Universidad del País Vasco
Euskal Herriko Unibertsitatea

BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

Behavioral adaptation to the temporal organization of speech: Implications for phonological and reading abilities

By Laura Díaz-Muñoz Manzanares

Master´s thesis supervised by:
Marie Lallier
Jose Pérez-Navarro

San Sebastián, September 2022

Abstract

Recent evidence suggests the importance that the speech envelope has for the acquisition of phonology and the comprehension of speech. It has also been proposed that sensitivity to stimuli statistics allows us to anchor and integrate events across long time intervals to process upcoming stimuli in a more efficient manner. Most of the research in this field has focused on acoustic features that are relevant for speech processing, such as tones or pitch. However, the effect of perceptual anchoring to temporal regularities in connected speech (such as those present in the amplitude modulation of the speech envelope) is to date unexplored. The main aim of the current study was to explore behavioral adaptations to different temporal regularities of speech over time, in order to shed light on the extent to which listeners exploit temporal speech regularities for efficient speech comprehension. Additionally, we explored whether individual differences in phonological and reading skills were related to the behavioral adaptations mentioned above. The results show that any departure from natural speech reduces intelligibility. However, after minimal exposure, the comprehension of the less intelligible form of speech improves, which highlights the flexibility of our envelope tracking system. Finally, we observed that this phenomenon was linked to better phonological short-term memory and reading skills.

Keywords: Perceptual anchoring; speech comprehension; isochronous speech; natural speech; amplitude modulations; phonological abilities

Table of contents

1. Introduction

In everyday life, we encounter speakers with different accents that challenge our understanding of the conversation. Despite this, we experience an adaptation over time to the particular characteristics of speech and comprehension becomes easier. The rapid and gradual adaptation to speech variations highlights the flexibility of our perceptual system, which exploits the regularities present in the environment to continuously create hypotheses about the possible timing of upcoming sensory events (Nobre et al., 2007). Thanks to the repetition of events, we are able to learn about the frequency with which certain events are repeated in relation to others and extract patterns from them, a process known as statistical learning (Saffran et al., 1996). Although statistical learning is a general learning mechanism to which human learners are highly sensitive (Romberg & Saffran, 2010), it was first identified as a key element in infant lexical acquisition (Saffran, 2003). The same process has also been proposed as an important learning ability used in phonological and broader speech comprehension development (Gomez & Gerker; 2000; Maye et al., 2002; Kuhl, 2004; Werker et al. 2012; Banai & Ahissar, 2017).

There is evidence that infants implicitly analyze the statistical distribution of sounds that they hear in their native language and use these statistics to establish phonetic categories of that particular language (Maye et al., 2002; Kuhl, 2004). This type of statistical learning has been named "distributional learning" (Werker & Yeung, 2005; Werker et al., 2012) and is required for categorical speech perception. Thanks to the categorization of speech sounds, infants can group perceptually distinct sounds into the same phoneme category showing no sensitivity to intermediate sounds, even if they belong to different dimensions, such as talkers or contexts (Kuhl, 2004). At the same time, the quality of these speech sound representations plays a critical role in the acquisition of letter-to-sound associations required for the development of literacy (Ramus, 2003; Vellutino et al., 2004). Thereby, sensitivity to statistical regularities in speech is considered an early marker of normal reading development and is linked to reading skills (Gabay et al., 2015; Treiman et al., 2018; Banai & Yifat, 2012; Moav-Scheef et al., 2015).

Several independent observations have suggested that people with dyslexia are less sensitive to stimuli repetitions and thus have difficulties when learning their regularities, which may impair speech categorization. Most of these studies have only focused on speech categories, as they were atypically acquired by individuals with dyslexia

(Serniclaes et al., 2001; Bogliotti, 2003; Horlyck et al. 2012). However, some theoretical accounts proposed that this deficit is not speech specific, but rather could be due to general temporal processing impairments (Nittrouer et al., 2011; Vandermosten et al., 2011). This is exemplified in the work undertaken by Tallal (1980), who reported that reading problems were related to a general auditory deficit in temporal processing and not from a specific phonological deficit. Tallal designed a paradigm based on a two-tone identification task that required participants to identify tone pairs as high/low or low/high. The performance of children with reading disorders was worse than in controls, especially when the time interval between the two tones was brief, as it occurs when processing phonological units such as syllables in connected speech.

In order to understand these findings in relation to the statistical learning deficits in dyslexia when processing speech, Ahissar and colleagues (2006) designed a new version of Tallal's task. Their first paradigm included a reference stimulus that was presented in every trial, while the second paradigm did not contain a reference stimulus, which made the comparison between stimuli necessary in every trial. Their results showed that in the second paradigm (non-reference stimuli) dyslexics and control performed similarly when comparison was required. However, in the first paradigm (reference stimuli), only controls performed significantly better. Taken together, these observations suggest that people with dyslexia were unable to benefit from using a small, consistent stimulus set and improve their performance over time. Therefore, Ahissar et al.'s concluded that dyslexia might give rise to difficulties in forming perceptual anchors, a process that is tightly linked to the concept of statistical learning of regularities, which allows subjects to process more efficiently upcoming speech stimuli (Ahissar, 2006, 2007).

The study described above is the basis for the "anchoring deficit" hypothesis of dyslexia (Ahissar, 2007). This hypothesis proposes that individuals with developmental reading disorders will not benefit from stimuli regularities over time to the same extent as controls. Specifically, the "anchoring deficit" hypothesis proposes that the implicit memory of people with dyslexia decays faster than that of good readers. This faster decay reduces the temporal window over which they build refined representations of highly frequent stimuli, yielding an impoverished acquisition of speech categories, as exposed in recent empirical studies (Jaffe-Dax et al., 2017; Jaffe-Dax & Ahissar, 2017). Thus, people with dyslexia have shown difficulties processing and integrating the regularities

of stimuli across long time intervals, but they would not have a general deficit in statistical learning.

The anchoring deficit hypothesis has provided a useful framework for characterizing speech processing deficits that may arise in developmental dyslexia. However, its empirical accounts were restricted to the processing of acoustic stimuli such as single tones and syllables that detach to a considerable extent from speech as it takes place in natural contexts. For this reason, our study aims to extend the perceptual anchoring theory to connected speech perception.

Another more recent research field has contributed to our understanding of the speech deficits observed in dyslexia, through studies on the functional role of brain oscillatory activity in processing connected speech. Several theories propose that ongoing cortical oscillations synchronize to external stimuli at diverse timescales, such a process has been described as neural entrainment (Giraud & Poeppel, 2012; Ghitza, 2011) whereby brain oscillatory activity aligns to and tracks the timing of an external stimulus of close rhythmicity. In the particular case of connected speech, brain oscillatory activity entrains to the amplitude modulations (AMs) of the speech signal. AMs are fluctuations of acoustic intensity that create quasi-regular rhythm modulations coinciding with the rate of occurrence of critical linguistic units, for example, phonemes occur approximately every 80 ms, syllables appear at rates of roughly 200 ms, and stressed syllables at around 500 ms (Ghitza & Greenberg, 2009; Goswami, 2011). These speech variations of energy over time (pressure) give rise to the amplitude envelope pattern (Rosen, 1992), which coincides to an extent with the frequency of predominant brain oscillations at rest. Relatedly, speech comprehension is linked to the tracking of delta and theta rate AMs, as shown by electroencephalography (EEG) and magnetoencephalography (MEG) experiments (Doelling et al., 2014; Peelle & Davis, 2012; Molinaro & Lizarazu, 2018). For example, Doelling et al. (2014) removed the temporal fluctuations that occur at the syllabic rate, which impacted comprehension negatively. By artificially restoring the syllabic information, comprehension was regained to the level of natural speech.

In view of these studies, during the last years, researchers have explored which acoustic cues in the speech envelope are crucial to process for efficient speech comprehension. Although all the information contained in the amplitude envelope plays an important part in speech perception, multiple studies have pointed to the slow rhythms

(below 8 Hz) as the main contributors to speech intelligibility (Wood et al., 2009; Hickok & Poeppel, 2007; Abrams et al., 2008; Giraud & Poeppel, 2012; Ríos Lopez et al., 2017; Ghitza, 2011). It has been proven that the slow amplitude fluctuations that mark stress and syllabic rhythms help to segment syllables and are correlated with the acquisition of adequate phonological representations (Goswami, 2011; Leong et al., 2011; Goswami et al., 2014; Lallier et al., 2017, 2018). Indeed, studies with infants have shown that they use slow speech rhythms to segment the speech signal into its constituents (Ramus, et al. 2000) and to differentiate among languages (Mehler et al., 1988). In the same vein, there is evidence that links the use of child-directed speech, which contains enhanced temporal structure (rhythm), with the acquisition and support of a phonological system (Leong et al., 2014; Leong et al., 2017; Pérez-Navarro et al., 2022). Thus, when speaking to infants and children, mothers pattern their syllables more regularly providing salient acoustic markers relevant to extracting prosodic and syllabic phonological units, essential for language development.

Supporting the idea that the perception of speech rhythms is related to the acquisition of solid phonological representations, it has been consistently found that dyslexic children and adults are less sensitive to the slow amplitude changes (rise times) in the linguistic signal due to atypical neural entrainment to low-frequency components of speech (Goswami et al., 2002; Goswami, 2011; Pasquini et al., 2007; Lizarazu et al., 2015; Molinaro et al., 2016; Lallier et al., 2017). For example, Goswami et al. (2002) found that individuals with reading impairments had difficulties in rhythm detection. In particular, they identified significant differences between dyslexic and normally reading children in amplitude envelope onset detection, i.e., children with dyslexia showed poorer amplitude modulation beat detection than controls. With the aim of finding the underlying causes of the temporal processing difficulties found in dyslexics, Goswami (2011) proposed the *temporal sampling framework* that provides a neurological basis for rhythmic and phonological deficits in dyslexia. This theory suggests that the core deficit of dyslexia would arise from atypical tracking of the speech amplitude envelope at low-frequency bands corresponding to prosodic information, stressed syllables, and syllables.

The dyslexia data has offered a new perspective on temporal structure and speech processing, stressing the relevance of amplitude modulations of the speech envelope for the acquisition of phonology. Given the hypothesis that temporal AM regularities are critical for speech comprehension, then variations in intelligibility may result from

changes in the distributional temporal properties of AM in speech, and also affect perceptual anchoring mechanisms. However, it is still unclear how different temporal organizations of the speech envelope could affect speech comprehension. Recently, Aubanel and Schwartz (2020) have characterized the role of different temporal regularities on the intelligibility of speech in noise for French and English. They hypothesized that isochrony in speech, the division of speech units in equal time intervals (Lehiste 1977; Pike 1945), could present some advantages. In isochronous speech, the speech units are presented at a regular pace, which would make them more predictable and lead to maximum entrainment, alleviating the need for constant phase-reset subprocesses. In their experiment, they compared isochronous speech with natural speech timing, which has been also suggested to play an important role in speech intelligibility (with departures from natural timing being detrimental for speech recognition). They included a third condition: anisochronous speech, in which the timing of speech units bear the same degree of temporal distortion from naturally and isochronous timed speech. Their results go against the hypothesis that isochronous speech is easier to track, as isochronous speech was always less intelligible than naturally timed speech. However, syllabic isochronousness played an additional beneficial role since the more isochronous naturally timed syllables were, the better the sentence was recognized. These results inform the aims of the present study and suggest that perceptual isochrony (i.e., temporal regularity) in speech could play a role in speech intelligibility, and therefore in the perceptual anchoring that takes place during connected speech processing.

2. The current research work

2.1. Research aims

In the section above, it has been described the importance that the speech envelope has for the acquisition of phonology and intelligibility of speech. It has also been suggested that the sensitivity to stimuli statistics allows us to anchor and integrate events across long time intervals to process upcoming stimuli in a more efficient manner. Most of the research in this field has focused on acoustic features that are relevant for speech processing, such as tones or pitch. However, the effect of the perceptual anchoring of AM speech modulation regularities during connected speech is to date unexplored. The main aim of this study was to characterize behavioral adaptations - as indexed by speech comprehension - to changes in the regularities of AM speech distribution over time, in

order to shed light on the extent to which listeners learn and exploit temporal speech regularities to serve efficient speech comprehension. Given that perceptual anchoring is impaired in dyslexic individuals, we also wanted to explore initially whether phonological and reading abilities would be related to individual differences in the process of perceptual anchoring to the timing of speech.

The aims explained above led us to materialize three main research questions to which we will try to give an answer along this Master's thesis:

1) How does an individual adapt to diverse temporal organizations of speech for efficient comprehension?

2) Does such adaptation benefit from the repetition over time of different AM distributional properties of speech to boost comprehension?

3) How is this adaptation related to an individual's phonological and reading skills?

In order to do so, we used the three experimental speech conditions used in Aubanel and Schwartz (2020) (i.e., isochronous, anisochronous and natural) in a speech in noise task and monitored the speech comprehension skills of the listeners over the course of 198 trials to measure perceptual anchoring mechanisms. We hypothesized that, at the start of the experiment, natural speech would show the greatest intelligibility, reflecting the participant's expertise with this type of speech. Additionally, we predicted that across time, isochronous speech would lead to higher speech comprehension indexes as participants learn to harness enhanced temporal regularities. Furthermore, we expected participants with better phonological and reading skills to experience greater benefits across time from the isochronous condition because of better temporal anchoring skills.

3. Methodology

3.1. Participants

Twenty-one participants (15 females) took part in the present study (*mean age*= 24.47; *SD*= 3.38). Eighteen additional participants were tested but their responses were not recorded due to technical failure. All participants were native speakers of Spanish and reported no hearing difficulties. As the Basque Country is a bilingual region, where both Spanish and Basque are official languages, we controlled for their proficiency in Spanish,

the target language in this experiment. Thus, to assess their Spanish language skills, they completed the Basque, English and Spanish Test dataset (BEST) (de Bruin et al., 2017), which includes information from four different subtests divided in three language blocks (Basque, English and Spanish). The subtests are the following: An interview rated from 1 (lowest level) to 5 (native or native-like level) by a native speaker of Spanish and Basque; a lexical decision task; a picture-naming task and a short questionnaire about their language history and knowledge of the three languages. Participants obtained a high percentage of correct responses (*mean*= 99.86 and *SD*= 0.44) in the Spanish subtest of BEST, which indicates they were fluent speakers of Spanish.

The Basque Center on Cognition Brain and Language (BCBL) ethical committee approved the experiment (following the principles of the Declaration of Helsinki). All participants were recruited from the internal BCBL database and were remunerated for their voluntary participation in the study. Moreover, they gave their written consent at the beginning of the experimental session.

3.2. Stimuli and procedure

Experimental task

The aim of this task was to characterize behavioral adaptations as indexed by speech comprehension (dependent variable) to changes in the regularities of temporal organization of speech over time. We adapted the experimental paradigm created by Aubanel and Schwartz (2020), where utterances were divided into three different temporal organizations. The first condition was naturally produced sentences (NAT) which, in the case of our study, were obtained from the Harvard corpus of Spanish (Sharvard, Aubanel et al., 2014). This resource provides 700 phonemically balanced Spanish sentences inspired by the original Harvard corpus in English (Rothauser et al. 1969) along with its recordings from a female native peninsular Spanish talker. Each sentence contains five keywords that were used for scoring. The same sentences from the Sharvard corpus were used for the isochronous condition (ISO). In this case, the naturally produced sentences were retimed to an isochronous form by accelerating or slowing down adjacent speech portions corresponding to the syllables, creating a syllabic regular timing. A third condition of anisochronous sentences (ANI) was included, in which syllables were temporally modified to create an irregular rhythm of speech. For the anisochronous condition, we randomly assigned to each syllable the same compression factors that were

used to retime natural into isochronous speech. Finally, sentence stimuli were noise-masked to shift comprehension to below-ceiling levels, which is particularly relevant for the case of natural speech given the ease with which simple sentences can be comprehended without noise.

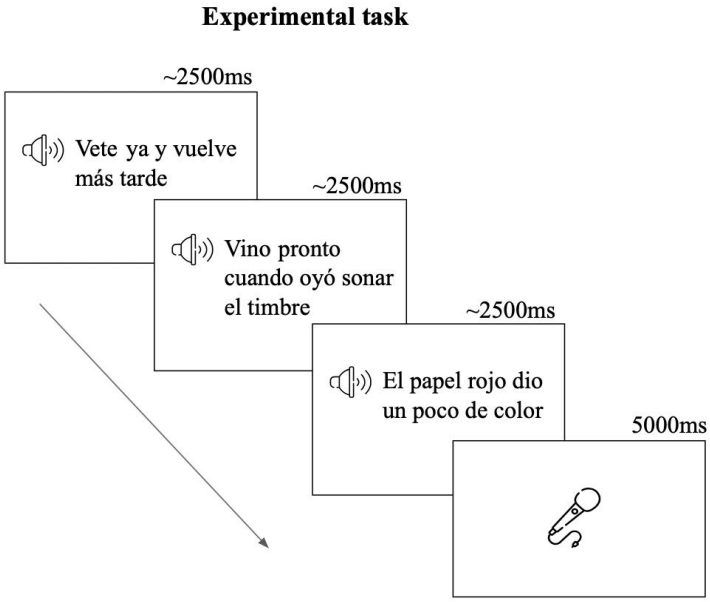

Figure 1. Structure of one trial in the experimental task. In this example, only three sentences are presented. The actual task included trials with three and four sentences.

Regarding the procedure, first, we conducted a pilot study with 10 adult native speakers of Spanish with three different levels of noise masking (i.e., speech-to-noise ratios of 0, -1, and -2 dB). The main aim of this pilot was to achieve below-ceiling comprehension across the 3 experimental conditions and test the paradigm behaviorally. The selected level of noise for the experiment was -1 dB. Then, the experimental task took place. Sentences were presented auditorily to the participants while their electrophysiological brain activity was recorded with magnetoencephalographic (MEG). For the purpose of this Master thesis, the MEG data was not used. The stimuli consisted of 231 sentences per condition, with a total of 693 sentences that were divided into 66 trials per condition and were presented to the participants in a pseudo-randomized order. The purpose of this pseudorandomization was to avoid having adjacent trials from the same condition which could add a confounding variable, producing uncontrolled learning effects and influencing the results. The trials (Figure 1) consisted of three or four, this variation was established in order to avoid participants anticipating the sentences they

were required to repeat. After one trial, participants were asked to repeat the last sentence that they listened. We took the proportion of recognized keywords as an index of speech comprehension. In the case of this study, comprehension was used as an indicator of successful cortical processing, following the established link between intelligibility and entertainment (Peelle et al., 2013).

Phonological processing and reading assessment

After the experimental task, participants were presented with three different tasks in order to assess their phonological and reading abilities. They were tested in a quiet room in a session that lasted less than half an hour. The phonological and reading tasks were the following:

- **Reading fluency**

In order to assess individual reading ability in our sample, participants were asked to read a short meaningless Spanish text called "La Alondra", an adaptation of the French Alouette reading test (Lefavrais, 1965). The stimulus was presented on a A4 paper (21 cm×29.7 cm) and consisted of 278 words that were organized in an unconventional manner, i.e., it was syntactically correct but not semantically coherent. The aim of this text is to assess mechanistic reading that cannot rely on the support of semantics, but more purely on the visual and phonological decoding of the script.

Participants were instructed to read as fast and correctly as possible. Time (in seconds) and the number of errors were recorded. Reading fluency was defined by the number of correct words read per second, thus the following reading index was calculated:

$$Reading\ fluency\ = \frac{Total\ number\ of\ correctly\ read\ words}{Total\ time\ (seconds)}$$

- **Phonemic deletion (phonemic awareness)**

In this task, participants were listening over speakers to a pseudoword together with one specific phoneme present in that pseudoword. They were instructed to repeat pseudowords removing such target phonemes. There was a total of 26 pseudowords with three, four or five syllables, which respected Spanish phonotactic rules. The novel words

were recorded by a female native Spanish speaker. The number of correct answers was recorded and converted into percentages. Accuracy was coded as: 1 – correct response, 2 - wrong word (e.g., 'estuño' instead of 'restuño').

- **Pseudoword repetition (phonological short-term memory)**

Participants listened to 24 pseudowords over speakers and were instructed to repeat them as well as they could. Items varied from four to nine syllables (four words of four, five, six, seven, eight and nine syllables) and their structure respected the Spanish phonotactic rules. The words were recorded by a female native Spanish speaker. The number of correctly repeated pseudowords was recorded and converted into percentage. Accuracy was coded as: 1 – correct response, 2 - wrong word (e.g., 'fibonegáforo' instead of 'fibolenugáforo').

Outlier treatment

We used the interquartile range (IQR) method to determine which scores from the phonological and reading data were outliers. Observations that were more than 1.5 IQR below the first quartile or more than 1.5 IQR above the third quartile were considered an outlier and removed from the following analyses. Thus, two participants were removed from the phoneme deletion task and one from the reading task.

4. Results

We used linear mixed effect (LME) models to assess the influence of different speech temporal organizations (i.e., natural, isochronous, anisochronous, —our experimental manipulations), time on task (i.e., trial index within the main experimental task), and phonological and reading tasks on speech comprehension. We included each participant as a random intercept in the models to meet the within-participant design of our study. We used the *lmer* function of the lme4 package (v.1.1.28, Bates et al., 2015) to fit the LMEs. Additionally, we used the *anova* function (in base R) to test the omnibus main effects and interactions of our predictors.

Given that we were looking for anchoring learning effects, we first observed whether they were increasing trends in comprehension across the experiment. Overall performance (i.e., across all conditions) showed a general decrease at approximately the 50th trial of each condition (Figure 2). The observed decreasing trends may be an effect

of time on task since the total duration of the experimental paradigm was approximately two hours, which may have a negative impact on participants' attention resources causing task disengagement. Therefore, we limited our further analyses to the first 75% items of each condition in order to be able to assess our hypothesized learning/anchoring effects.

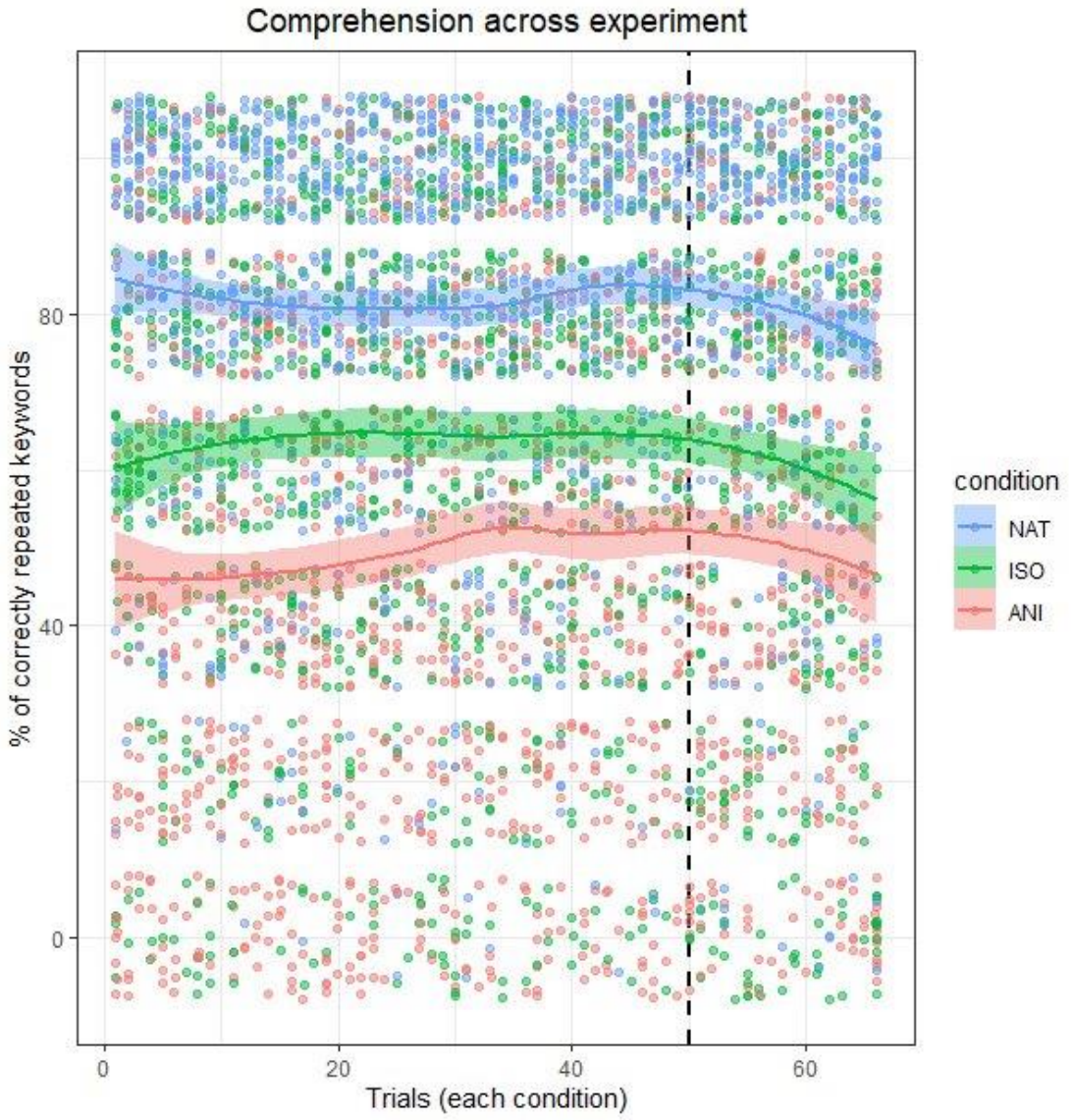

Figure 2. Comprehension across the 66 trials per condition (NAT, ISO, ANI) in the experimental task. The vertical black dotted line in the 50th trial marks the point where there is a general decline in performance across the three conditions. Colored points represent each trial per participant. The colored fitted lines represent the locally estimated regression of performance as a function of time on task per condition.

4.1. General differences between NAT, ISO, and ANI

There was a significant effect of speech temporal organization on participants' comprehension (i.e., the mean percentage of correctly repeated keywords per sentence), $F(2, 40) = 318.75$, $p < 0.001$. Such an effect of the different temporal conditions was due to natural speech conveying a higher comprehension than isochronous, $t(40) = 15.06$, $p < 0.001$ ($\beta = 17.98$, SE = 1.19, CI [15.57 20.39]), and anisochronous speech, $t(40) = 27.59$, $p < 0.001$ ($\beta = 32.95$, SE = 1.19, CI [30.54 35.37]). Also, participants' comprehension was significantly more accurate in isochronous than anisochronous condition, $t(40) = 12.54$, $p < 0.001$ ($\beta = 14.97$, SE = 1.19, CI [12.56 17.38]). This particular effect is visible in Figure 3, which shows intelligibility results as the proportion of keywords correctly recognized by participants.

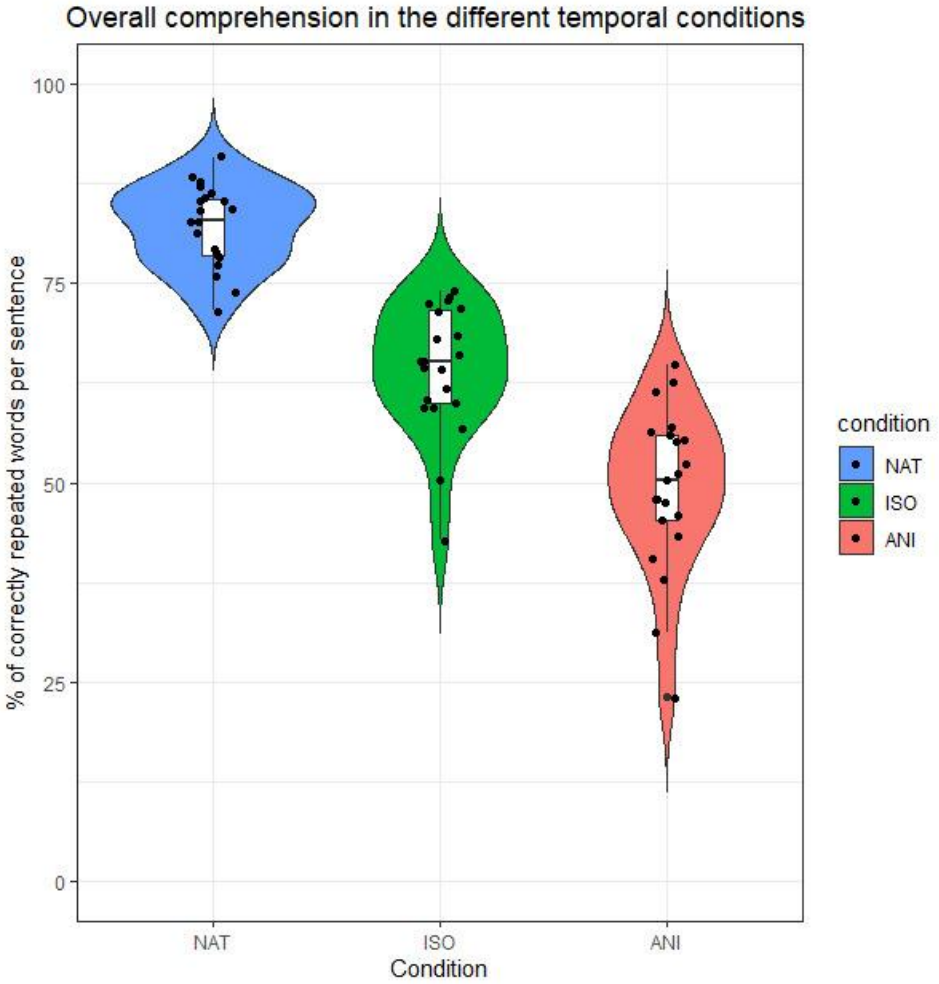

Figure 3. Overall comprehension within the different temporal conditions (NAT, ISO, ANI). The violin plots show the distribution of comprehension (in blue for the natural condition, green for the isochronous condition and red for the anisochronous condition).

The median is represented with a horizontal black line within the boxplot, the interquartile range as the white box and whiskers. Black points mark each participant's mean percentage of correctly repeated keywords in each condition.

4.2. Anchoring along the experiment (performance as a function of time within each condition).

We did not find a significant effect of time on task on overall speech comprehension (i.e., across all temporal conditions), $F(1, 280.53) = 0.22$, $p > 0.05$. Next, we followed the same modeling approach as in the first analysis (section 3), i.e., participants as random intercepts and time on task as random slopes, to assess changes in comprehension within each temporal condition across the experiment. There was indeed a significant interaction of time on task by condition, $F(3, 2605.2) = 161.97$, $p<0.001$.

To assess further the underlying causes of such interaction, we fitted three independent submodels to assess the size and direction of the relationship between time on task and comprehension within each particular temporal condition. There was no significant effect of time on task on comprehension for the natural condition, $F(1, 177.41) = 0.47$, $p>0.05$, neither for the isochronous condition, $F(1, 19.448) = 0.81$, $p>0.05$. However, there was a significant effect for the anisochronous condition, $F(1, 129.65) = 5.85$, $p = 0.017$, showing that participants' comprehension increases with time within the anisochronous condition (see Figure 4).

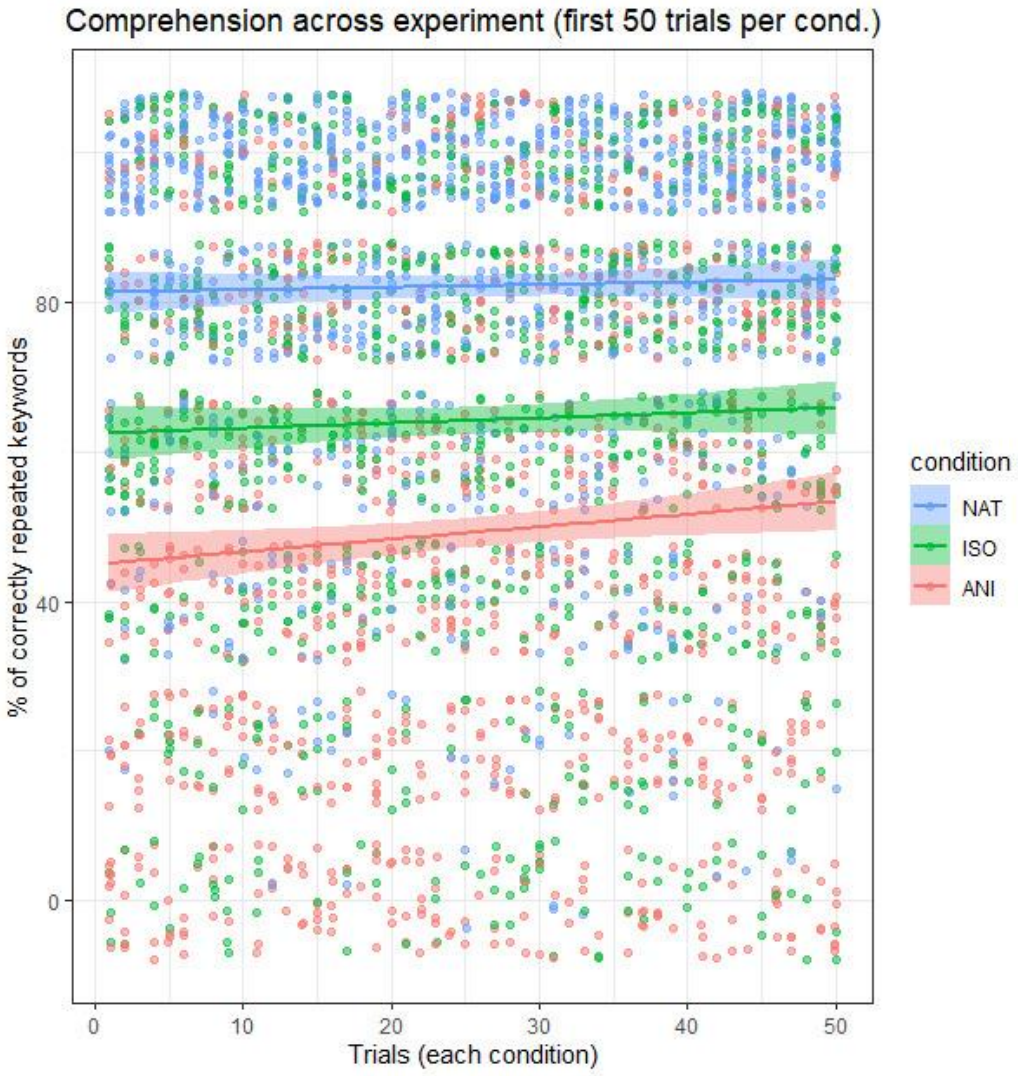

Figure 4. Comprehension across the experiment within the temporal conditions (NAT, ISO, ANI). This figure shows the first 50 trials per condition. The colored lines represent the linear fit for all participants. Colored points represent the mean of the trials across conditions for each participant.

4.3. Is there a relationship between phonology, reading and (overall) performance in each condition?

Pseudoword repetition task

We assessed whether there was a relationship between pseudoword repetition and comprehension across the three temporal conditions (i.e., the mean comprehension of each participant in each condition in the whole experiment, i.e., average across all trials independently of the time point in the task). There was indeed a significant relationship between pseudoword repetition and the experimental conditions, F (3, 29.869) = 89.34,

p<0.001. This was due to the natural condition being the most robustly linked to pseudoword repetition, t(20.451) = 4.92, p < 0.001 (β = 44.01, SE = 8.95). There was no significant interaction between the isochronous, t(20.541) = 1.957, p > 0.05 and anisochronous conditions, t(20.451) = -0.44, p > 0.05 and pseudoword repetition. Therefore, participants that had better performance in pseudoword repetition performed significantly better in the natural condition but not necessarily in the two other conditions (Figure 5).

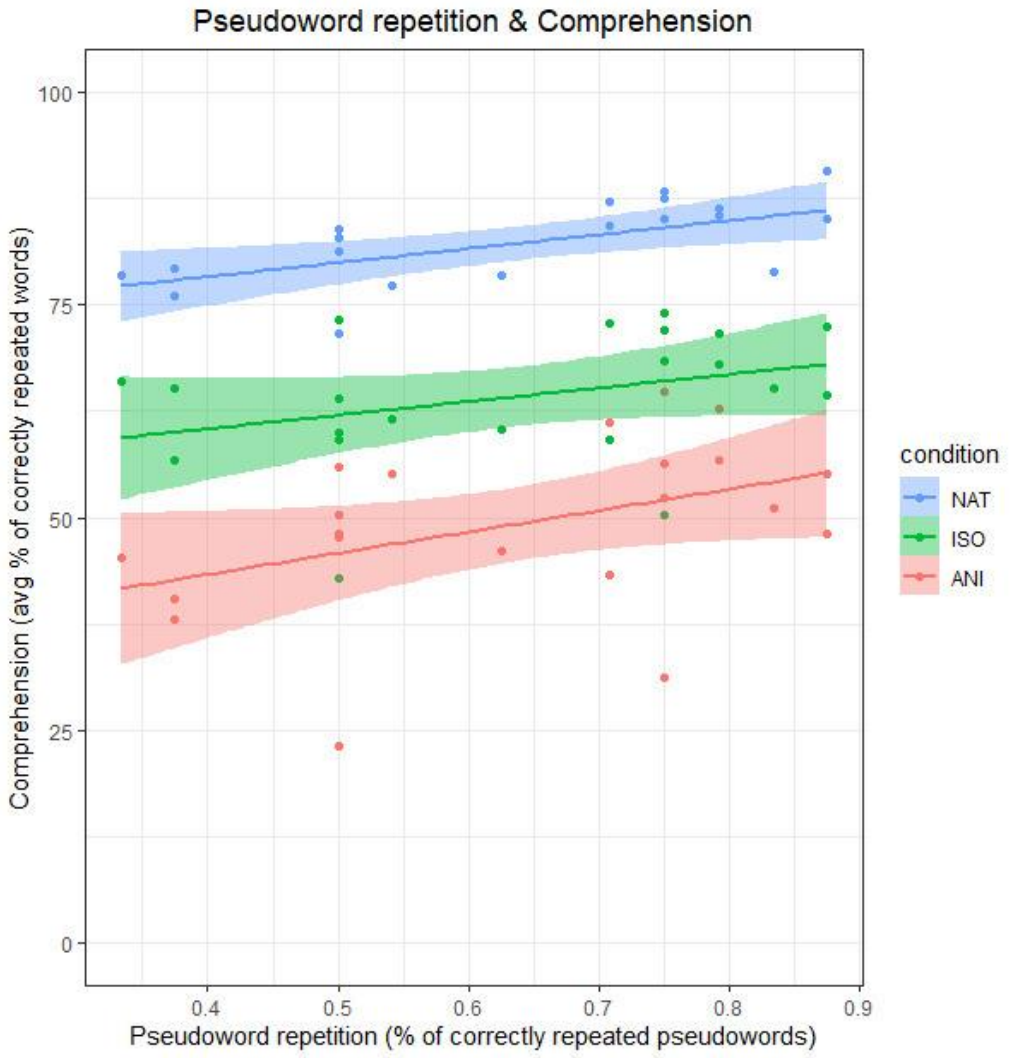

Figure 5. Horizontal lines represent the fit of a linear model for each of the conditions between pseudoword repetition and comprehension. Colored points mark the mean score for each participant in comprehension and pseudoword repetition.

Phoneme deletion task

We also evaluated whether there was a relationship between the scores obtained in the phoneme deletion task and the different temporal conditions when predicting better

performance in comprehension. A significant interaction between phonemic deletion and the temporal conditions was found, F (3, 29.721) = 85.798, p<0.001. In this case, phoneme deletion was positively correlated with the natural condition, t(20.280) = 3.57, p < 0.001 (β = 29.38, SE = 8.22), i.e., participants with a better comprehension of the natural condition showed also a better performance in the phoneme deletion task (see Figure 6). In contrast, such effect was not observed in the isochronous condition, t(20.280) = 0.82, p > 0.05 (β = 6.73, SE = 8.22) nor the anisochronous condition, t(20.280) = -1.40, p > 0.05 (B = -11.52, SE = 8.22).

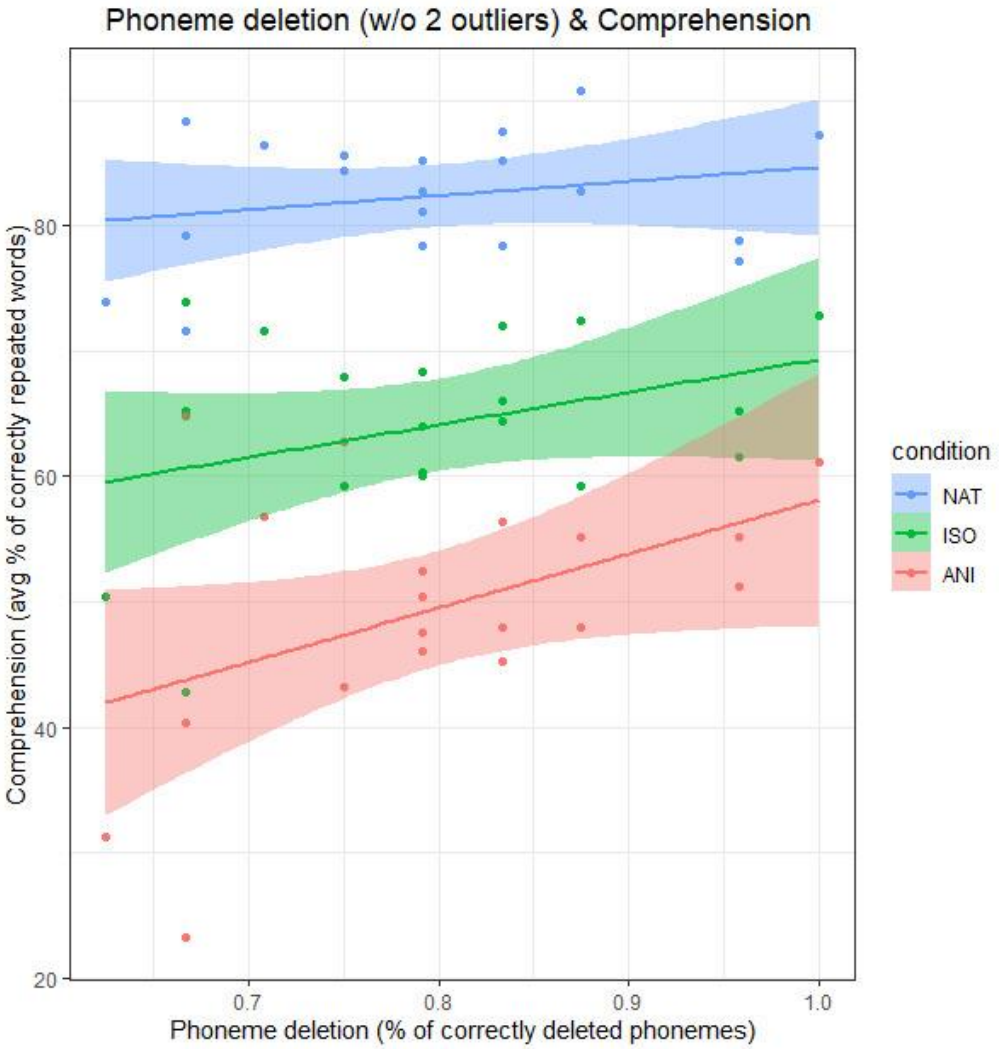

Figure 6. Horizontal lines represent the fit of a linear model for each of the conditions between phoneme deletion and comprehension across conditions. Colored points mark the mean score for each participant in comprehension and phoneme deletion. In this figure, two outliers were removed based on their score with respect to a group central value.

Reading task

Similarly, there was a significant interaction between reading and temporal conditions when predicting comprehension, $F(3, 27.674) = 138.13$, $p<0.001$. Again, this interaction was driven by reading scores being significantly related to comprehension in the natural condition, $t(18.283) = 2.55$, $p < 0.001$ ($\beta = 10.77$, $SE = 4.23$) (Figure 7). As in the phonological tasks, reading scores did not predict significantly the comprehension in isochronous, $t(18.238) = 0.86$, $p >0.05$ ($\beta = 3.63$, $SE = 4.23$) nor anisochronous conditions, $t(18.238) = -0.56$, $p >0.05$ ($\beta = -3.36$, $SE = 4.23$).

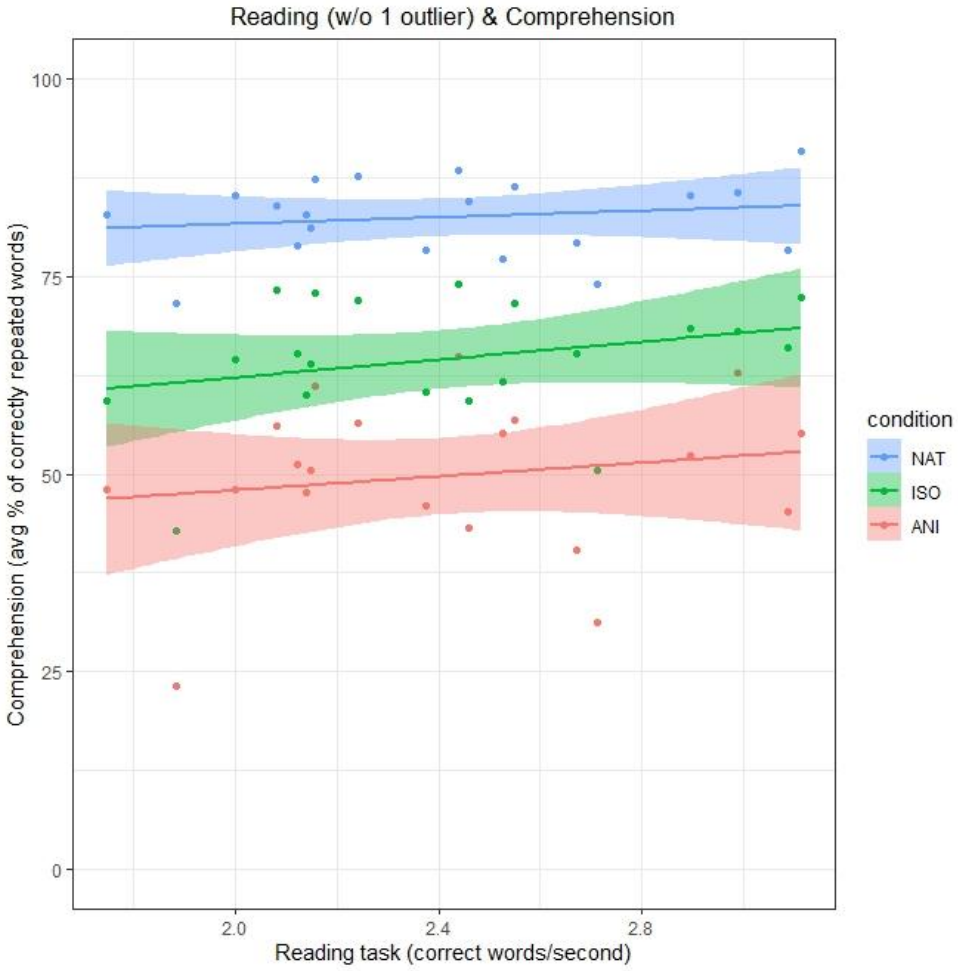

Figure 7. Horizontal lines represent the fit of a linear model for each of the conditions between reading and comprehension across conditions. Colored points mark the mean score for each participant in comprehension and reading. In this figure, one outlier was removed based on the interquartile range method.

4.4. Is there a relationship between phonology and the anchoring that takes place as a function of time for each speech condition?

To specifically test our hypotheses on perceptual anchoring mechanisms, we then explored whether phonological and reading skills were related to changes in performance *across time* for each of the three speech conditions. Perceptual anchoring was indexed by the slope of the comprehension performance as a function of time for each participant and condition: steeper slopes reflect bigger changes in performance across time, with bigger positive values translating into stronger perceptual anchoring.

Pseudoword repetition

We assess whether there was a relationship between pseudoword repetition and the anchoring that takes place across the experiment for each temporal condition. There was indeed a significant relationship between both, $F(3, 59) = 11.87$, $p < 0.001$. This is due to the anisochronous condition being the most robustly linked to pseudoword repetition $t(59) = 2.65$, $p = 0.01$ ($\beta = 0.17$, $SE = 0.06$). These main effects are visible in Figure 8, which shows that the participants that learned better from the anisochronous condition were the ones with better performance in pseudoword repetition. However, there was not a significant interaction between anchoring in the isochronous, $t(59) = 0.24$, $p > 0.05$ and natural conditions, $t(59) = -0.73$, $p > 0.05$ and pseudoword repetition.

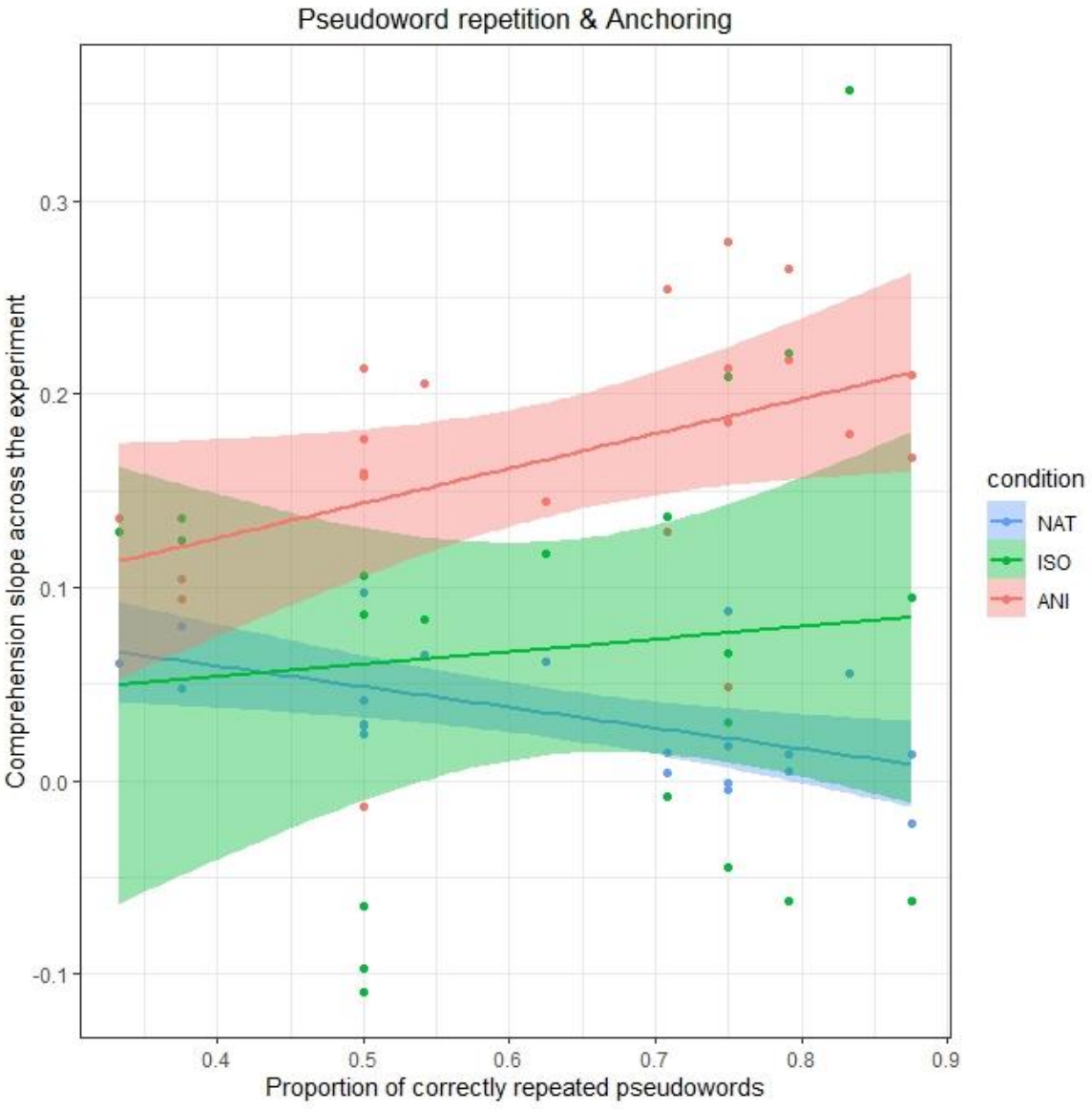

Figure 8. Horizontal lines represent the fit of a linear model for each of the conditions between pseudoword repetition and anchoring across conditions. Colored points mark the mean score for each participant in anchoring and pseudoword repetition.

Phoneme deletion task

Next, we analyzed the relationship between phoneme deletion and the different temporal conditions when predicting learning across the experiment. Although there was indeed a significant relationship between both factors, $F(3, 53) = 10.23$, $p<0.001$, none of the conditions was positively correlated with anchoring across the experiment. i.e., it is not possible to predict the performance in the phoneme deletion task through the anchoring effect of the different temporal conditions in the experimental task. In figure 9, we can observe increasing trends for the anisochronous condition, which are related to

its marginally significant results, t(53) = 1.87, p = 0.07 (β = 0.19, SE = 0.10). In contrast, not such effect is observed for the isochronous condition, t(53) = 0.69, p > 0.05 (β = 0.07, SE = 0.10) and the natural condition, t(53) = 0.20, p > 0.05 (B = 0.021, SE = 0.10).

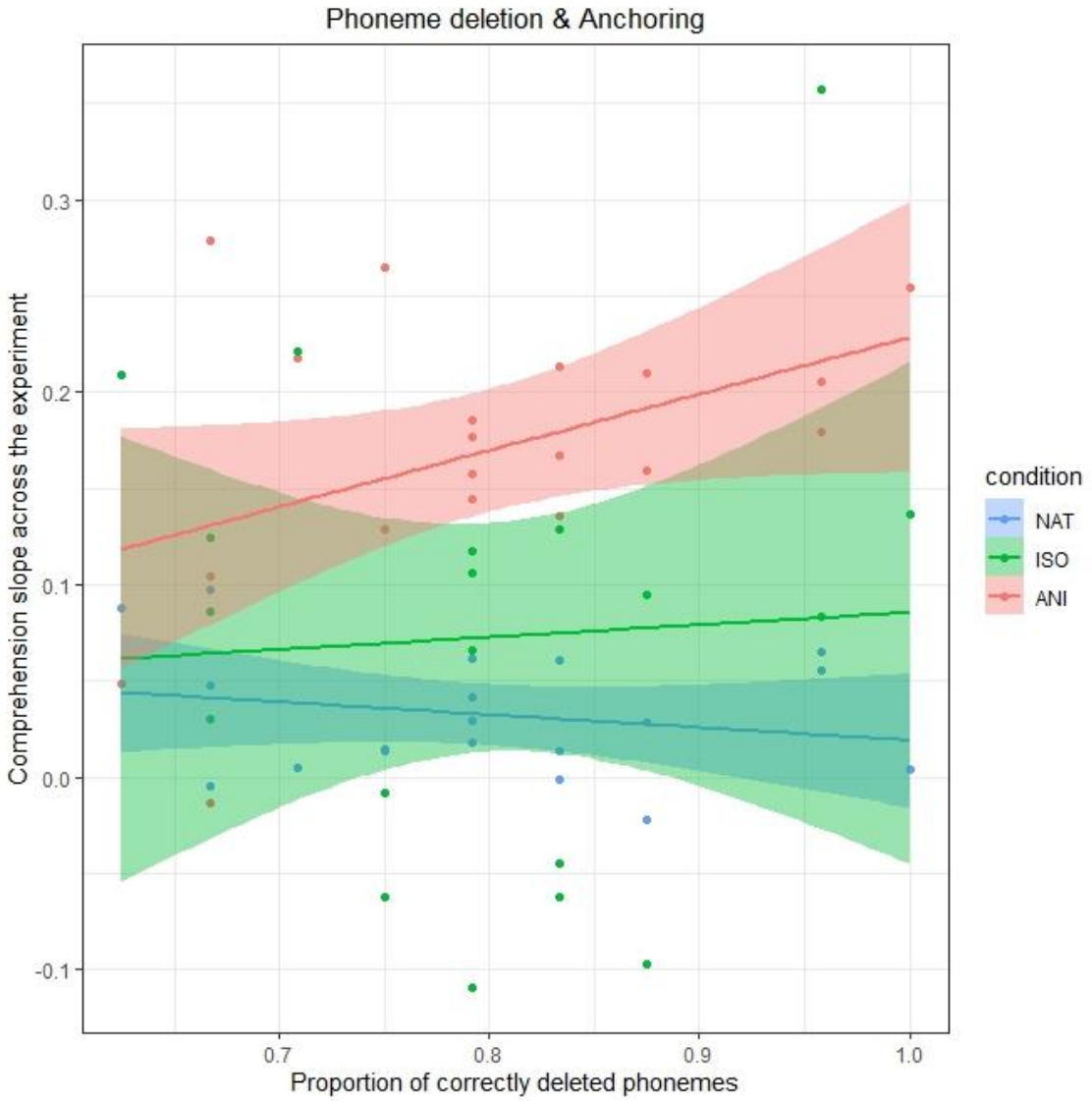

Figure 9. Horizontal lines represent the fit of a linear model for each of the conditions between the proportion of correctly deleted phonemes and the comprehension slope across the experiment. Colored points mark the mean score for each participant in anchoring and phoneme deletion.

Reading task

We also assess whether there was a relationship between the scores obtained in the reading task and the different temporal conditions when predicting anchoring effects. There was indeed a significant relationship between both, F (3, 56) = 10.66, p<0.001. This

is due to a significant effect between the anisochronous condition and reading t(56) = 2.198, p = 0.03 (β = 0.06, SE = 0.02). These main effects are visible in Figure 10, which shows that the participants that learned better from the anisochronous condition were the ones with better performance in reading. However, there was not a significant relationship between reading and anchoring in the isochronous, t(56) = 0.71, p > 0.05 and natural conditions, t(56) = 2.20, p > 0.05.

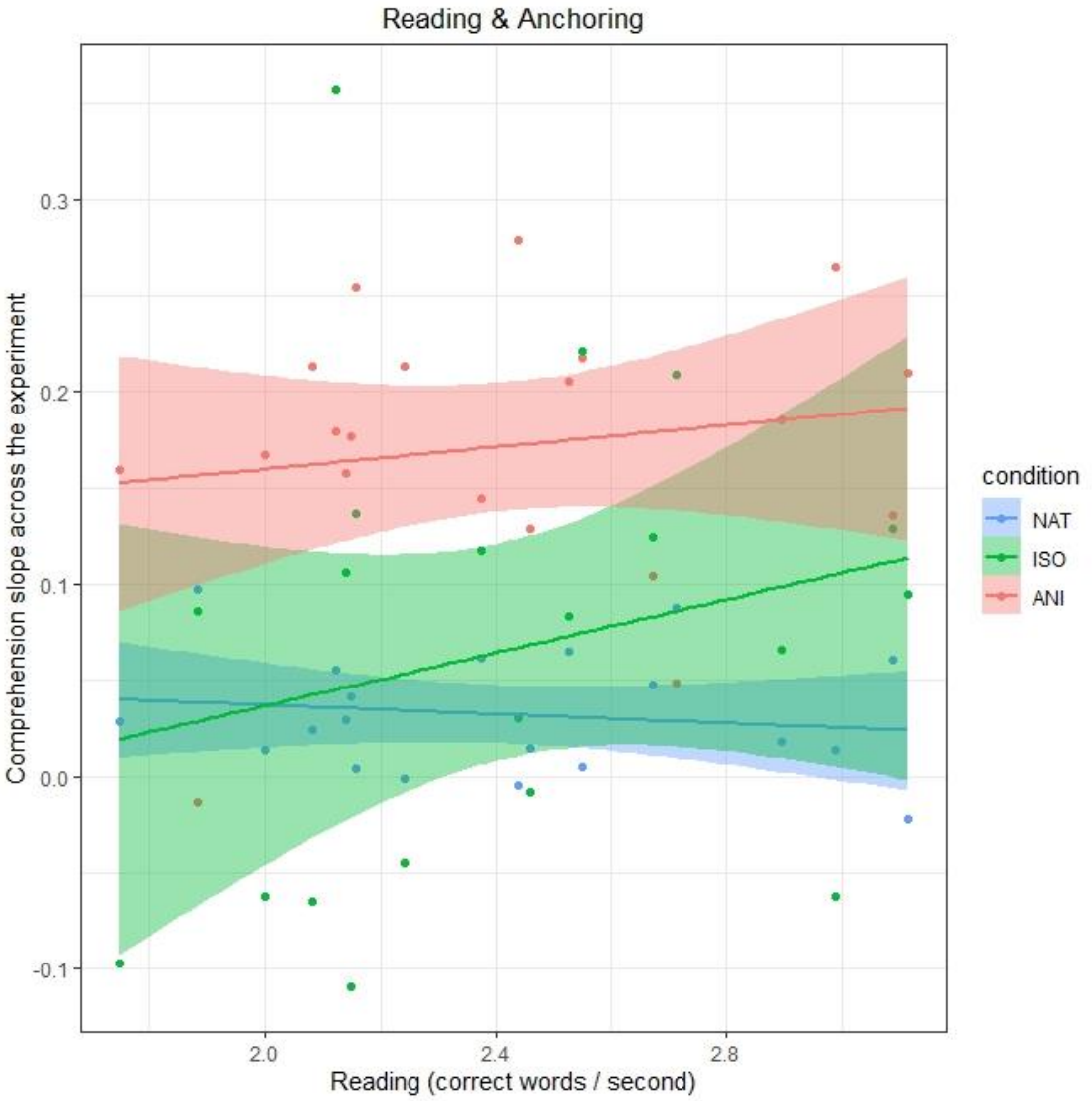

Figure 10. Horizontal lines represent the fit of a linear model for each of the conditions between the performance in the reading task and the comprehension slope across the experiment. Colored points mark the mean score for each participant in anchoring and reading.

5. Discussion

The main aim of this study was to explore behavioral adaptations, indexed by comprehension of speech presented in noise, to different temporal regularities of speech over time. Namely, our intention was to provide an empirical framework on how listeners learn and exploit the temporal regularities of speech for efficient comprehension. Additionally, we explored whether individual differences in phonological and reading skills were related to the behavioral adaptations mentioned above. To this end, 21 adult native speakers of Spanish completed two sets of tasks. The aim of the main experimental task was to characterize behavioral adaptations to different temporal organizations of speech over time. For this purpose, we presented sentences in three different speech conditions characterized by different regularities in the distribution of speech AM: natural, isochronous, and anisochronous speech. The second set consisted of three tasks that were used to assess the phonological and reading abilities of the participants. The aim was to determine whether performance on these tasks could predict the anchoring that took place across the speech-in-noise comprehension task, i.e., the slope of the comprehension for each participant and condition as a function of time.

In the following sections, we will discuss the answer to the three research questions raised in the Introduction focusing on the integration of our results within current neurophysiological accounts of speech perception and reading development. The results should be interpreted with caution due to the limitations of the current research that are also discussed in this chapter. Finally, we expose several recommendations for future research.

1) How does an individual adapt to diverse temporal organizations of speech?

In our first analyses, we examined the effect of three different temporal distributions of speech AMs on comprehension. The results showed that, in Spanish, natural speech conveys an overall higher intelligibility than isochronous and anisochronous speech, replicating the findings from Aubanel and Schwartz (2020) in English and French. These results provide evidence that is contrary to an extent to the isochronousness hypothesis (Abercrombie, 1987; Cummins, 2012), which suggests that isochronous speech would have a beneficial effect on intelligibility due to the division of speech units into equal duration intervals. This hypothesis proposes that thanks to isochrony, syllables would

become more predictable and lead to maximum entrainment, alleviating the need for constant phase resetting (Peelle & Davis, 2012). By contrast, our data showed the detrimental effect that both isochronous and anisochronous temporal distortions from naturally timed speech has on intelligibility. In line with our findings, different works on speech rhythm claim that it is not acoustic isochrony that guides speech comprehension, but perceptual isochrony (Studdert-Kenney, 1980; Brown et al., 2011; Dilley et al. 2012). Acoustic isochrony refers to the production of phonological units at approximately equal time intervals (Lehiste, 1977), instead perceptual isochrony is the phenomenon through which listeners perceive those units as occurring more regularly than they actually do. This could explain that the use of an isochronous speech did not benefit speech comprehension in our experiment even in adverse environments, such as speech in noise. However, the use of perceptually isochronous speech (natural condition) did play a role in speech comprehension, probably helping listeners to identify where stressed syllables occurred. In this context, there is also substantial evidence that listeners rely on acoustic information at the syllable level contained in slow amplitude modulations (Ghitza & Greenberg, 2009; Aubanel & Cooke, 2013; Peelle & Davis, 2012; Doelling et al., 2014; Molinaro & Lizarazu, 2018). Therefore, any alteration of the speech rate at the syllable structure would reduce speech intelligibility, as has been shown in our results. Taking all these studies together, we suggest that the temporal statistics of natural speech are actively used by listeners to encode essential information and are a key factor in speech processing.

Another vein of research that can provide a new insight into our findings suggests that natural speech is best recognized due to the two-stage speech perception process. In this approach proposed by Ghitza (2011), the first stage of speech perception includes the template-like temporal registration of the speech signal, while the second stage involves the use of that information registered to access already stored representations. Thus, it could be possible that under the first stage, isochronous speech gains an advantage over natural speech due to its boosted temporal regularities. However, in the second stage, the access to listeners' knowledge about timing statistics learned through repeated exposure to spoken language, makes natural speech offset the advantage of isochronous timing. This knowledge of the natural timing patterns would allow listeners to make predictions about the ongoing speech stream, while in the case of isochronous speech the recognition will suffer as the signal does not match the well-established pattern.

Finally, our results also showed that participants' comprehension was significantly more accurate overall for the isochronous than for the anisochronous form of speech. Again, the results are similar to the findings in Aubanel and Schwartz (2020) to an extent, as they only observe an advantage of the isochronous speech in the English condition. However, no such difference was observed in French. It may be the case that English, as a stress-timed language, takes a greater benefit from the isochronous condition, making its syllables more similar to syllable-timed languages such as Spanish and French. Thus, participants could better track acoustic information at the syllabic level, which has been suggested to be a key factor in successful speech comprehension. We should also take into consideration the different experimental designs, sample size and analysis procedures that could have contributed to these discrepancies.

2) Does the temporal adaptation to the AM temporal regularities of speech boost comprehension after repeated encounters?

One of the aims of our study was to explore the effect of perceptual anchoring (Ahissar, 2006) to AM speech modulations regularities, a mechanism that has not been explored yet in the temporal domain of speech, nor in its connected usual form. Additionally, this process is tightly linked to statistical learning, which allows individuals to adapt and process upcoming speech stimuli in a more efficient manner. We hypothesized that isochronous forms of speech would present an advantage in comprehension when repeated over time due to their enhanced temporal regularities that enable cortical mechanisms to harness them. We also hypothesized that anisochronous sentences would convey a challenge to our comprehension and no improvement with time, as their syllabic temporal statistics are irregular and randomly distributed across sentences. However, contrary to our hypothesis, we only observed a significant positive effect of time on task on the anisochronous speech, as it was the only condition that led to significantly increased comprehension at the end relative to the beginning of the experiment. These results suggest that anisochronous speech produces the greater learning effects from the three conditions. Based on the findings of other studies (Wightman & Kistler, 2005; Schmidt et al., 2021), our results could be attributed to the need to allocate greater resources with less intelligible speech. For example, Schmidt and colleagues (2021) manipulated speech intelligibility using noise-vocoded speech and showed that low levels of intelligibility in speech required a more dominant track of the modulation rate (acoustic information) of the acoustic envelope, shifting away from the

syllabic rate (linguistic information). On the contrary, intelligible speech would produce a greater synchronization of neural activity to the amplitude envelope of speech and, consequently, would be easier to track. In regard to the natural and isochronous forms of speech, it is possible that participants avoid the use of the same tracking mechanisms that were used in the anisochronous condition, due to their high levels of intelligibility. According to the results, the comprehension percentage of both conditions at the beginning of the experimental task was above 50% of correct answers, in contrast with the anisochronous condition (below 50% of correct answers).

Individuals´ adaptation to anisochrony in connected speech is still unexplored to date however, our findings could be compared to similar results in adaptation to foreign accents. There is existing evidence that listeners readily adapt to foreign accents after minimal exposure, despite the dissimilarity to native speech (Trude et al. 2013). Although comprehension of foreign accents is initially taxed due to the discrepancies with the typical acoustic characteristics associated with the language (Schirru & MacKay, 2003), various studies have shown that performance improves after minimal exposure (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, et al., 2009). This ability to adapt to variability in speech perception, highlights the flexibility of the envelope tracking system, demonstrating that repeated exposure to regular stimuli, as in the isochronous conditions, is not required to experience an adaptation to some particular characteristics of speech. We suggest that sensitivity to stimuli statistics plays a role in this process, allowing our envelope tracking system to anchor to the more salient units of speech and integrate them across long time intervals, making the processing of upcoming stimuli more efficient.

In conclusion, contrary to our hypothesis, participants did not exploit the temporal regularities of the isochronous condition over time, as was expected. Instead, the anisochronous condition was the form of speech that led to maximum learning throughout the experiment.

3) How is the adaptation to speech temporal statistics related to an individual's phonological and reading skills?

We evaluated whether anchoring to the different temporal conditions across the experiment could be predicted by phonological and reading abilities. We observed a relationship between the overall performance of participants in the experimental task and

their phonological and reading abilities. Specifically, we found that phoneme deletion, pseudoword repetition and reading were robustly linked to the overall comprehension of the natural condition. This could mean that participants who process more efficiently natural speech in noise, have better mechanisms to exploit the timing statistics of this natural condition to which they are constantly exposed and in which they are experts. Thanks to their ability to recognize and manipulate the sound structure of speech in Spanish, they would be able to perform better on the different phonological and reading tasks. In fact, it has been suggested that phonological awareness and reading may be related to adults and children's ability to understand speech in noise (Dole et al., 2012; Ziegler et al. 2009; Lewis et al., 2010). Supporting this idea, it has been consistently found that dyslexic children and adults are less sensitive to the slow amplitude changes (rise times) in the linguistic signal due to atypical neural entrainment to low-frequency components of speech (Goswami et al., 2002; 2011). Thus, impaired AM processing could be one other cause of the speech-in-noise deficit observed in dyslexia. This could suggest why individual differences in phonological and reading skills explain the degree to which participants benefit from the natural condition.

In addition, we assessed whether there was a significant correlation between phonology and the anchoring effect across the experiment. Indeed, pseudoword repetition and reading were significantly correlated to the anchoring that took place in the anisochronous condition. However, no significant relation was found between anchoring in any of the conditions and the phoneme deletion task. A possible explanation for these results is that the pseudoword repetition task measures phonological short-term memory and, similarly, the repetition of sentences in noise during the experimental task requires the use of phonological short-term memory to remember and repeat the last sentence that the participants listened. Therefore, we propose that subjects with better phonological short-term memory took more advantage and learned better from the anisochronous condition over time. Moreover, we should also consider that pseudoword repetition is a phonological task that targets syllables, however, the phoneme deletion task targets phonemes. As was mentioned in the introduction, acoustic information at the syllabic level plays an important role in the successful comprehension of speech. In that way, participants who have better phonological awareness at the syllabic level, show better anchoring to the anisochronous form of speech, which is the condition posing more difficulty and the one that requires a greater focus on syllable-rate information. However,

we should also consider that a higher number of participants could result in a significant relationship between phoneme deletion and anchoring in the anisochronous condition, as the results were marginally significant. In addition, the relation between anchoring and reading performance could also be explained if we take into account the multifactorial nature of reading (Peyrin et al., 2012), where a successful awareness of syllables plays an important role in reading decoding.

Finally, we have not been able to prove that participants with better phonological and reading skills experience greater benefits across time from the isochronous condition, our initial hypothesis. It could be the case that participants with better phonological and reading skills do not take advantage of the isochronous condition due to an initially enhanced performance over this condition, as it was mentioned before.

6. Limitations and futures directions

The empirical results reported here should be considered in light of some limitations that could be addressed in future research. Our modest sample size is one of the most salient limitations of the current study. Although our initial sample size was 39 participants, due to technical issues 18 subjects were removed. In order to draw valid and more precise conclusions, we will continue collecting new data in the future.

Another caveat was the experimental manipulation, which was targeted toward the temporal dimensions of speech but did not take into account linguistic characteristics, such as word frequency or lexical and syntactic predictability. While we selected our stimuli from a corpus that is particularly designed for psycholinguistic research in which sentences are controlled for semantic and syntactic complexity (Sharvard, Aubanel et al., 2014), our results are to an extent subject to biases and confounding linguistic effects that may have influenced our model estimates. Hopefully in the short-run, we will be able to analyze whether these linguistic features could have an impact on comprehension.

Furthermore, we also believe that our results would be best complemented by rhythm metrics to quantify the departure, for each sentence in any temporal condition, from naturalness, isochrony or anisochrony at the syllable level, e.g., the more isochronous naturally timed syllables are, the better the sentence is comprehended. These analyses could have best modelled the relations between our variables to predict the

comprehension and anchoring effects of the temporal conditions across the experiment. Therefore, we cannot discard that our findings regarding a better overall comprehension of the natural condition across the experiment (Section 4.3) have something to do with a robust isochrony within the syllables of the natural condition.

Due to time limitations, the MEG data that was collected during the experimental task is not included in this Master´s thesis. In the future, this data will allow us to explore whether cortical tracking of speech benefits from specific temporal regularities over time and will extend our conclusions on how cortical adaptation to different temporal conditions is related to individual differences in phonological and reading abilities. In addition, these analyses could provide partial support to theories claiming that entrainment to the slow rhythms leads to the formation of proper phonological representations and reading (Goswami, 2011, 2018; Lallier et al., 2018, 2017).

Furthermore, in order to explore the benefits of temporal regularities in speech processing, we believe that it would be necessary to modify the block design created for this experiment. Although we designed trials that were presented in a pseudo-randomized order to avoid uncontrolled learning effects, it may be that the precaution we took could have masked any potential anchoring effect for the isochronous condition. Therefore, it should be considered to design blocks where each temporal condition is repeated along their respective block. Thus, it would be possible to alleviate the need for constant phase resetting between conditions and could allow future studies to observe greater learning effects thanks to maximum entrainment.

Finally, once we conceive an experimental design that provides more informative evidence about the perceptual anchoring hypothesis and given that this mechanism has been suggested to be impaired in dyslexic individuals, the ultimate goal of this study would be its implementation in different populations with developmental reading disorders. More specifically, these future directions could not only give us new insights into how people with dyslexia learn and adapt to different temporal regularities in speech across time, but also could be translated into educational practice, inspiring new approaches and helping literacy instruction. In fact, there are already several behavioral programs (Cogo-Moreira et al., 2011; Kraus & Chandrasekaran, 2010; Patel, 2011) that have obtained acceptable results in improving phonological awareness in children

through the training of different rhythms, proving that the perception of slow rhythms in speech is a key factor in the development of appropriate phonological awareness and reading.

7. Conclusions

This Master's thesis yields novel evidence on the role of different temporal regularities on speech comprehension over time. To our knowledge, the set of experiments that constitutes this work is the first to explore the mechanisms of perceptual anchoring to AM speech modulation regularities in the temporal domain of connected speech. Our interpretation of the results speaks first and foremost for the advantage of naturally timed speech versus isochronous and anisochronous formats in comprehension, suggesting that any departure from the natural speech will reduce intelligibility. Furthermore, the findings of anchoring to anisochronous speech highlight the flexibility of our perceptual system, which rapidly learns to integrate stimuli across time even when they are irregular and randomly distributed. We suggest that low intelligibility contexts, as is the case of the anisochronous condition, require greater resources that track the amplitude modulations of this form of speech. However, further studies are needed to shed light on such effects. Finally, regarding the contribution of phonological and reading skills to the prediction of anchoring along the experiment, a possible interpretation of our results is that participants who process more efficiently natural speech in noise, have better mechanisms to exploit the temporal statistics of this natural condition in which they are experts. Moreover, our data suggest that phonological short-term memory and reading skills are linked to improved learning of anisochronous speech over time.

8. References

Abercrombie, D. (1967). *Elements of General Phonetics* (0 ed.). Columbia Univ@press.

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. *Journal of Neuroscience*, *28*(15), 3958–3965. https://doi.org/10.1523/jneurosci.0187-08.2008

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*(11), 458–465. https://doi.org/10.1016/j.tics.2007.08.015

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

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

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Aubanel, V., & Cooke, M. (2013). Strategies adopted by talkers faced with fluctuating and competing-speech maskers. *The Journal of the Acoustical Society of America*, *134*(4), 2884–2894. https://doi.org/10.1121/1.4818757

Aubanel, V., & Schwartz, J. L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76594-1

Banai, K., & Ahissar, M. (2017). 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

Banai, K., & Yifat, R. (2012). Anchoring in 4- to 6-year-old children relates to predictors of reading. *Journal of Experimental Child Psychology*, *112*(4), 403–416. https://doi.org/10.1016/j.jecp.2012.04.008

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

Bogliotti, C. (2003). Relation between categorical perception of speech and reading acquisition. *Paper presented at the 15th International Congress of Phonetic Sciences.*

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189–1196. https://doi.org/10.3758/s13423-011-0167-9

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cogo-Moreira, H., Andriolo, R. B., Yazigi, L., Brandão De Ávila, C. R., & Mari, J. (2011). Music education for improving reading skills in children and adolescents with dyslexia. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.cd009133

Cummins, F. (2012). Oscillators and Syllables: A Cautionary Note. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00364

Dilley, L. C., Wallace, J., & Heffner, C. C. (2012). Perceptual isochrony and fluency in speech by normal talkers under varying task demands. *Understanding Prosody*, 237–258. https://doi.org/10.1515/9783110301465.237

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. https://doi.org/10.1016/j.neuroimage.2013.06.035

Dole, M., Hoen, M., & Meunier, F. (2012). Speech-in-noise perception deficit in adults with dyslexia: Effects of background type and listening configuration. *Neuropsychologia*, *50*(7), 1543–1552. https://doi.org/10.1016/j.neuropsychologia.2012.03.007

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. https://doi.org/10.1016/s0167-6393(02)00128-0

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired Statistical Learning in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934–945. https://doi.org/10.1044/2015_jslhr-l-14-0324

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., & 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*(1–2), 113–126. https://doi.org/10.1159/000208934

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

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/s1364-6613(00)01467-4

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., Power, A. J., Lallier, M., & Facoetti, A. (2014). Oscillatory â€œtemporal samplingâ€• and developmental dyslexia: toward an over-arching theoretical

framework. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00904

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

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Horlyck, S., Reid, A., & Burnham, D. (2012). The Relationship Between Learning to Read and Language-Specific Speech Perception: Maturation versus Experience. *Scientific Studies of Reading*, *16*(3), 218–239. https://doi.org/10.1080/10888438.2010.546460

Jaffe-Dax, S., Frenkel, O., & Ahissar, M. (2017a). Dyslexics' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *eLife*, *6*. https://doi.org/10.7554/elife.20557

Jaffe-Dax, S., Frenkel, O., & Ahissar, M. (2017b). Dyslexics' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *eLife*, *6*. https://doi.org/10.7554/elife.20557

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

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

Lallier, M., Lizarazu, M., Molinaro, N., Bourguignon, M., Ríos-López, P., & Carreiras, M. (2018). From Auditory Rhythm Processing to Grapheme-to-Phoneme Conversion: How Neural Oscillations Can Shed Light on Developmental Dyslexia. *Literacy Studies*, 147–163. https://doi.org/10.1007/978-3-319-90805-2_8

Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2016). Amodal Atypical Neural Oscillatory Activity in Dyslexia. *Clinical Psychological Science*, *5*(2), 379–401. https://doi.org/10.1177/2167702616670119

Lefavrais, P. (1965). Description, définition et mesure de la dyslexie. Utilisation du test "L´Alouette". *Revue de Psychologie Appliquée*, 15, 33-34.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*(3), 253–263. https://doi.org/10.1016/s0095-4470(19)31139-8

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. (2016). 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., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, *64*(1), 59–73. https://doi.org/10.1016/j.jml.2010.09.003

Lewis, D., Hoover, B., Choi, S., & Stelmachowicz, P. (2010). Relationship Between Speech Perception in Noise and Phonological Awareness Skills for Children With Normal Hearing. *Ear & Hearing*, *31*(6), 761–768. https://doi.org/10.1097/aud.0b013e3181e5d188

Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P. M., Lerma-Usabiaga, G., & Carreiras, M. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Human Brain Mapping*, *36*(12), 4986–5002. https://doi.org/10.1002/hbm.22986

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111. https://doi.org/10.1016/s0010-0277(01)00157-3

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Moav-Scheff, R., Yifat, R., & Banai, K. (2015). Phonological memory and word learning deficits in children with specific language impairment: A role for perceptual context? *Research in Developmental Disabilities*, *45-46*, 384–399. https://doi.org/10.1016/j.ridd.2015.08.010

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., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *37*(8), 2767–2783. https://doi.org/10.1002/hbm.23206

Nittrouer, S., Shune, S., & Lowenstein, J. H. (2011). What is the deficit in phonological processing deficits: Auditory sensitivity, masking, or category formation? *Journal of Experimental Child Psychology*, *108*(4), 762–785. https://doi.org/10.1016/j.jecp.2010.10.012

Nobre, A., Correa, A., & Coull, J. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470. https://doi.org/10.1016/j.conb.2007.07.006

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory Processing of Amplitude Envelope Rise Time in Adults Diagnosed With Developmental Dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286. https://doi.org/10.1080/10888430701344280

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00142

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

Pérez-Navarro, J., Lallier, M., Clark, C. A., Flanagan, S. A., & Goswami, U. (2022). Local temporal regularities in child-directed speech in Spanish. *Preprints*. https://doi.org/10.31234/osf.io/4q3de

Peyrin, C., Lallier, M., Démonet, J., Pernet, C., Baciu, M., le Bas, J., & Valdois, S. (2012). Neural dissociation of phonological and visual attention span disorders in developmental dyslexia: FMRI evidence from two case reports. *Brain and Language*, *120*(3), 381–394. https://doi.org/10.1016/j.bandl.2011.12.015

Pike, K. L. (1945). *The intonation of American English*. University of Michigan Press.

Ramus, F. (2003). Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*(2), 212–218. https://doi.org/10.1016/s0959-4388(03)00035-7

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language Discrimination by Human Newborns and by Cotton-Top Tamarin Monkeys. *Science*, *288*(5464), 349–351. https://doi.org/10.1126/science.288.5464.349

Redford, M. A. (2019). *The Handbook of Speech Production*. Wiley-Blackwell.

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. *WIREs Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R. (2003). Statistical Language Learning. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schmidt, F., Chen, Y. P., Keitel, A., Rösch, S., Hannemann, R., Serman, M., Hauswald, A., & Weisz, N. (2021). Neural speech tracking shifts from the syllabic to the modulation rate of speech as intelligibility decreases. *BioRxiv*. https://doi.org/10.1101/2021.03.25.437033

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual Discrimination of Speech Sounds in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research*, *44*(2), 384–399. https://doi.org/10.1044/1092-4388(2001/032)

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306. https://doi.org/10.1121/1.3101452

Studdert-Kennedy, M. (1980). Speech Perception. *Language and Speech*, *23*(1), 45–66. https://doi.org/10.1177/002383098002300106

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

Temporal information in speech: acoustic, auditory and linguistic aspects. (1992). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

Treiman, R. (2018). Statistical Learning and Spelling. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 644–652. https://doi.org/10.1044/2018_lshss-stlt1-17-0122

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of Memory and Language*, *69*(3), 349–367. https://doi.org/10.1016/j.jml.2013.05.002

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Wouters, J., & Ghesquière, P. (2011). Impairments in speech and nonspeech sound categorization in children with dyslexia are driven by temporal processing difficulties. *Research in*

Developmental Disabilities, *32*(2), 593–603. https://doi.org/10.1016/j.ridd.2010.12.015

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*(11), 519–527. https://doi.org/10.1016/j.tics.2005.09.003

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How Do Infants Become Experts at Native-Speech Perception? *Current Directions in Psychological Science*, *21*(4), 221–226. https://doi.org/10.1177/0963721412449459

Wood, C., Wade-Woolley, L., & Holliman, A. J. (2009). Phonological awareness: Beyond phonemes. In C. Wood & V. Connelly (Eds.), Contemporary perspectives on reading and spelling (pp. 7–23). New York: Routledge.

Wightman, F. L., & Kistler, D. J. (2005a). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Wightman, F. L., & Kistler, D. J. (2005b). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732–745. https://doi.org/10.1111/j.1467-7687.2009.00817.x