Local temporal regularities in child-directed speech in Spanish

Jose Pérez-Navarro[1,2*], Marie Lallier[1], Catherine Clark[1,2], Sheila Flanagan[3] & Usha Goswami[3]

[1]BCBL, Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, Spain

[2]University of the Basque Country UPV/EHU, Donostia-San Sebastian, Spain

[3]Centre for Neuroscience in Education, Department of Psychology, University of Cambridge, Cambridge, United Kingdom

[*]Corresponding author: josejpereznavarro@gmail.com

Abstract

Purpose: To characterize the local (utterance-level) temporal regularities of child-directed speech (CDS) that might facilitate phonological development in Spanish, classically termed a syllable-timed language.

Method: 18 female adults addressed their 4-year-old children versus other adults spontaneously and also read aloud (CDS versus ADS). We compared CDS and ADS speech productions using a spectro-temporal model (S-AMPH, Leong & Goswami, 2015), obtaining three temporal metrics: 1) distribution of modulation energy, 2) temporal regularity of stressed syllables, and 3) syllable rate.

Results: CDS was characterized by 1) significantly greater modulation energy in the lower frequencies (0.5-4 Hz), 2) more regular rhythmic occurrence of stressed syllables, and 3) a slower syllable rate than ADS, across both spontaneous and read conditions.

Discussion: CDS is characterized by a robust local temporal organization (i.e., within utterances) with amplitude modulation bands aligning with delta and theta electrophysiological frequency bands respectively showing greater phase synchronization than in ADS, facilitating parsing of stress units and syllables. These temporal regularities, together with the slower rate of production of CDS, might support

the automatic extraction of phonological units in speech and hence support the phonological development of children.

Introduction

Under typical listening conditions, humans effortlessly process and comprehend speech as it unfolds over time. Several theories suggest that cortical oscillations (the relatively regular synchronous firing of neuronal populations) in the auditory and broader language regions synchronize to the speech signal at several timescales (Ghitza, 2011; Giraud & Poeppel, 2012). Such synchronization mechanisms allow the temporal processing of speech and facilitate its comprehension. Neurophysiological research corroborates this view by showing cortical tracking of speech acoustic cues that map onto linguistic syllables and prosodic patterns (e.g., Ding & Simon, 2012; Doelling et al., 2014; Molinaro & Lizarazu, 2018; Peelle et al., 2013). Moreover, there is direct evidence that links an efficient cortical tracking of prosodic (Rimmele et al., 2021; delta band oscillations, 0.5 – 4 Hz) and syllabic (Doelling et al., 2014; theta band oscillations, 4 – 8 Hz) acoustic cues in the speech signal with speech comprehension. While most of the evidence about the oscillatory mechanisms for tracking acoustic regularities in speech comes from proficient adult populations, infants' and children's abilities to track the temporal cues of speech have also been studied (e.g. Attaheri et al., 2022; Gervain & Werker, 2013; Ríos-López et al., 2017; Tallal, 1980). However, there is currently little evidence concerning whether the temporal regularities of child-directed speech (CDS) are enhanced (as compared to adult-directed speech, ADS) in order to support and guide the emergence of a phonological system. There is also little evidence concerning which statistical forms this temporal enhancement may take. Answering this question is crucial for a comprehensive developmental framework that considers how the brains of infants and children exploit the temporal regularities of the speech they are typically addressed with to achieve proficient language comprehension.

Several studies have highlighted the presence of temporal regularities within the prosodic and syllabic timescales, which inform the aims of the present study. At the syllabic level, the rate of approximately 5 syllables per second (5 Hz) is common across languages (Ding et al., 2017; Greenberg et al., 2003).

At the prosodic stress level, Tilsen and Arvaniti (2013) showed that amplitude envelope-based methods (similar to those used in the present study) could capture stress regularities in spontaneous utterances. In the same vein, Inbar et al., (2020) found that prosodic units (termed ‘intonation units’ in their study) produced by adult speakers appear at a roughly constant rate of ~1 Hz. Interestingly, Stehwien and Meyer (2021) analyzed an annotated corpus of radio newscasts in German to show that the prosody of intonational phrases (mapping onto utterances) determined the periodicity of their nested subordinate phrases, suggesting that prosody could have a determining role in shaping the local temporal regularities of adult-directed speech. Overall, the evidence suggests that there is a close overlap between the rhythms of quasi-regular speech units such as stressed syllables and syllables and the timescales at which neurophysiological mechanisms operate to subserve their processing (see Poeppel & Assaneo, 2020 for a comprehensive review on the rhythms of speech production and perception).

While it is well established that human neurocognitive abilities subtending the extraction and segmentation of phonological units in speech fine-tune and gain language specificity during the early years of life (for reviews see Kuhl, 2004; Skeide & Friederici, 2016; Werker & Hensch, 2015), it is still unclear how the speech inputs directed to infants and children provide them with robust temporal statistics that can support this phonological tuning. Of particular interest for the present study are the low-frequency temporal statistics present in the amplitude envelope of the speech signal, governed by amplitude modulations (AMs) centered at different temporal rates. AMs are systematic intensity changes in the speech signal, mainly taking place at the delta (~2 Hz) and theta (~5 Hz) rate bands of AM, that help to signal the occurrence of linguistic units like prosodic phrasing (~1000 ms) and syllables (~200 ms) respectively (Ding, Patel, et al., 2017; Greenberg, 2006; Greenberg et al., 2003). Such temporal fluctuations in the amplitude envelope of the speech signal, particularly the AM rise times (rates of change for these AM bands), provide salient acoustic markers relevant to extracting prosodic and syllabic phonological units, while faster modulations (~35 Hz) are thought to contribute to the extraction of phonemic information (Poeppel et al., 2008). The identification of phonological units in the speech signal is crucial for phonological and reading development (Ziegler & Goswami, 2005). Behavioral evidence, in line with the evidence on the cortical tracking of speech (e.g., Doelling et al., 2014; Rimmele et al., 2021), highlights the functional role of tracking delta and theta AMs in sentence segmentation and syllabic parsing respectively (Ghitza, 2012, 2017). A key functional role for delta and theta AMs is also in line with the Temporal Sampling (TS) Framework (Goswami, 2011), a developmental framework for language acquisition centered on phonology. TS theory proposes that the automatic alignment of endogenous brain rhythms with AM-governed rhythm patterns in speech is critical for linguistic and phonological development, and that this unconscious neural alignment (or sampling) process may be atypical in developmental dyslexia, which is characterized by both phonological and amplitude rise time difficulties.

Coherent with the TS hypothesis, two bodies of evidence attest the key role of tracking low frequency speech AMs for phonological development. Firstly, multiple studies across languages have shown that impairments in AM sensitivity accompany the atypical phonological development characteristic of developmental dyslexia (e.g., Goswami et al., 2002, 2010; Leong & Goswami, 2014; Surányi et al., 2009; see Hämäläinen et al., 2012 for a systematic review). Secondly, sensitivity to AMs during the first years of life is a predictor of outcomes in fundamental language domains, such as phonological awareness (Goswami, Wang, et al., 2010; Vanvooren et al., 2017), vocabulary (Kalashnikova et al., 2019), and reading abilities (Vanvooren et al., 2017). In addition, recent longitudinal studies show that cortical oscillatory tracking of prosodic information is present in infants from 4 months, and increases during early childhood (Ríos-López et al., 2020; Attaheri et al., 2022), suggestive of the relevance of delta-band speech tracking for language development. Ríos-López et al., (2021) show that a bigger delta-band cortical tracking of speech in pre-reading children indeed predicts better reading skills one year later, after the beginning of formal reading instruction.

Previous evidence shows that adults adapt their speech complexity to children’s linguistic abilities and communicative feedback, in order to facilitate comprehension (Kalashnikova et al., 2020; Lam & Kitamura, 2012; Smith & Trainor, 2008). There is abundant evidence concerning the spectral (pitch) characteristics of infant-directed speech (IDS), which are exaggerated to make it a phonetically-salient and engaging register to address language-learning individuals (Dilley et al., 2020; Fernald, 1985; Kuhl et al., 1997; Trainor & Desjardins, 2002; Werker et al., 2007; Werker & McLeod, 1989; see Fernald, 2000 for a review). The enhanced spectral characteristics of IDS are well-established, however less is known regarding potential temporal adaptations that may take place when addressing infants and children. Two well-known temporal features of IDS are a slower speech rate and shorter utterances (Fernald et al., 1989; Fernald & Simon, 1984; Leong et al., 2017). It may be the case that CDS could also provide especially regular temporal statistics to facilitate identification of and access to phonological units in speech and thereby to facilitating the emergence of a proficient phonological system. Such a hypothesis was initially explored by Leong and Goswami (2015) in relation to the AM organization of CDS in English, typically regarded as a stress-timed language (i.e., a language characterized by certain regularity in the timing of stressed syllables); and further tested by contrasting IDS and ADS in English (Leong et al., 2017). In the latter study, Leong et al. (2017) showed that IDS differed from ADS in its temporal organization, especially regarding two critical aspects. One was the higher prominence of delta band modulation energy in IDS compared to ADS: the modulation spectrum revealed relatively more power in the delta band for IDS than for ADS. This feature is likely linked to enhanced prosody in IDS, providing more salient temporal information relevant to extracting phonological information at slower timescales (e.g., intonation phrases, words, and stressed syllables) to a learning individual. The second feature was that stressed syllables were more regularly spaced in IDS than ADS, shown by significantly greater phase synchronization (rhythmic alignment) of delta-rate and theta-rate AMs (~2Hz and ~5 Hz respectively) in IDS. This was interpreted as providing a predictable temporal skeleton to facilitate the infant's attentional and perceptual access to syllables during early stages of language learning.

However, to date, there is no study concerning the potential benefit that the temporal organization of CDS (in contrast to ADS) could provide during pre-school years, nor to what extent such temporal organization is present in non-stress timed languages like Spanish. Languages like Spanish are characterized by salient syllabic timing, and thus have been traditionally categorized as syllable-timed languages, (see Ramus et al., 1999, and Varnet et al., 2017, for instances of supporting evidence; but also Arvaniti, 2009; Turk & Shattuck-Hufnagel, 2013 for opposing views). Here we focus on kindergarten, a stage in which phonological abilities (e.g., phonological awareness and phonological short-term memory) are explicitly taught, as they will support later reading acquisition (e.g., Caravolas et al., 2001; Muter et al., 2004). We investigated whether the temporal regularities of CDS differed from those of ADS in Spanish, by directly contrasting the temporal statistics of the two speech registers within the same study for the first time. If CDS shows similar salient temporal features to English, in principle this could signal the presence of language-universal temporal statistics that may facilitate learning, particularly regarding an emergent phonological system. To this purpose, we focused on three temporal features of speech: the modulation spectrum, the temporal regularity of the placement of stressed syllables and syllable rate. We studied the two features —modulation spectrum and the temporal regularity of the placement of stressed syllables—that Leong et al., (2017) already found distinctive in IDS in English, a stress-timed language. The modulation spectrum for each speech register was computed and the area under the curve (AUC) was compared in delta versus theta bands for CDS and ADS respectively. Our aim was to discern whether in Spanish, the two speech registers can be differently categorized as more prosody-salient (greater AUC in the delta-rate AM band) or syllable-salient (greater AUC in the theta-rate AM band). To characterize the regularity with which syllables were stressed in CDS in contrast to ADS, we analyzed the temporal alignment between delta and theta AM bands in terms of AM phase alignment (rhythmic synchronicity). To this purpose, we used the spectral-amplitude modulation phase hierarchy (S-AMPH) model developed by Leong and Goswami (2015). The S-AMPH model allows us to decompose the amplitude envelope of the speech signal and measure the temporal alignment between different AM bands nested within the signal in different words and phrases in terms of their phase synchronization (see Figure 1 for a phrasal example). Of particular interest for our study, delta-theta phase alignment plays a crucial role in the perception of prosodic patterns in English and has been proposed as a novel statistic for the language-learning brain (Leong & Goswami, 2015). Greater delta-rate to theta-rate AM phase synchronization is thought to help to identify prosodic patterning by specifying strong versus weak syllables (Leong et al., 2014). When both AM bands peak together, a strong syllable is heard. When a trough in the slower delta-rate AM band (centered on ~2 Hz in the speech materials used by Leong et al., 2014) coincides with a peak in the faster theta-rate AM band (centered on ~4 Hz in Leong et al., 2014), a weak syllable is heard. Whether the same is true in Spanish is currently unknown. Finally, we analyzed syllable rate. Our goal was to extend previous findings of CDS being more slowly paced than ADS (Biersack et al., 2005; Sjons et al., 2017), and to investigate the potential links between a putative slower speech rate in CDS and its expected enhanced temporal regularities. In summary, the role of sensitivity to AM information for efficient speech processing and language development is well supported. In the present study, we take a step further and explore whether specific AM regularities in the acoustic signal of CDS in Spanish (AUC in delta versus theta AM bands, delta-rate to theta-rate AM phase synchronization, and speech rate) are enhanced in comparison to ADS, with the assumption that developing an emergent phonological system should benefit from the presence of salient temporal statistics in the input. By testing Spanish, classically considered to be a syllable-timed language, our results should provide developmental evidence regarding the possibly universal relevance of AM phase relations to extracting phonological grain sizes in language learning. Further, our data can offer a comprehensive link between the cumulative knowledge from the cognitive neuroscience of language about cortical tracking of speech and universal processes in language acquisition.

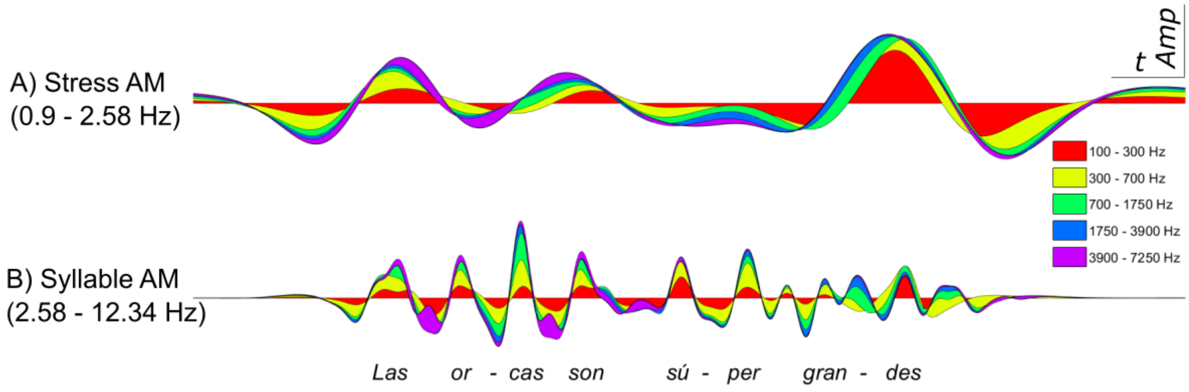


Figure 1. Example of S-AMPH model's spectro-temporal decomposition of an utterance ("*Las orcas son súper grandes"*, *Whales are super big*). **A**) Stress AM band (delta range, 0.9 – 2.58 Hz). **B**) Syllable AM band (theta range, 2.58 – 12.34 Hz). To estimate prosodic and syllabic salience, amplitude modulation is extracted from Stress (A) and Syllable (B) bands respectively. To estimate the regularity of stressed syllables, we calculated the phase alignment between 1 cycle of Stress (A) and 2 cycles of Syllable (B) AMs.

Method

Participants and conditions

We recorded the CDS and ADS speech productions of 18 female Spanish-speaking adults (*mean age* = 39.06 years; *SD* = 5.39). All participants had attained higher education and lived in urban areas of the Basque Country. 16 participants can be considered monolinguals (exposed to Spanish more than 70% of their time) and 2 participants can be considered Spanish-Basque bilinguals (exposed to their second language, Basque, at least 30% of their time). We selected them based on Spanish being the language they used to address others in the vast majority of interactions (*mean use of Spanish* = 87.5 %; *SD* = 9.20). For CDS speech productions, participants were accompanied by their 4 year-old children (*N* = 18, 6 females, *mean age* = 4.1 years; *SD* = .35), with the aim of generating as ecologically valid CDS productions as possible, like those that could happen in everyday life (Lam & Kitamura, 2012; Smith & Trainor, 2008). The purpose of having 4-year-old children as addressees of CDS was to ensure that children were mature enough to understand the purpose and, therefore, be attentive and quiet during the CDS recordings (~20 minutes). In the ADS productions, participants addressed one of the experimenters (*N* = 2, 1 female, mean age = 28.1 years; SD = .4). For each speech register, participants were asked to (i) address their child or the adult interlocutor in spontaneous speech monologues—the critical spontaneous CDS and ADS conditions—, and (ii) read to their interlocutors—baseline reading CDS and ADS conditions. Although our main purpose was to study spontaneous speech, we added baseline reading conditions to control for potential participant variability in their spontaneous productions (see Hirose & Kawanami, 2002) as well as for discerning whether CDS shows boosted temporal regularities regardless of its production context. Each participant thus took part in four speaking conditions: spontaneous CDS, read CDS, spontaneous ADS, and read ADS. Participants were provided with several topics to facilitate their spontaneous productions to children (e.g., animals and pets, family trips, anecdotes that their children liked, etc.) and adults (e.g., participant's studies, working life, how they spent their leisure and family time, etc.). Elicitation instructions were minimal in order to generate speech productions as ecologically valid as possible, and were the following: "please, talk/read to the child/adult about any of the mentioned topics in an engaging way. Let us know if you run out of ideas, and we will suggest a few new topics.” We recorded each participant during between 9 and 10 minutes per speaking condition, to get at least 8 minutes of analyzable continuous speech signal (i.e., after removing noisy and silent segments) per condition.

Speech recordings

Speech was recorded in a soundproof room while participants and addressees were seated in front of each other, with a cardioid microphone (Sennheiser e 840) at approximately 10 centimeters from each speakers’ head. Continuous speech (single channel, 44.1 kHz, 16-bit PCM) was segmented into utterances based on their terminal intonation contour, and at the start of pauses longer than 2 seconds between productions, according to widely used standard criteria (Miller, 1981). Additionally, utterances with more than two coordinate clauses were segmented before the second conjunction (i.e., “*and*”), to avoid spuriously lengthening due to clausal chaining (Rice et al., 2006). Utterances containing false starts, repetitions and reformulations were either excluded or trimmed to their correct formulation to limit the impact of those factors in our temporal metrics (Tree, 1995). In total, participants provided 5070 utterances. We excluded from further analyses 645 utterances (12.72 % of the data set) shorter than 2 seconds, as they do not provide enough information for reliable low-frequency (~1 Hz) AM estimations. Thus, the final dataset was composed of 1084 spontaneous CDS, 1400 read CDS, 1067 spontaneous ADS, and 874 read ADS utterances. After segmentation, the volume levels of each utterance were z-scored prior to our temporal analyses.

Temporal analyses

We used a spectro-temporal acoustic model (S-AMPH, Leong & Goswami, 2015) that allowed us to characterize the multiscale temporal hierarchy of amplitude modulation information in speech. To achieve this, the S-AMPH model reduces the dimensions of the speech signal into three AM bands in two main steps. First, band-pass filtering the z-scored utterances into 5 spectral bands (band edge frequencies: 100, 300, 700, 1,750, 3,900, and 7,250 Hz) through a series of adjacent zero-phase finite impulse response (FIR) filters. Second, each spectral band signal was Hilbert filtered, and subsequently band-pass filtered through an additional series of 3 AM bands: delta (0.9 – 2.58 Hz), theta (2.58 – 12.34 Hz) and beta/low-gamma (12.34 – 40 Hz). The ranges of our AM bands, determined by the signal-driven model construction of S-AMPH for English, map closely onto the frequency bands typically linked with prosodic (delta, 0.5 – 4 Hz), syllabic (theta, 4 – 7 Hz) and phonemic (beta/low gamma, 12 – 50 Hz) timescales respectively (e.g., Giraud & Poeppel, 2012). These timescales were mapped for each of the 5 different spectral bands, which are color coded in Figure 1. Figure 1 depicts the output of the model for the delta and theta bands for a single phrase. Visual inspection of Figure 1 shows that some peaks in the delta band correspond to peaks in the theta band. In these cases, phase synchronization indices (PSI values) would be larger, indicating the likely presence of a stressed syllable. The figure also shows that typically the S-AMPH modeling produces one theta band peak per syllable in a given utterance.

We estimated both the modulation (AM) spectrum and the phase synchronization between AM bands to test whether CDS and ADS differed in the distribution of their modulation rates and in their phase relations regarding delta-rate and theta-rate AMs. The modulation spectrum analysis approximately indicates whether we can categorize each register as more prosody-prominent versus syllable-prominent respectively. Since utterances that were too long, too short, or that contained long pauses could bias modulation rate estimates, we limited the modulation spectrum analyses to utterances in the range of 2 to 6 seconds and excluded utterances with silences longer than 1 second. To characterize the modulation spectra of our speech materials, we Hilbert filtered each utterance’s 5 bands resulting from the first S-AMPH step and passed them through a FIR filterbank with 24 log-spaced channels ranging from 0.9 to 40 Hz. We then computed mean power across modulation channels for each frequency band, followed by the power difference from the mean (in dB) for each modulation channel. We used the average power difference from the mean of the 5 spectral bands for further statistical analyses of the modulation spectrum.

The phase synchronization index (PSI) estimates the rhythmic relations between the adjacent delta-rate and theta-rate AM bands (A and B respectively in Figure 1). Cross-frequency PSI quantifies phase alignment between two oscillators of different frequencies (Tass et al., 1999; see S1). This is achieved by adjusting the *n:m* ratio, in which *n* and *m* are the number of cycles of the lower (delta in this study) and higher (theta) frequency oscillators, respectively. PSI values range between 0 (no phase synchronization) and 1 (perfect phase synchronization). The n:m ratio that best accommodated delta-theta PSI for our Spanish materials was 1:2 (see S2). Therefore, PSI results are computed using the 1:2 ratio. S-AMPH model also extracts a beta/low gamma (12.34 – 40 Hz) AM band, mapping onto phonemes/onset-rime units. Given that the hypotheses of the present study address low frequency (< 12 Hz) modulations, we did not further analyze such higher frequency beta/low gamma AM band. However, it is noteworthy that 1:2 was also the ratio that best suited theta-beta/low gamma phase alignment, which is in line with a previous S-AMPH analysis of IDS and ADS in English (Leong et al., 2017). Since we obtained 5 delta-theta PSIs per utterance (one per spectral band), we averaged them and conducted our statistical analyses on mean PSI.

We computed syllable rate to assess whether CDS is slower paced than ADS, as well as whether the speed at which utterances are produced contributes to their temporal regularity. Syllable rate was semiautomatically computed in Praat (Boersma & Weenink, 2021), based on the acoustic algorithm developed by de Jong and Wempe (2009). Volume parameters were adapted to the decibel (dB) levels of each participant’s recording to obtain reliable syllable rate estimates regardless of between-participant loudness and pitch differences. We validated a subset of 1584 (38 % of all utterances) of automatic syllable rate metrics with their corresponding manually annotated syllable rate indexes, estimated by trained native speakers, showing indeed a high correlation between manually annotated and automatically detected syllable rate ($r(1582) = .95$, $p < .001$; S3).

Results

In order to assess the influence of speech register (CDS, ADS), and speaking condition (spontaneous speech, read speech) on each of our temporal measures (distribution of modulation energy, phase synchronization, and syllable rate), we used linear mixed effect (LME) models. Given the within-participant structure of our study, we included each participant as a random intercept in the model. We used the *lmer* function of lme4 package (v.1.1.28, Bates et al., 2015) as well as *anova* function to test the omnibus main effects and interactions of our predictors.

Modulation Spectrum (Prosodic salience)

To operationalize our planned analyses concerning the peak locations of the modulation spectra (Figure 2A), we calculated the area under the curve (AUC), defined as the linear transformation of each frequency band's difference in dBs from mean power. Delta and theta segments of the modulation spectrum differed greatly in their AUC (Figure 2), as previously shown by other studies (e.g., Ding, Patel, et al., 2017). Overall (i.e., across registers and conditions), AUC was significantly bigger in theta than in delta, $t(36) = 25.62$, $p < 0.001$ ($\beta = 0.260$, $SE = 0.010$, CI [0.240 0.280]). Therefore, we circumscribed our planned analyses to each of the AM bands separately. LME showed that, within **delta**, there were significant effects of speaking register, $F(1, 54) = 11.45$, $p = 0.001$, condition, $F(1, 54) = 51.43$, $p < .001$, and an interaction between register and condition, $F(1, 54) = 23.39$, $p < 0.001$. This pattern of results reveals a bigger delta AUC in CDS than in ADS, $t(54) = 5.81$, $p < 0.001$ ($\beta = 0.066$, $SE = 0.011$, CI [0.043 0.089]), as well as in spontaneous than in read speech, $t(54) = 8.49$, $p < 0.001$ ($\beta = 0.097$, $SE = 0.011$, CI [0.074 0.120]). In the **theta** segment of the modulation spectrum, LME also yielded a significant effect of speaking register, $F(1, 54) = 20.51$, $p < 0.001$, condition, $F(1, 54) = 98.07$, $p < 0.001$, as well as an interaction between both factors, $F(1, 54) = 18.79$, $p < 0.001$. However, the theta segment of the modulation spectrum was characterized by the inverse pattern relative to delta, namely ADS showing a bigger theta AUC than CDS, $t(54) = 6.27$, $p < 0.001$ ($\beta = 0.027$, $SE = 0.004$, CI [0.019 0.036]), as well as read speech showing a bigger theta AUC than spontaneous speech, $t(54) = 10.07$, $p < 0.001$ ($\beta = 0.044$, $SE = 0.004$, CI [0.035 0.052]). Indeed, for **theta**, the modulation spectrum of all conditions peaked at around 5 – 6 Hz, corresponding to the syllable rate (as previously shown across languages; Ding, Patel, et al., 2017; Greenberg et al., 2003).

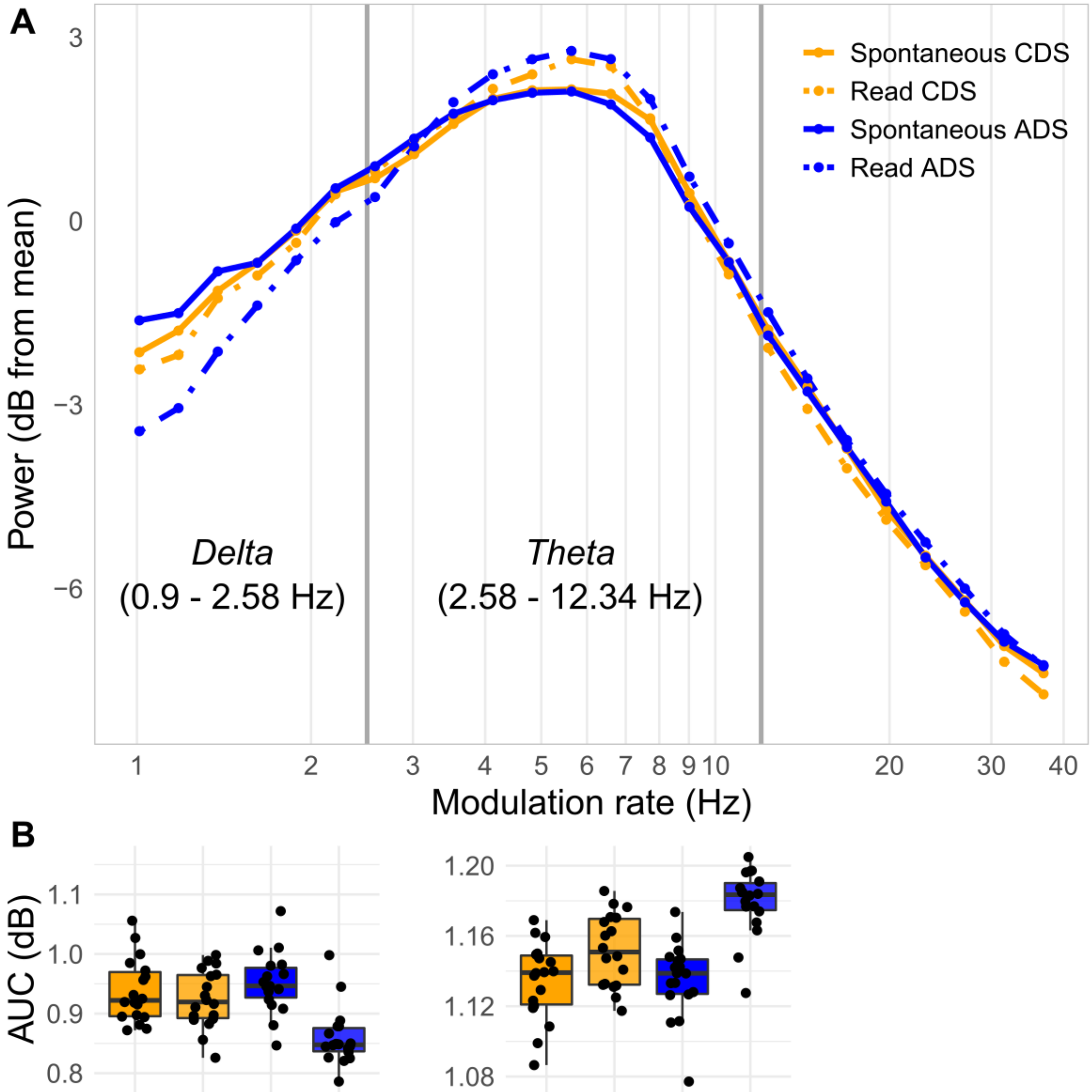


Figure 2. A) Modulation spectra of the four speaking conditions. The vertical grey lines divide the signal-derived modulation rates of the S-AMPH model that we used to define delta and theta bands, to which we subset our PSI and AM spectrum analyses. **B**) Area under the curve (AUC) of the delta (left) and theta (right) bands of spontaneous CDS, read CDS, spontaneous ADS, and read ADS respectively from left to right. The horizontal lines between conditions represent significant differences in AUC, adjusted for multiple comparisons (** $p < .01$; **** $p < .0001$). **C**) AUC of canonical theta band (4 - 7 Hz). Significant differences between speaking conditions are represented as in section B. Dots in Panels B and C represent mean AUC values per participant.

Thus, CDS spontaneous and read speech had significantly greater modulation energy (i.e., bigger delta AUC) than ADS spontaneous and read speech respectively, suggestive of more salient prosodic structure in CDS. The results for spontaneous speech are in line with the IDS-ADS prosodic differences in IDS in English demonstrated by Leong et al. (2017). The data for read CDS are completely novel. Moreover, and in line with the differences between read and spontaneous materials that have been reported with respect to prosody (e.g., Hirose & Kawanami, 2002; Howell & Kadi-Hanifi, 1991), our results suggest that when reading to or spontaneously addressing adults in a syllable-timed language, a greater syllabic salience (i.e., bigger theta AUC) is found.

Regularity of stressed syllables (delta-theta phase synchronization, PSI)

Delta-theta PSI values in the different spectral bands demonstrated a similar pattern across speech registers (S4). Therefore, we first computed an LME model with mean PSI values as the dependent variable. The LME yielded a significant effect of speaking register, $F(1, 54) = 26.82$, $p < .001$, showing that CDS is characterized by higher delta-theta phase synchronization than ADS, $t(54) = 2.49$, $p = 0.016$ ($\beta = 0.011$, $SE = 0.004$, CI [0.002 0.020]) (Figure 3). There was no significant effect of speaking condition (spontaneous vs. read) nor interaction between register and condition ($p > .05$).

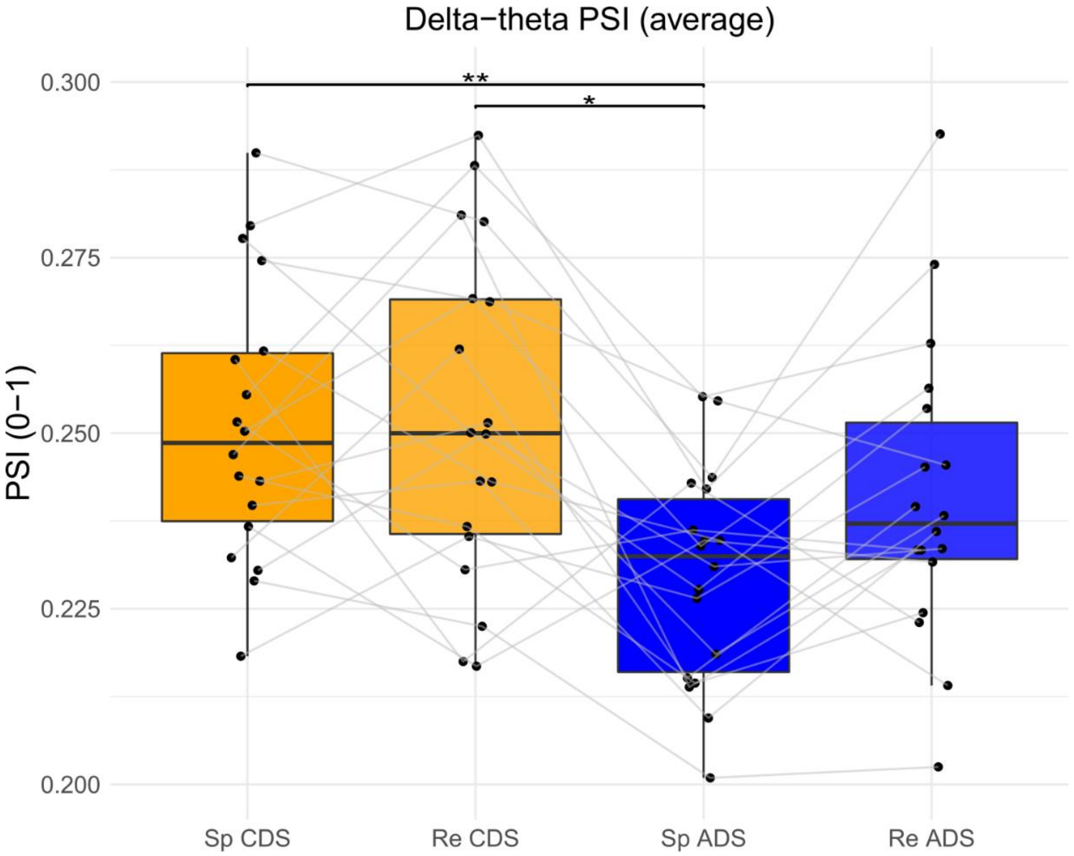


Figure 3. Mean delta-theta PSI. Gray lines connect participants' mean PSI across conditions. Horizontal lines within each box represent median PSI. Upper and lower hinges mark the first and third quartile, and whiskers show 1.5 * inter-quartile range. Bonferroni-corrected significant differences are represented with * (*p* < .05) and ** (*p* < .01).

Syllable rate

The LME model on syllable rate yielded significant effects of speech register ($F(1, 54) = 8.32$, $p = .006$) and condition ($F(1, 54) = 7.93$, $p = .007$), but no interaction between these factors ($p > 0.05$). These main effects are visible in Figure 4, which shows the higher syllable rate of ADS relative to CDS, $t(54) = 2.203$, $p = 0.032$ ($\beta = 0.262$, *SE* = 0.119, *CI* [0.025 0.499]), and of read speech relative to spontaneous speech, $t(54) = 2.155$, $p = 0.036$ ($\beta = 0.256$, *SE* = 0.119, *CI* [0.019 0.493]). Figure 4 also shows that there is much less variability in the speech rate of read speech, and interestingly, particularly of speech read to children (CDS). This suggests that readers spontaneously adapt their speech when reading to children to make it highly predictable. It should be noted that the method we used to calculate syllable rate yields slightly smaller values than manual annotation or other typically used calculations. Accordingly, we multiplied our syllable rate values by 1.28 as stated in the method's manuscript (de Jong & Wempe, 2009). This confirmed an overlap with the peak of the modulation spectrum in the theta band for each register and condition (Figure 2).

Next, we analyzed the relationship between syllable rate and the temporal regularity of the utterances. The negative correlations between syllable rate and delta-theta PSI were significant (Figure 5). This shows that the slower paced utterances were the most temporally organized utterances in our dataset.

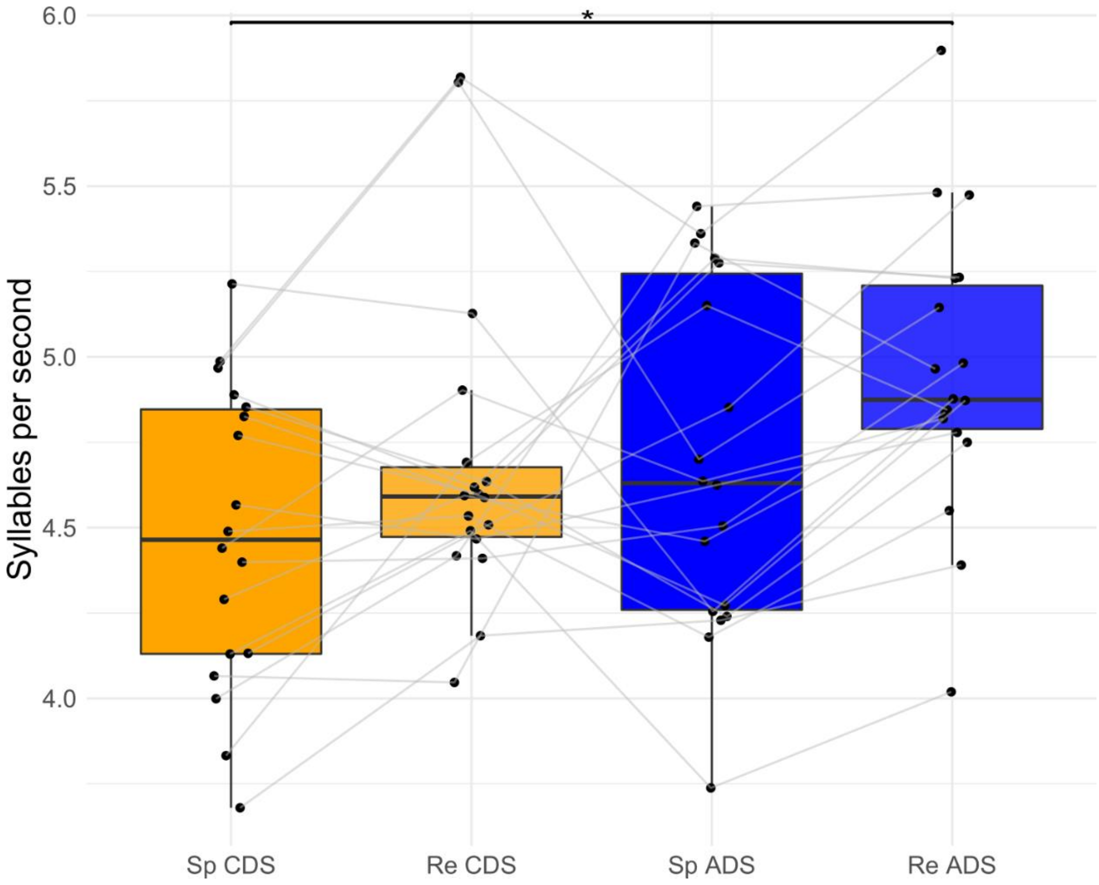


Figure 4. Syllable rate across speaking conditions. Gray lines connect participants' mean syllable rates across conditions. Horizontal lines within each box represent median syllable rate. Upper and lower hinges mark the first and third quartile, and whiskers show 1.5 * inter-quartile range. Bonferroni-corrected significant differences are represented with * ($p < .05$).

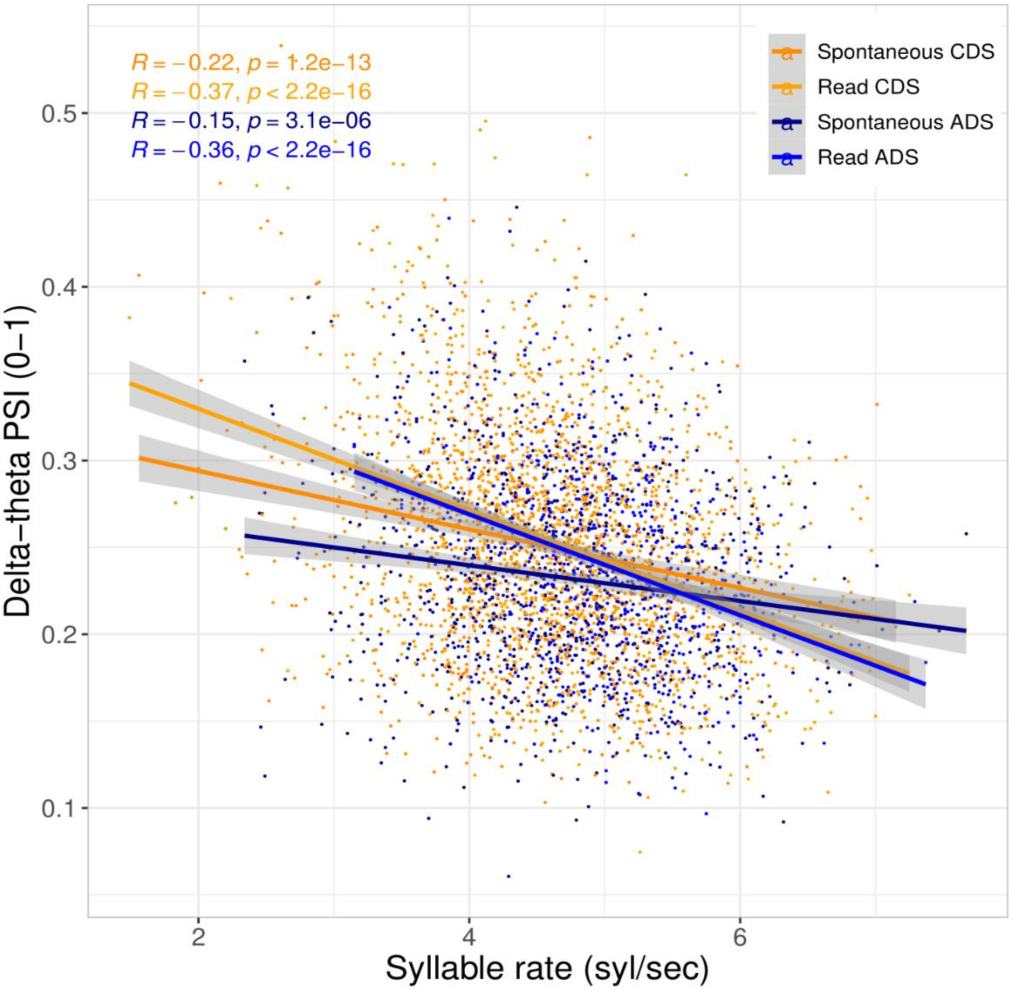


Figure 5. Correlation between syllable rate and delta-theta PSI. The four lines indicate the slopes of fitted linear models for each speaking condition. Top left: Pearson correlation coefficients and p-values for each speaking condition.

Discussion

In the present study, we investigated both spontaneous and read CDS and ADS in Spanish with the objective of contrasting them in terms of temporal regularities. Our within-participant design allowed us to investigate whether adults flexibly adapt their spontaneous speech productions to boost speech temporal regularities when addressing 4-year-old children rather than other adults. Using three temporal metrics, we found that CDS in Spanish carries more regular temporal statistics than ADS. First, CDS has significantly more modulation energy in the delta band than ADS, whether it is spoken spontaneously or whether the adult is reading to the child. Second, CDS contains more regularly stressed syllables than ADS, as shown by the greater phase alignment of the delta-rate and theta-rate AM bands in the CDS registers. Third, CDS shows a slower syllable rate relative to ADS. Adults slow down when speaking to children, as might be expected when addressing language-learning individuals. Indeed, read CDS also showed a notably narrower range than the other registers regarding syllable rate, suggesting that when reading to young children, temporal information becomes highly predictable. This may help to explain why early story reading is such an important contributor to language development (Attig & Weinert, 2020).

The modulation spectrum analyses (Figure 2) for CDS suggested that it has significantly more modulation energy in the delta band than ADS. This is in line with prior IDS data in English (Leong et al., 2017), classically considered a stress-timed language. The fact that we also found enhanced prosodic salience in CDS in Spanish, typically termed a syllable-timed language, is consistent with the idea that IDS and CDS boost suprasegmental information to aid the mapping of phonological units by language-learning individuals (Fernald, 2000). Neurophysiological studies show that infants and children rely more on suprasegmental/prosodic than syllabic information for tracking and segmenting continuous speech (Attaheri et al., 2022; Ríos-López et al., 2020), which may help to explain the enhanced delta band modulation energy in Spanish CDS. Despite this enhancement of delta-band modulations, and as expected, the modulation spectrum for this syllable-timed language peaked in the theta band for both Spanish CDS and ADS, as has previously been reported across languages for ADS (e.g., Ding, Patel, et al., 2017; Greenberg et al., 2003). However, ADS showed significantly more modulation energy in the theta band compared to CDS. This might indicate that the temporal regularities of ADS are more systematic at the syllabic timescale, either because syllables are a fundamental temporal landmark for adult neurocognitive speech processing abilities (Doelling et al., 2014; Ghitza, 2012), or because the receivers and producers of ADS are literate (Araujo et al., 2018), or both.

Our findings of higher delta-theta PSIs in CDS utterances suggest that stressed syllables are temporally more regularly placed in CDS than in ADS. Indeed, prior speech modelling work has shown that delta-theta AM phase relationships underpin speech rhythm perception (Leong et al., 2014), with AM peak synchronization helping to determine the perceived metrical patterning of utterances such as trochaic versus iambic. Given the syllable-timed nature of Spanish, the greater predictability of syllable stress may help the phonological mapping of Spanish by language-learning individuals. Our data thus contribute to the current evidence on continuous speech rhythmicity, by showing that, at utterance level, CDS is more rhythmic than ADS. Our findings concerning rhythmicity are also in line with previous adult studies that have contextualized the temporal regularities of speech within local (utterance level) stress patterns (Arvaniti, 2009; Nolan & Jeon, 2014; Tilsen & Arvaniti, 2013). Indeed, there is recent evidence for local prosodic stress regularities in ADS in different languages (e.g., Inbar et al., 2020; Stehwien & Meyer, 2021). The slower syllable rate in CDS relative to ADS is also of relevance when comparing temporal statistics. In summary, CDS appears to offer a continuous speech stream that is easier to segment via slower speech rate (fewer syllables per second), greater rhythmicity (predictability of occurrence of stressed syllables), and the enhancement of delta-band speech information. In line with this interpretation, previous evidence shows that while adult neurocognitive mechanisms adapt to different speech rates within the 4 – 7 Hz range (e.g., Foulke & Sticht, 1969; Ghitza, 2011; Lizarazu et al., 2019), children’s comprehension abilities benefit from slower speech rates (Berry & Erickson, 1973; Haake et al., 2014; Montgomery, 2004; Riding & Vincent, 1980). The adapted temporal statistics in Spanish CDS demonstrated here could thus aid comprehension by children as well as facilitating the development of a phonological system. The enhanced local (utterance-level) temporal regularities of CDS, whether it is read or spoken, provide a set of temporal statistics that can be exploited by children’s neurocognitive mechanisms for statistical (Romberg & Saffran, 2010) and distributional learning (see Banai & Ahissar, 2018 for a review). Sensitivity to these AM-related statistics would enable a listening child to build increasingly more robust phonological representations at word and syllable level. Indeed, the mapping of speech temporal statistics is known to be inefficient in individuals with phonological deficits such as dyslexia (Ahissar et al., 2006; Banai & Ahissar, 2018; Goswami, 2011; Leong & Goswami, 2017). Previous studies with adults have also shown prosodic (Inbar et al., 2020) and syllabic (Ding, Patel, et al., 2017) regularities across languages. However, our results highlight that greater temporal synchronization between delta-rate and theta-rate AMs may be a specific characteristic of

CDS (in contrast to ADS). This finding is consistent with an ‘acoustic-emergent’ perspective regarding phonological development from infancy onwards (Leong & Goswami, 2015).

Regarding potential cross-language universality, it is notable that Leong et al., (2017) found the same enhanced delta-rate to theta-rate AM synchronisation as found here in English IDS when compared to English ADS, hence in a stress-timed language. This may imply that there are certain key AM statistics that are universal concerning the temporal regularities present in the amplitude envelope of speech directed to young learners. As these two languages are typically grouped into two different rhythmic categories (i.e., English, stress-timed; and Spanish, syllable-timed), the current findings may point in principle towards an enhanced rhythmic organization of speech when addressing language learners, regardless of the rhythmic timing of a specific language. We propose that at the very early stages of language development, infants are presented with speech inputs that contain higher pitch (Fernald, 2000), enhanced delta-band modulation energy and prominent rhythm (delta-theta AM phase synchronization), the latter providing temporal landmarks to begin the task of speech segmentation in the form of identifying and predicting the placement of stressed syllables (Leong et al., 2017; Cutler & Norris, 1988). Thus, infants can rely on salient spectro-temporal information that is boosted in IDS to orient their attention to acoustic cues relevant to extracting phonological information. As lexical knowledge develops and children progress in their word segmentation skills, the temporal regularity in CDS is exploited to parse the stress patterns characterizing whole-word phonological forms. This may be of particular relevance in languages that, like Spanish, have a greater proportion of multi-syllabic words than English. Once an efficient language processing system has developed, ADS can then contain less regular temporal statistics, as such regular statistics are not required to aid segmentation. Indeed, adults can adapt their linguistic processing via the over-learned temporal predictions of proficient language models (e.g., Molinaro et al., 2021; Ten Oever & Martin, 2020).

However, additional cross-linguistic evidence in languages belonging to other rhythmic categories (e.g., the mora-timed rhythms of Japanese), as well as in languages in which lexical stress is completely predictable (e.g., French), or has different degrees of unpredictability (e.g., Basque) is required to test this cross-language developmental hypothesis. Such studies could help to further our understanding of the possible enhancement of delta-theta phase synchronization in CDS and its potential role in phonological development. In addition to cross-linguistic evidence, cross-cultural investigations are needed to contextualize these findings regarding the temporal regularities of child-adapted speech, as there are also cultural and socioeconomic factors that shape the quantity and quality of IDS and CDS (Cristia et al., 2019; Schick et al., 2022; see Cristia, 2022 for a systematic review). Although there is evidence of the maturation of cortical tracking of delta-rate versus theta-rate AMs in infants (Attaheri et al., 2021) and children (Ríos-López et al., 2020), as well as about the potential role that it has for phonological development and reading acquisition (Ríos-López et al., 2021), further studies are needed to fill the gap regarding the emergence of cortical tracking of syllables from infancy and during childhood, and how this may be aided by the temporal regularities of IDS and CDS.

In closing, our data are also relevant to evaluating the Temporal Sampling hypothesis of developmental dyslexia, which suggests that there is a specific link between delta- and theta-rate AM sensitivity and phonological development during the first years of life across languages (Goswami, 2011; Goswami et al., 2016; Goswami, Wang, et al., 2010; Vanvooren et al., 2017). Longitudinal neurophysiological evidence in Spanish shows that cortical tracking of speech in children relies mainly on prosodic (delta band) acoustic information (Ríos-López et al., 2020, 2021), and a similar pattern is found longitudinally for infants in English (Attaheri et al., 2022). Indeed, a recent study by Menn et al. (2022) in Dutch found that the cortical tracking of the delta AM rate in infants predicted their later vocabulary knowledge. Our findings are consistent with such evidence, showing that enhanced temporal regularities within the delta frequency band (0.5-2.5 Hz, the timescale of stressed syllables) occur more reliably in CDS (as previously shown in English IDS, Leong et al., 2017) than in ADS. Moreover, our results are broadly in line with the Temporal Sampling hypothesis from the perspective of the importance of temporal AM statistics <10 Hz for phonological development. Future studies that directly compare IDS, CDS and ADS could help to delineate the developmental sequence of the temporal regularities that an emerging phonological system needs to map in order to aid comprehension and language learning.

CRediT authorship contribution statement

Jose Pérez-Navarro: Conceptualization, Formal Analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. **Marie Lallier:** Conceptualization, Writing – original draft, Writing – review & editing. **Catherine Clark:** Investigation. **Sheila Flanagan:** Formal Analysis, Methodology, Software, Visualization, Writing – review & editing. **Usha Goswami:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors do not have known competing financial interests or personal relationships that can influence the investigation reported in this article.

Data availability statement

All code used for processing, analysis and visualization is available at the Open Science Framework project's repository. Given that participants provided personal information in their speech recordings, we only share some audio snippets that do not compromise their anonymity. Nonetheless, all preprocessed files are available at the project's repository to ensure the reproducibility of our analyses.

Acknowledgements

We want to thank the participants and their children for their volunteer contribution to our study. We also thank Tatsuya Daikoku, João Araújo, Marina Kalashnikova, Elena Aguirrebengoa, and Patricia Jiménez for their collaboration in different aspects of the study. This study was supported by the FPI grant BES-2016-078125 by Ministerio Español de Economía, Industria y Competitividad MINECO and Fondo Social Europeo; through project RTI2018-096242-B-I00 (MCIU/AEI/FEDER, UE) funded by Ministerio de Ciencia, Innovación y Universidades (MCIU), the Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER); by the Basque Government through the BERC 2018-2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490.

References

Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. https://doi.org/10.1038/nn1800

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Attaheri, A., Choisdealbha, Á. N., Liberto, G. M. D., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2021). *Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants* (p. 2020.10.12.329326). https://doi.org/10.1101/2020.10.12.329326

Attig, M., & Weinert, S. (2020). What Impacts Early Language Skills? Effects of Social Disparities and Different Process Characteristics of the Home Learning Environment in the First 2 Years. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.557751

Banai, K., & Ahissar, M. (2018). Poor sensitivity to sound statistics impairs the acquisition of speech categories in dyslexia. *Language, Cognition and Neuroscience*, *33*(3), 321–332. https://doi.org/10.1080/23273798.2017.1408851

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berry, M. D., & Erickson, R. L. (1973). Speaking Rate: Effects on Children’s Comprehension of Normal Speech. *Journal of Speech and Hearing Research*, *16*(3), 367–374. https://doi.org/10.1044/jshr.1603.367

Biersack, S., Kempe, V., & Knapton, L. (2005). *Fine-Tuning Speech Registers: A Comparison of the Prosodic Features of Child-Directed and Foreigner-Directed Speech*. 4.

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.46, retrieved 15 May 2021 from http://www.praat.org/) [Computer software].

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The Foundations of Spelling Ability: Evidence from a 3-Year Longitudinal Study. *Journal of Memory and Language*, *45*(4), 751–774. https://doi.org/10.1006/jmla.2000.2785

Cristia, A. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *e13265*. https://doi.org/10.1111/desc.13265

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development*, *90*(3), 759–773. https://doi.org/10.1111/cdev.12974

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Dilley, L. C., Lehet Matthew, Wieland Elizabeth A., Arjmandi Meisam K., Kondaurova Maria, Wang Yuanyuan, Reed Jessa, Svirsky Mario, Houston Derek, & Bergeson Tonya. (2020). Individual Differences in Mothers’ Spontaneous Infant-Directed Speech Predict Language Attainment in Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2453–2467. https://doi.org/10.1044/2020_JSLHR-19-00229

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*, 481. https://doi.org/10.3389/fnhum.2017.00481

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews, 81,* 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

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 Pt 2*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A. (2000). Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition. *Phonetica*, *57*(2–4), 242–254. https://doi.org/10.1159/000028477

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & 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. https://doi.org/10.1017/S0305000900010679

Foulke, E., & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, *72*(1), 50–62. https://doi.org/10.1037/h0027575

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*, 1490. https://doi.org/10.1038/ncomms2430

Ghitza, O. (2011). Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00130

Ghitza, O. (2012). On the Role of Theta-Driven Syllabic Parsing in Decoding Speech: Intelligibility of Speech with a Manipulated Modulation Spectrum. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00238

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, *32*(5), 545–561. https://doi.org/10.1080/23273798.2016.1232419

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

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

Goswami, U., Cumming, R., Chait, M., Huss, M., Mead, N., Wilson, A. M., Barnes, L., & Fosker, T. (2016). Perception of Filtered Speech by Children with Developmental Dyslexia and Children with Specific Language Impairments. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00791

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995–1019. https://doi.org/10.1007/s11145-009-9186-6

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*(16), 10911–10916. https://doi.org/10.1073/pnas.122368599

Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2010). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, *23*(2), 325–337. https://doi.org/10.1162/jocn.2010.21453

Greenberg, S. (2006). A Multi-Tier Framework for Understanding Spoken Language. In *Listening to speech: An auditory perspective* (pp. 411–433). Lawrence Erlbaum Associates Publishers.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics*, *31*(3), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

Haake, M., Hansson, K., Gulz, A., Schötz, S., & Sahlén, B. (2014). The slower the better? Does the speaker's speech rate influence children's performance on a language comprehension test? *International Journal of Speech-Language Pathology*, *16*(2), 181–190. https://doi.org/10.3109/17549507.2013.845690

Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. T. (2012). Basic Auditory Processing Deficits in Dyslexia: Systematic Review of the Behavioral and Event-Related Potential/ Field Evidence. *Journal of Learning Disabilities*, *46*(5), 413–427. https://doi.org/10.1177/0022219411436213

Hirose, K., & Kawanami, H. (2002). Temporal rate change of dialogue speech in prosodic units as compared to read speech. *Speech Communication*, *36*(1), 97–111. https://doi.org/10.1016/S0167-6393(01)00028-0

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*(2), 163–169. https://doi.org/10.1016/0167-6393(91)90039-V

Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports*, *10*(1), 15846. https://doi.org/10.1038/s41598-020-72739-4

Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at three years: A significant relationship. *Developmental Science*, *0*(ja), e12836. https://doi.org/10.1111/desc.12836

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

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Lam, C., & Kitamura, C. (2012). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science*, *15*(2), 212–221. https://doi.org/10.1111/j.1467-7687.2011.01118.x

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. https://doi.org/10.1016/j.heares.2013.07.015

Leong, V., & Goswami, U. (2015). Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0144411

Leong, V., & Goswami, U. (2017). Difficulties in auditory organization as a cause of reading backwardness? An auditory neuroscience perspective. *Developmental Science*, *20*(6), e12457. https://doi.org/10.1111/desc.12457

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*, *1*(2), 78–90. https://doi.org/10.1162/OPMI_a_00008

Leong, V., Stone, M. A., Turner, R. E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America*, *136*(1), 366–381. https://doi.org/10.1121/1.4883366

Lizarazu, M., Lallier, M., & Molinaro, N. (2019). Phase‒amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, *1453*(1), 140–152. https://doi.org/10.1111/nyas.14099

Menn, K., Ward, E., Braukmann, R., Boomen, C. van den, Buitelaar, J., Hunnius, S., & Snijders, T. (2022). *Neural Tracking in Infancy Predicts Language Development in Children With and Without Family History of Autism*. PsyArXiv. https://doi.org/10.31234/osf.io/gxpkm

Miller, J. F. (1981). *Assessing Language Production in Children: Experimental Procedures*. University Park Press.

Molinaro, N., & Lizarazu, M. (2018). Delta(but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, *48*(7), 2642–2650. https://doi.org/10.1111/ejn.13811

Molinaro, N., Lizarazu, M., Baldin, V., Pérez-Navarro, J., Lallier, M., & Ríos-López, P. (2021). Speech-brain phase coupling is enhanced in low contextual semantic predictability conditions. *Neuropsychologia*, 107830. https://doi.org/10.1016/j.neuropsychologia.2021.107830

Montgomery, J. W. (2004). Sentence comprehension in children with specific language impairment: Effects of input rate and phonological working memory. *International Journal of Language & Communication Disorders*, *39*(1), 115–133. https://doi.org/10.1080/13682820310001616985

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, Rimes, Vocabulary, and Grammatical Skills as Foundations of Early Reading Development: Evidence From a

Longitudinal Study. *Developmental Psychology*, *40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130396. https://doi.org/10.1098/rstb.2013.0396

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071–1086. https://doi.org/10.1098/rstb.2007.2160

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean Length of Utterance in Children With Specific Language Impairment and in Younger Control Children Shows Concurrent Validity and Stable and Parallel Growth Trajectories. *Journal of Speech, Language, and Hearing Research*, *49*(4), 793–808. https://doi.org/10.1044/1092-4388(2006/056)

Riding, R. J., & Vincent, D. J. T. (1980). Listening Comprehension: The effects of sex, age, passage structure and speech rate. *Educational Review*, *32*(3), 259–266. https://doi.org/10.1080/0013191800320303

Rimmele, J. M., Poeppel, D., & Ghitza, O. (2021). Acoustically driven cortical delta oscillations underpin prosodic chunking. *ENeuro*. https://doi.org/10.1523/ENEURO.0562-20.2021

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2020). Development of neural oscillatory activity in response to speech in children from 4 to 6 years old. *Developmental Science*, e12947. https://doi.org/10.1111/desc.12947

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2021). Right-hemisphere coherence to speech at pre-reading stages predicts reading performance one year later. *Journal of Cognitive Psychology*, *0*(0), 1–15. https://doi.org/10.1080/20445911.2021.1986514

Ríos-López, P., Molnar, M. T., Lizarazu, M., & Lallier, M. (2017). The Role of Slow Speech Amplitude Envelope for Speech Processing and Reading Development. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01497

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Schick, J., Fryns, C., Wegdell, F., Laporte, M., Zuberbühler, K., Schaik, C. P. van, Townsend, S. W., & Stoll, S. (2022). The function and evolution of child-directed communication. *PLOS Biology*, *20*(5), e3001630. https://doi.org/10.1371/journal.pbio.3001630

Sjons, J., Hörberg, T., Östling, R., & Bjerva, J. (2017). Articulation Rate in Swedish Child-Directed Speech Increases as a Function of the Age of the Child Even When Surprisal is Controlled for. *Interspeech 2017*, 1794–1798. https://doi.org/10.21437/Interspeech.2017-1052

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Smith, N. A., & Trainor, L. J. (2008). Infant-Directed Speech Is Modulated by Infant Feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Stehwien, S., & Meyer, L. (2021). *Rhythm Comes, Rhythm Goes: Short-Term Periodicity of Prosodic Phrasing*. PsyArXiv. https://doi.org/10.31234/osf.io/c9sgb

Surányi, Z., Csépe, V., Richardson, U., Thomson, J. M., Honbolygó, F., & Goswami, U. (2009). Sensitivity to rhythmic parameters in dyslexic children: A comparison of Hungarian and English. *Reading and Writing*, *22*(1), 41–56. https://doi.org/10.1007/s11145-007-9102-x

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, *9*(2), 182–198. https://doi.org/10.1016/0093-934X(80)90139-X

Tass, P., Kurths, J., Rosenblum, M., Weule, J., Pikovsky, A., Volkmann, J., Schnitzler, A., & Freund, H.-J. (1999). Complex Phase Synchronization in Neurophysiological Data. In C. Uhl (Ed.), *Analysis of Neurophysiological Brain Functioning* (pp. 252–273). Springer. https://doi.org/10.1007/978-3-642-60007-4_14

Ten Oever, S., & Martin, A. E. (2020). Linguistic constraints modulate speech timing in an oscillating neural network. *BioRxiv*, 2020.12.07.414425. https://doi.org/10.1101/2020.12.07.414425

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639. https://doi.org/10.1121/1.4807565

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340. https://doi.org/10.3758/BF03196290

Tree, J. E. F. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, *34*(6), 709–738. https://doi.org/10.1006/jmla.1995.1032

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93–118. https://doi.org/10.1515/lp-2013-0005

Vanvooren, S., Poelmans, H., De Vos, A., Ghesquière, P., & Wouters, J. (2017). Do prereaders' auditory processing and speech perception predict later literacy? *Research in Developmental Disabilities*, *70*, 138–151. https://doi.org/10.1016/j.ridd.2017.09.005

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, *142*(4), 1976–1989. https://doi.org/10.1121/1.5006179

Werker, J. F., & Hensch, T. K. (2015). Critical Periods in Speech Perception: New Directions. *Annual Review of Psychology*, *66*(1), 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

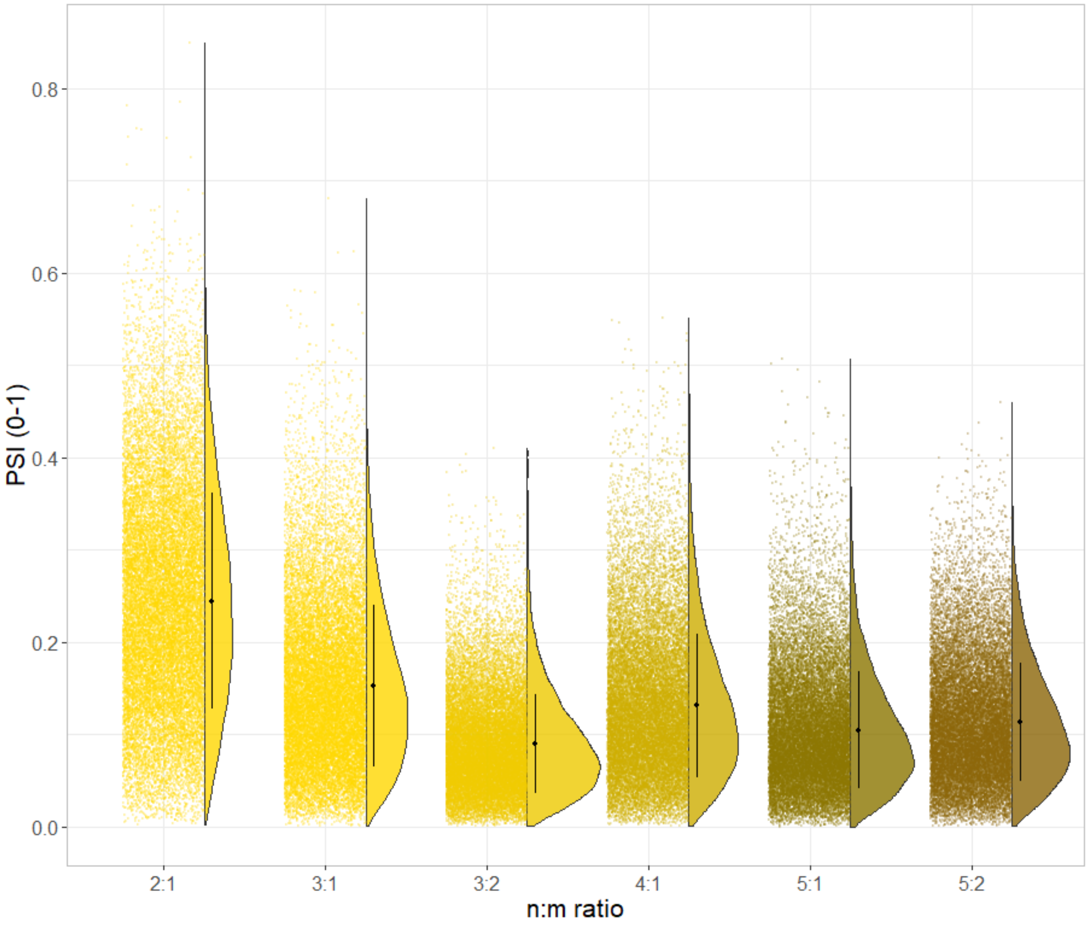
Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(2), 230–246. https://doi.org/10.1037/h0084224

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*(1), 147–162. https://doi.org/10.1016/j.cognition.2006.03.006

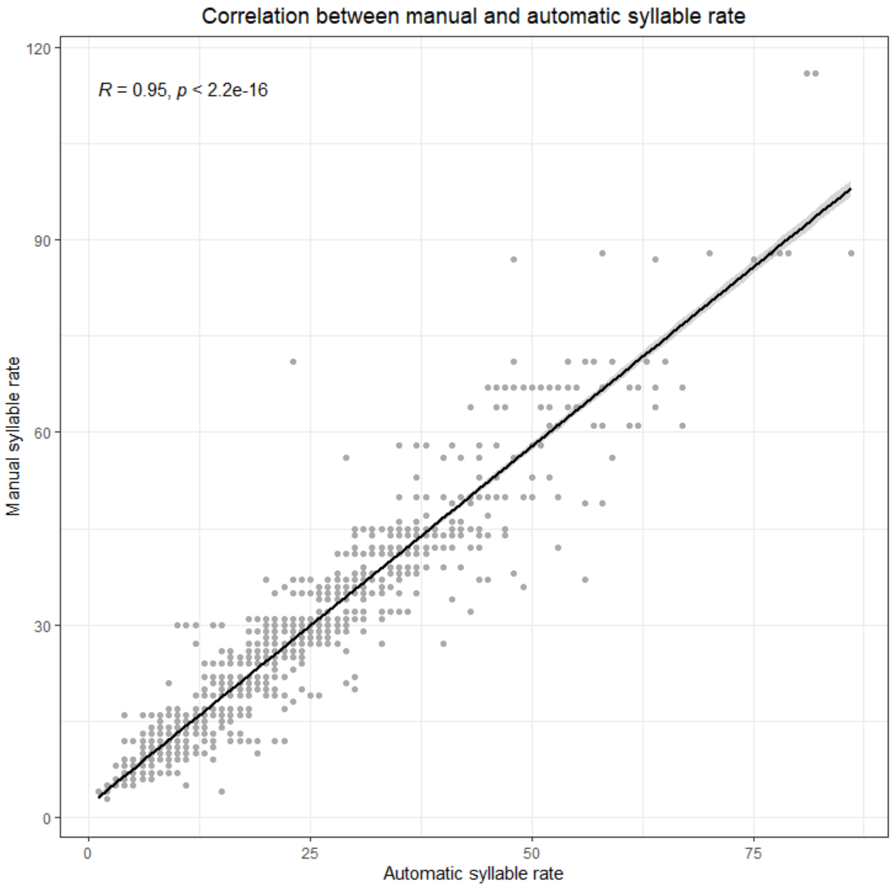
Supplemental materials

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

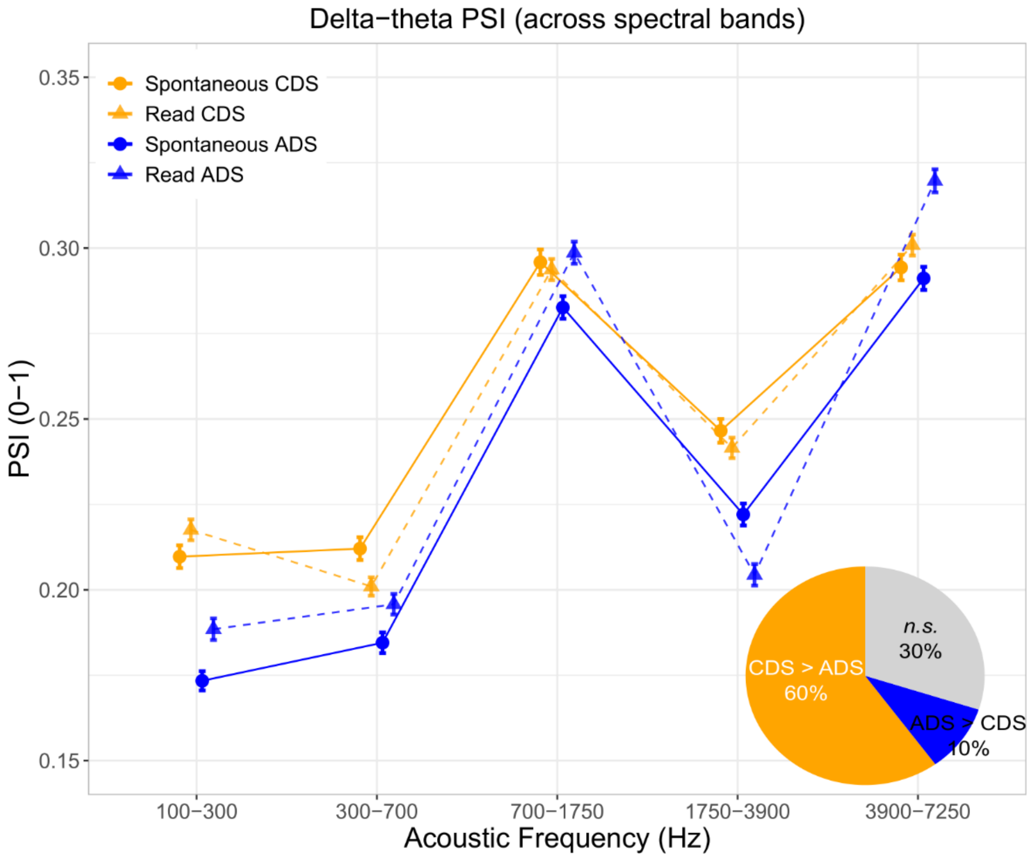
S1. Formula of cross-frequency phase synchronization index (PSI).



S2. Delta-theta PSI values across different n:m ratios. The point within each distribution represents the mean, and the bars represent standard deviations.



S3. Pearson correlation between manual and automatic syllable rate computation. Correlation index at the top left corner.



S4. Delta-theta PSI across speaking conditions and spectral bands. Error bars represent standard error of the mean. Pie plot in the bottom right corner represents the percentage of Tukey HSD contrasts for which PSI is higher for CDS (orange), ADS (blue), or resulted in a non-significant difference (gray, adjusted p-threshold = .05).