syllable rate in the IDS conditions. This scaling factor was then used to resample each IDS utterance to the appropriate length for that speaker (e.g., if the stress rate for IDS was half that of ADS, then the IDS sample was compressed so that it was half of its original length). Identical signal processing steps and statistical analyses were then conducted with the rescaled data.

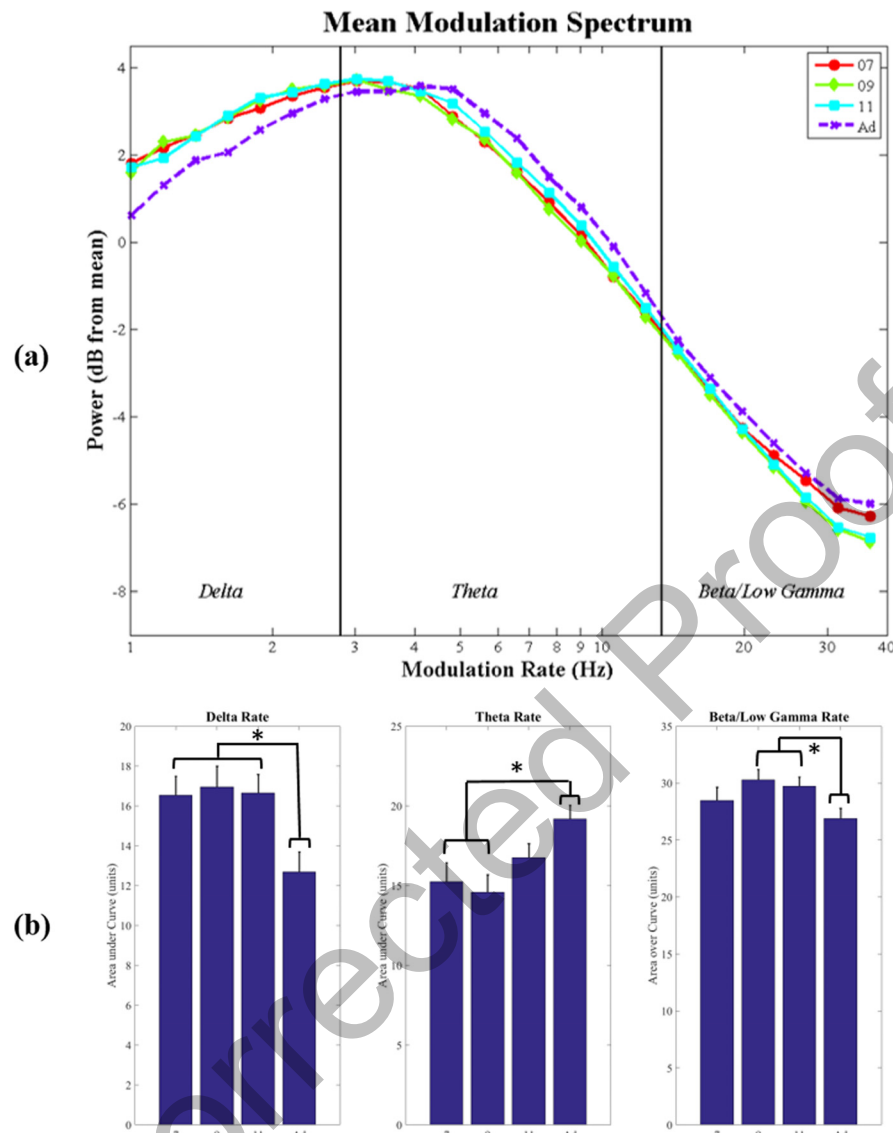
## RESULT

### Natural Speech

**Modulation Spectrum** Figure 2 (panel a) shows the mean modulation spectrum of the speech amplitude envelope for each speaking condition. Compared to ADS, the modulation spectrum of IDS contained more energy in the delta band (0.9–2.5 Hz), whereas the modulation spectrum of ADS showed the typical profile of most energy in the theta band, as previously reported (Greenberg, Carvey, Hitchcock, & Chang, 2003). To assess whether these differences were statistically significant, the area under the curve corresponding to the three modulation bands was computed (shown in Figure 2, panel b) and assessed using a repeated measures ANOVA with modulation rate (three levels) and speaking condition (four levels) as within-subjects factors. There was a significant interaction between Rate and Condition,  $F(6, 138) = 7.78$ ,  $p < .0001$ , showing that the relative amount of modulation energy in the delta, theta and beta/low-gamma bands differed between IDS and ADS. Tukey post-hoc analysis revealed that all three IDS conditions contained significantly more modulation energy than ADS in the delta band ( $p < .05$  for all). Conversely, for the theta band, ADS contained significantly more modulation energy than IDS at 7 and 9 months ( $p < .05$  for both), but not at 11 months ( $p = .059$ ). In the beta/low-gamma band, ADS contained significantly more modulation energy than 9- and 11-month IDS ( $p < .05$  for both) but did not differ from 7-month IDS ( $p = .81$ ). Finally, modulation energy did not differ between the three IDS conditions for any AM band ( $p > .93$  for all).

**Phase Synchronization (PSI)** Next, the phase synchronization between AM bands in the mothers' IDS was assessed using the PSI between pairs of AM bands in the S-AMPH representation of the speech envelope. Mean PSI values are shown in Figure 3. Each subplot shows the PSI value (ranging from 0 [no synchronization] to 1 [perfect synchronization]) on the  $y$ -axis, and acoustic frequency on the  $x$ -axis. The figure shows more synchronization in IDS than ADS for delta-theta AM band synchronization in low-frequency spectral bands (100–300 Hz and 300–700 Hz), but less synchronization for theta-beta/low-gamma AM bands. Moreover, for delta-theta band phase synchronization in the middle spectral frequency band (700–1,950 Hz), typically corresponding to vocalic energy, PSI values increase with increasing age of the addressee (7m < 9m < 11m < ADS).

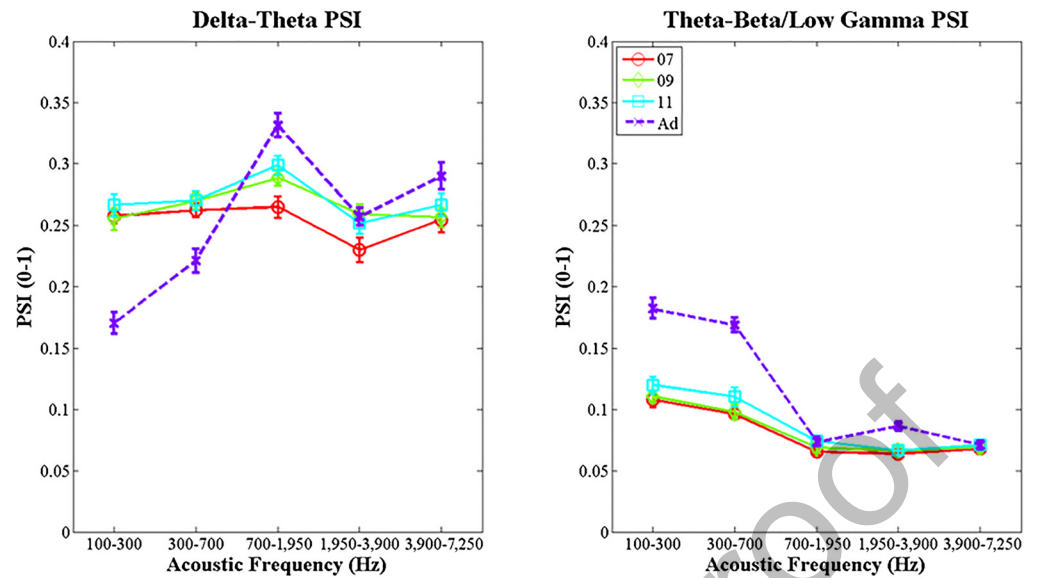
To compare the conditions statistically, two repeated measures ANOVAs were performed, taking either delta-theta band PSIs or theta-beta/low-gamma band PSIs as the dependent variable. In each ANOVA, speaking condition (four levels) and spectral frequency band (five levels) were the within-subjects factors. For the delta-theta band ANOVA, there were significant main effects of Condition,  $F(3, 69) = 3.07$ ,  $p < .05$ , and Spectral Band,  $F(4, 92) = 27.76$ ,  $p < .001$ . The interaction between Condition and Spectral Band was also significant,  $F(12, 276) = 17.92$ ,  $p < .001$ . Post-hoc analysis (Newman-Keuls) indicated that for the two lowest spectral frequency bands (100–300 Hz; 300–700 Hz), PSI scores for all three IDS conditions were significantly higher than for ADS ( $p < .01$  for all comparisons). By contrast, for the middle and highest spectral frequency bands (700–1,950 Hz; 3,900–7,250 Hz), PSI scores for all IDS conditions were significantly lower than for ADS ( $p < .05$  for all comparisons). The PSI scores for the 1,950–3,950 spectral frequency band did not differ between IDS and ADS. Finally, for the middle spectral frequency band (700–1,750 Hz), there was a graded effect between IDS at the different ages. Delta-theta AM band phase synchronization in IDS to 7-month-olds was lower (less synchronized) than in IDS to 11-month-olds ( $p < .05$ ). However, it did not differ



**Figure 2.** Panel (a). Mean modulation spectrum for infant-directed speech (IDS) (“07,” “09,” and “11” months) and adult-directed speech (ADS) (“Ad”) conditions, computed for each of five spectral frequency bands (100–300 Hz; 300–700 Hz; 700–1,750 Hz; 1,750–3,900 Hz; 3,900–7,250 Hz) and then averaged across bands and speakers. The x-axis shows the modulation rate, and the y-axis the power in dB (normalized with respect to the mean power of each frequency band in each sample). Vertical lines indicate the boundaries used to delineate the area under the curve for Spectral-Amplitude Modulation Phase Hierarchy (S-AMPH) delta-, theta-, and beta/low-gamma-rate amplitude modulation (AM) bands. Panel (b). Area under the curve computed for the modulation spectrum in the delta, theta, and beta/low-gamma ranges for each speaking condition (“7” = 7 month IDS, “9” = 9 month IDS, “11” = 11 month IDS, “Ad” = ADS). Note that for the beta/low-gamma range, as the curve falls below zero, the area over the curve is reported instead. Here, a smaller numerical value denotes greater power in the beta/low-gamma range.

statistically from delta-theta AM band phase synchronization in IDS to 9-month-olds ( $p = .14$ ). This suggests that the rhythmic regularity of IDS increases as a function of infant age.

For the theta-beta/low-gamma AM band ANOVA, there were again significant main effects of Condition,  $F(3, 69) = 45.84$ ,  $p < .001$ , and Spectral Band,  $F(4, 92) = 119.88$ ,



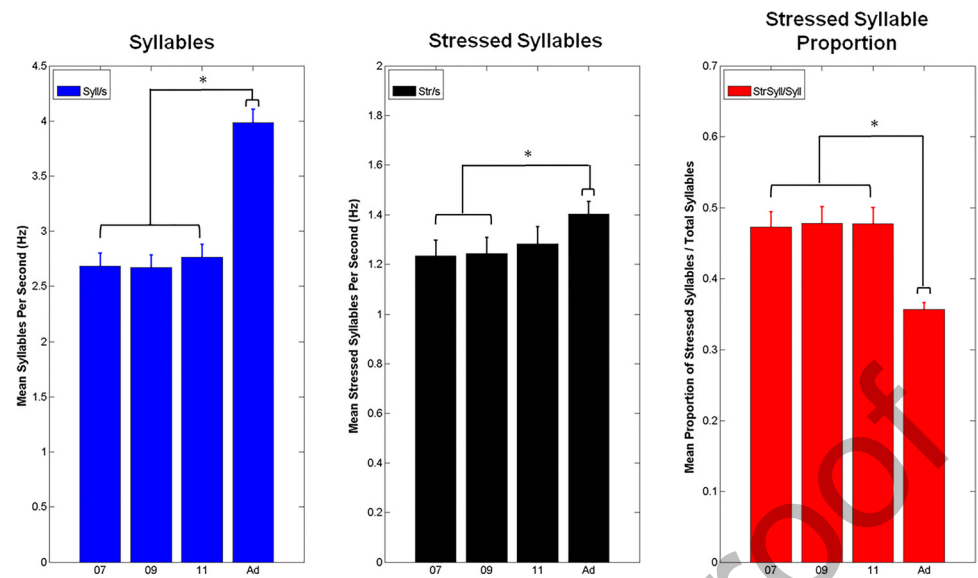
**Figure 3.** Phase Synchronization Index (PSI) values computed for the delta-theta amplitude modulation (AM) bands (left) and the theta-beta/low-gamma AM bands (right), for infant-directed speech (IDS) collected at each different age (different-colored solid lines) and adult-directed speech (ADS) (dotted line). The *x*-axis indicates the frequency (spectral) band from which the AMs originate. The *y*-axis shows the PSI value. Error bars indicate the standard error of the mean.

$p < .001$ , and a significant interaction between Condition and Spectral Band,  $F(12, 276) = 18.21$ ,  $p < .001$ . However, Newman-Keuls post-hoc analysis revealed a *reverse symmetry* to the results obtained for the delta-theta AM band comparisons. For the majority of spectral frequency bands (first, second, and fourth; 100–300 Hz; 300–700 Hz; 1,950–3,900 Hz), the IDS conditions showed lower PSI scores than ADS ( $p < .001$  for all comparisons). However, PSI scores in the middle and highest spectral frequency bands (700–1,950 Hz; 3,900–7,250 Hz) did not differ between speaking conditions ( $p > .05$  for all comparisons).

#### Control Analysis: Rate-Normalized Data

**Syllables and Stressed Syllables per Second** Figure 4 shows the mean number of syllables and stressed syllables produced per second for IDS and ADS, along with the mean proportion of stressed syllables as a function of the total number of syllables produced. We used repeated measure ANOVAs to compare speaking conditions in each case. For the number of syllables produced, the ANOVA showed a significant main effect of speaking condition,  $F(3, 69) = 74.42$ ,  $p < .0001$ . Tukey post-hoc analysis showed that ADS had a significantly greater number of syllables per utterance than the three IDS conditions ( $p < .001$ ), which did not differ significantly from each other. For the number of stressed syllables produced, the ANOVA again showed a significant main effect of speaking condition,  $F(3, 69) = 3.42$ ,  $p < .05$ . Tukey post-hoc analysis revealed significantly more stressed syllables per second in ADS compared to IDS at 7 months and 9 months ( $p < .05$ ), but not IDS at 11 months. Finally, the ANOVA for the proportion of stressed syllables produced in the different speaking conditions also showed a significant effect of condition,  $F(3, 69) = 28.33$ ,  $p < .00001$ . Here, post-hoc tests showed that ADS contained a significantly *lower* proportion of stressed syllables than IDS at any age ( $p < .001$  for all comparisons). Therefore, although ADS contained more syllables and more stressed syllables than IDS within a given period of time (i.e., a *quantitative* rate effect), there was a *qualitative* difference between the speaking styles. Infant-directed speech contained a





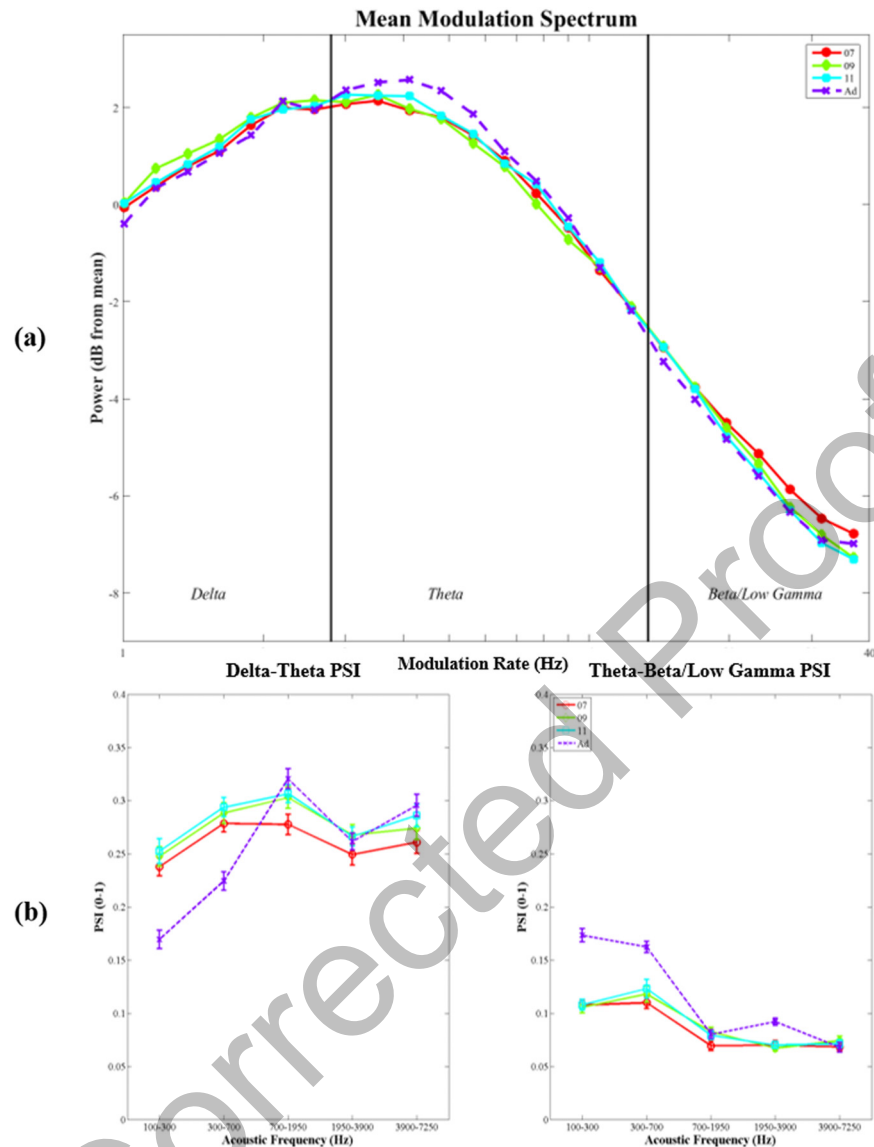
**Figure 4.** Mean number of syllables per second, stressed syllables per second, and proportion of syllables that were stressed for the infant-directed speech (IDS) and adult-directed speech (ADS) data for each speaking condition (“07” = 7 month IDS, “09” = 9 month IDS, “11” = 11 month IDS, “Ad” = ADS).

higher proportion of stressed to unstressed syllables. In IDS, nearly half (~48%) of all syllables were stressed, whereas in ADS, only around one-third (~36%) of syllables were stressed.

#### Modulation Spectrum

The modulation spectra for the rate-normalized control data are shown in Figure 5 (panel a). Comparison with Figure 2 suggests that rate-normalization has not changed the pattern of IDS/ADS difference. A  $3 \times 4$  ANOVA (Modulation Rate  $\times$  Condition) showed the predicted interaction between Rate and Condition,  $F(6, 138) = 2.16, p < .05$ , one-tailed. Post-hoc analysis (Fisher LSD) showed that IDS at 9 months contained significantly more modulation energy in the delta band than in ADS ( $p < .05$ ). For the theta band, ADS contained significantly more energy than IDS at 7 and 9 months ( $p < .05$  for both), but not at 11 months ( $p = .077$ ). Finally, in the beta/low-gamma band, ADS contained significantly more energy than IDS at 7 months ( $p < .05$ ). Therefore, the patterns observed with natural speech were broadly conserved after the rate-normalization procedure.

**Phase synchronization (PSI)** The PSI values obtained for the rate-normalized data are also shown in Figure 5 (panel b). Two  $4 \times 5$  (Speaking Condition  $\times$  Spectral Band) ANOVAs were again performed, again taking either delta-theta band PSIs or theta-beta/low-gamma band PSIs as the dependent variable. For the delta-theta band ANOVA, there were significant main effects of Condition,  $F(3, 69) = 5.82, p < .01$ , and Spectral Band,  $F(4, 92) = 41.41, p < .0001$ , and a significant interaction,  $F(12, 276) = 9.88, p = .0001$ . Tukey post-hoc analysis indicated that ADS had significantly lower PSI scores ( $p < .001$ ) than all three IDS conditions for the two lowest spectral frequency bands (100–300 Hz; 300–700 Hz). By contrast, for the middle spectral frequency band (700–1,950 Hz), PSI scores for ADS were significantly higher than for IDS, but at 7 months only ( $p < .01$ ).



**Figure 5.** Panel (a). Mean modulation spectrum for rate-normalized infant-directed speech (IDS) (“07”, “09,” and “11” months) and adult-directed speech (ADS) (“Ad”) conditions, computed for each of five spectral frequency bands (100–300 Hz; 300–700 Hz; 700–1,750 Hz; 1,750–3,900 Hz; 3,900–7,250 Hz) and then averaged across bands and speakers. The *x*-axis shows the modulation rate, and the *y*-axis the power in dB (normalized with respect to the mean power of each frequency band in each sample). Vertical lines indicate the boundaries used to delineate the area under the curve for Spectral-Amplitude Modulation Phase Hierarchy (S-AMPH) delta-, theta- and beta/low-gamma-rate amplitude modulation (AM) bands. Panel (b). Phase Synchronization Index (PSI) values computed for the rate-normalized data, for delta-theta AM bands (left) and the theta-beta/low-gamma AM bands (right), for IDS collected at each different age (different-colored solid lines) and ADS (dotted line). The *x*-axis indicates the frequency (spectral) band from which the AMs originate. The *y*-axis shows the PSI value. Error bars indicate the standard error of the mean.

For the theta-beta/low-gamma AM band ANOVA, there were again significant main effects of Condition,  $F(3, 69) = 27.76, p < .00001$ , and Spectral Band,  $F(4, 92) = 96.69, p < .0001$ , and a significant interaction,  $F(12, 276) = 15.30, p < .0001$ . Tukey post-hoc analysis revealed that ADS scores were significantly higher ( $p < .01$ ) than IDS scores at all ages in the first, second, and fourth spectral bands (100–300 Hz; 300–700 Hz; 1,950–3,900 Hz), but

did not differ for the third and fifth spectral bands (700–1,950 Hz; 3,900–7,250 Hz). Therefore, the results of the PSI analysis with rate-normalized data were again consistent with the PSI analysis conducted with natural speech.

## DISCUSSION

The temporal modulation structure of IDS differs significantly from that of ADS. First, there is significantly greater modulation energy in the delta band in IDS relative to ADS, while ADS shows greater energy in the theta band compared to IDS. This difference is not simply a consequence of IDS typically being spoken more slowly than ADS, as demonstrated by the rate-normalized control analysis. The higher delta power in IDS was accompanied by a higher proportion of syllables being stressed in IDS as compared to ADS, providing a more rhythmic input. Second, IDS showed significantly greater phase synchronization between the slower delta- and theta-rate AM bands compared to ADS. This finding was also mirrored in the control analysis with rate-normalized data. Accordingly, there are clear differences in the pattern of rhythmic synchronization between IDS and ADS. Greater phase synchronization between delta- and theta-rate AM bands accompanies greater rhythmic regularity in IDS: syllables are stressed more regularly.

These findings have important implications for the neural basis of language acquisition, at least for English. The temporal rate bands at these different timescales can be assumed to entrain neuronal oscillatory activity (Giraud & Poeppel, 2012). For adult speech, oscillations in the delta (1–3 Hz), theta (4–8 Hz), and beta/low-gamma bands (15–30 Hz / >30 Hz) respectively have been shown to be critical for encoding (frequency ranges from Poeppel, 2014). One proposal is that oscillatory entrainment to amplitude modulations in these bands helps to parse the speech signal into linguistically relevant units (*delta*—syllable stress patterns, *theta*—syllables, *beta*—onset-rime units, *low gamma*—phonetic information). Our data suggest that IDS is acoustically structured to facilitate neural entrainment to *delta band* (prosodic) information by the infant brain. Indeed, a large infant behavioral literature attests to the importance of stressed syllables in early word segmentation (e.g., Cutler & Norris, 1988; Echols, 1996). Further, at least two studies of infant entrainment to nonspeech AMs have revealed significant neuronal entrainment in this frequency range for German-learning infants, including newborns (Telkemeyer et al., 2009; Telkemeyer et al., 2011). The stronger phase synchronization between delta- and theta-rate AM bands that characterizes IDS would further facilitate the extraction of rhythmic patterning, for example, in distinguishing trochaic versus iambic prosodic structure (Leong, Stone, et al., 2014). The infant electrophysiological literature shows that infants are encoding the difference between trochees and iambs by 4 months of age (Weber et al., 2004). These differences in phase synchronization strength between slower (IDS) versus faster (ADS) timescales could also be related to recent demonstrations that phonetic contrasts in IDS are less clear than in ADS (Martin et al., 2015; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013).

Indeed, phase synchronization at faster modulation timescales was greater in ADS, and this effect was also preserved in the rate-normalized data. Perceptually, this would increase the rhythmic regularity of the acoustic relationship between syllables and phonemes, which could potentially reflect the acquisition of literacy. A more regular temporal patterning of phonemes within syllables may reflect the explicit awareness of phonemes in syllables that is a cognitive consequence of learning to read (see Ziegler & Goswami, 2005, for data across languages). Accordingly, ADS may show stronger acoustic patterning at faster temporal rates because most ADS speakers are literate. Experimental investigation of the temporal modulation structure of

ADS as produced by literate versus illiterate populations offers one method for exploring this possibility further (Araújo, Flanagan, Castro-Caldas, & Goswami, 2016).

### ACKNOWLEDGMENTS

We thank Maria Christou-Ergos for preparation of data for analysis. This work was supported by funding from the Australian Research Council (DP110105123) to DB and UG. Victoria Leong is now in the Division of Psychology, Nanyang Technological University, Singapore.

### AUTHOR CONTRIBUTIONS

VL, DB, and UG conceived and designed the study, MK and DB collected the data, VL analyzed the data, VL, MK, DB, and UG wrote the article.

### REFERENCES

- Araújo, J., Flanagan, S., Castro-Caldas, A., & Goswami, U. (2016, June). *The temporal modulation structure of illiterate versus literate speech*. Poster presented at the 9th National Symposium on Research in Psychology, Portuguese Psychological Association, Faro, Portugal.
- Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer. Version 5.3.51, <http://www.praat.org/>
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002) What's New Pussycat? On talking to babies and animals. *Science*, 296, 1435.
- Chait, M., Greenberg, S., Arai, T., Simon, J. Z., & Poeppel, D. (2015). Multi-time resolution analysis of speech: Evidence from Psychophysics. *Frontiers in Neuroscience*, 9, 214.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121.
- Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, 761–768.
- Echols, C. H. (1996). A role for stress in early speech segmentation. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 151–170). Mahwah, NJ: Lawrence Erlbaum.
- Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, 10, 279–293.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501.
- Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, 2, 130.
- Ghitza, O., Giraud, A.-L., & Poeppel, D. (2012). Neuronal oscillations and speech perception: Critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, 6, 340.
- Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66, 113–126.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15, 511–517.
- Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In S. Greenberg & W. Ainsworth (Eds.), *Understanding speech: An auditory perspective* (pp. 411–434). Mahwah, NJ: Lawrence Erlbaum.
- Greenberg, S., & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. *Proceedings of the 7th Eurospeech Conference on Speech Communication and Technology (Eurospeech-2001)*, 473–476.
- Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech: A syllable-centric perspective. *Journal of Phonetics*, 31, 465–485.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biology*, 11(12), e1001752.
- Jusczyk, P. W., Cutler, A., & Redanz, N.J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675–687.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., . . . Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277, 684–686.
- Leong, V. (2012). *Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models* (Doctoral dissertation). University of Cambridge. Retrieved August 20, 2014, from <http://www.cne.psychol.cam.ac.uk/pdfs/phds/vleong>
- Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, 308(100), 141–161.

- Leong, V., & Goswami, U. (2015). Acoustic-emergent phonology in the amplitude envelope of child-directed speech. *PLOS ONE*, 10(12), e0144411.
- Leong, V., & Goswami, U. (2016). Difficulties in auditory organization as a cause of reading backwardness? An auditory neuroscience perspective. *Developmental Science*. Advance online publication. doi:10.1111/desc.12457
- Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2014, September). *Infant-directed speech enhances rhythmic structure in the envelope*. Paper presented at Interspeech, Singapore.
- Leong, V., Stone, M., Turner, R., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America*, 136, 366–381.
- Liu, H. M., Kuhl, P., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science*, 6(3), F1–F10.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54, 1001–1010.
- Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, 26(3), 341–347.
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, 129(2), 362–378.
- Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*. Advance online publication. doi:10.1002/hbm.23206
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, 41, 245–255.
- Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, 28c, 142–149.
- Power, A. J., Colling, L. C., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, 160, 1–10. doi:10.1016/j.bandl.2016.06.006
- Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2012). Neural entrainment to rhythmically-presented auditory, visual and audio-visual speech in children. *Frontiers in Psychology*, 3, 216.
- Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, 7, 777.
- Schack, B., & Weiss, S. (2005). Quantification of phase synchronization phenomena and their importance for verbal memory processes. *Biological Cybernetics*, 92, 275e287.
- Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America*, 128(1), 389–400.
- Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., & Freund, H. J. (1998). Detection of n:m phase locking from noisy data: Application to magnetoencephalography. *Physical Review Letters*, 81, 3291.
- Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., . . . Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience*, 29, 14726–14733.
- Telkemeyer, S., Rossi, S., Nierhaus, T., Steinbrink, J., Obrig, H., & Wartenburger, I. (2011). Acoustic processing of temporally-modulated sounds in infants: Evidence from a combined NIRS and EEG study. *Frontiers in Psychology*, 2, 62.
- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53–71.
- 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>
- Weber, A., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research*, 18, 149–161.
- Wechsler, D. (2008). *WAIS IV administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131, 3–29.