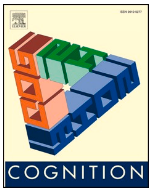
Cognition 245 (2024) 105737

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit



Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning

Ireri Gómez Varela [a,1], Joan Orpella [b,1], David Poeppel [b,c,d,f], Pablo Ripolles [b,d,e,f,2], M. Florencia Assaneo [a,*,2]

[a] Institute of Neurobiology, National Autonomous University of Mexico, Querétaro, Mexico
[b] Department of Psychology, New York University, New York, NY, USA
[c] Ernst Strüngmann Institute for Neuroscience, Frankfurt, Germany
[d] Center for Language, Music and Emotion (CLaME), New York University, New York, NY, USA
[e] Music and Audio Research Lab (MARL), New York University, New York, NY, USA
[f] Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany

ARTICLE INFO

Keywords:
Statistical learning
Auditory-motor synchronization
Speech
Prediction-based learning

ABSTRACT

Phonological statistical learning - our ability to extract meaningful regularities from spoken language - is considered critical in the early stages of language acquisition, in particular for helping to identify discrete words in continuous speech. Most phonological statistical learning studies use an experimental task introduced by Saffran et al. (1996), in which the syllables forming the words to be learned are presented continuously and isochronously. This raises the question of the extent to which this purportedly powerful learning mechanism is robust to the kinds of rhythmic variability that characterize natural speech. Here, we tested participants with arhythmic, semi-rhythmic, and isochronous speech during learning. In addition, we investigated how input rhythmicity interacts with two other factors previously shown to modulate learning: prior knowledge (syllable order plausibility with respect to participants' first language) and learners' speech auditory-motor synchronization ability. We show that words are extracted by all learners even when the speech input is completely arhythmic. Interestingly, high auditory-motor synchronization ability increases statistical learning when the speech input is temporally more predictable but only when prior knowledge can also be used. This suggests an additional mechanism for learning based on predictions not only about when but also about what upcoming speech will be.

1. Introduction

Statistical learning is the ability to implicitly extract the distributional properties of various inputs. It is a widespread phenomenon found in different cognitive domains such as vision, audition, reading, and event processing. *Phonological statistical learning* (phSL) is considered a fundamental mechanism for language acquisition, whereby a sensitivity to the transitional probabilities of syllables in continuous speech drives the learning of its constituent words (Erickson & Thiessen, 2015; Rebuschat & Williams, 2012; Romberg & Saffran, 2010).

Phonological SL studies typically comprise a familiarization phase, involving the continuous repetition of various trisyllabic pseudowords, followed by a two-alternative forced-choice test of the ability to recognize them. Several factors have been shown to modulate phSL performance. An example is prior linguistic knowledge or syllabic-level phonotactic probability; pseudowords formed by more likely syllable combinations in the participant's first language are easier to extract than those formed by less common combinations (Elazar et al., 2022; Siegelman et al., 2018). Another example is the ability to spontaneously synchronize one's own speech to isochronous auditory syllables (i.e., speech auditory-motor synchronization), a skill that appears to be bimodally distributed in the general population (Lizcano-Cortés et al., 2022). Previous work showed that, although both high and low auditory-motor synchronizers exhibited above-chance phSL, high

---

* Corresponding author.
E-mail address: fassaneo@inb.unam.mx (M.F. Assaneo).

[1] These authors contributed equally to this work.

[2] These authors jointly supervised this work.

https://doi.org/10.1016/j.cognition.2024.105737
Received 18 July 2023; Received in revised form 30 January 2024; Accepted 31 January 2024

synchronizers consistently outperformed low synchronizers (Assaneo et al., 2019; Orpella et al., 2022). Taken together, these results suggest the existence of two mechanisms supporting phSL: one related to the ability to exploit the statistical dependencies between phonological representations acquired through prior language exposure, and another related to the ability to exploit the rhythmic structure of the speech input. Consistent with this hypothesis, recent neuroimaging data implicates two distinct brain networks in phSL. On the one hand, a network comprising auditory and superior temporal cortex, whose activity correlates with phSL performance across participants (López-Barroso et al., 2015; Orpella et al., 2022), could play the role of integrating incoming speech information (e.g., syllable identity) with prior knowledge (e.g., plausible order) to form higher order representations (words). This is in line with the integrative role proposed for the superior temporal gyrus in speech perception (Bhaya-Grossman & Chang, 2022) and with recent ECoG data showing responses in the left superior temporal cortex at the rate of the pseudowords-to-be-learned emerging in the course of the phSL familiarization phase (Henin et al., 2021). On the other hand, activity in dorsal language stream areas, including frontal and inferior parietal cortex, which correlates with participants' degree of auditory-motor synchronization, has been shown to enhance the phSL of high auditory-motor synchronizers (Orpella et al., 2022). Given the implication of these areas in temporal prediction processes (Coull et al., 2011; Rimmele et al., 2018), a hypothesis is that high synchronizers leverage input rhythmicity to enhance learning through temporal prediction (Assaneo et al., 2019; Orpella et al., 2022).

Despite the various factors shown to modulate phSL, the effect of the rhythmic structure of the acoustic stimulus has not been tested systematically. Phonological SL learning studies have typically overlooked a potential role for syllable rhythmicity by presenting syllables isochronously (e.g., Lopez-Barroso et al., 2011; López-Barroso et al., 2015; Saffran et al., 1996). This choice in stimulus design is sometimes made explicitly for methodological purposes (e.g., for frequency-tagging of neural data (Getz et al., 2018; Henin et al., 2021) or pupil size (Marimon et al., 2022) synchronization metrics), but is more generally justified as a means of experimental control. Despite the remarkable temporal regularities at the syllable level across languages and speaking situations (Ding et al., 2017; Varnet et al., 2017), it is readily apparent that syllables in speech are not perfectly isochronous; when considering *specific instances* of natural speech (e.g., any single sentence vs. an average of sentences across a corpus) rhythmic variability is high (Nakatani et al., 1981). Accordingly, it is relevant to explore how phSL performance is affected when stimuli, as in natural speech, depart from perfect isochrony.

In the present study, we examined the effect of the speech input's rhythmic structure (i.e., syllabic temporal variability) on phSL and how this interacts with two other factors shown to impact phSL, individual differences in auditory-motor synchronization and how well the pseudowords to be learned adhere to statistical regularities in the participants' native language. We exposed high and low speech auditory-motor synchronizers to artificial languages with arhythmic, semi-rhythmic, or isochronous speech during learning. In addition, we manipulated the linguistic priors associated with the different artificial languages, such that the syllable order within the pseudowords to be learned had different levels of probability of occurrence in the participants' native language. If phSL is robust to temporal variability, we should observe successful phSL learning across participants (i.e., regardless of their auditory-motor synchronization abilities) even in conditions of irregular syllable temporal structure (arhythmic speech condition). Although reduced, PhSL should also remain above-chance when the words show a less probable syllabic order, indicating that learning is not simply driven by prior knowledge (i.e., syllabic-level phonotactic probability). Furthermore, the extent to which learners exploit different cues in the speech input (e.g., syllable rhythmicity and prior knowledge) can shed light on the nature of SL mechanisms. For example, high auditory-motor synchronizers may outperform low synchronizers when the input speech stream contains rhythmic cues (i.e., in semi-rhythmic and isochronous but not in arhythmic conditions) irrespective of how plausible the syllable order is. Because high synchronizers have been shown to differentially engage the dorsal language stream (Assaneo et al., 2019; Orpella et al., 2022), this could suggest the use of temporal prediction mechanisms that are thought to be channeled through this dorsal pathway (Rimmele et al., 2018) for phSL. Alternatively, if highs outperform lows only when both temporal cues and prior knowledge regarding syllable order can be leveraged, this might point to their use of predictions for learning containing both temporal and content information about upcoming speech (Orpella et al., 2021, 2020).

2. Materials and methods

2.1. Participants

A total of 300 participants, recruited from the general Mexican population through social media advertising, completed the main online study. Participants were assigned to two different groups (see Experimental design section). Half of the participants were assigned to subgroup 1 and the other half to subgroup 2. All participants were native Mexican Spanish speakers, with self-reported normal hearing and no neurological deficits. In line with previous work, we excluded participants if, during the speech-to-speech synchronization test (SSS-test), they spoke aloud instead of whispering, remained silent for >3 s, or if audio recordings were too noisy (Lizcano-Cortés et al., 2022). Due to this exclusion criteria, the data from 36 participants from subgroup 1 and 52 from subgroup 2 was removed from further analysis. A total of 212 participants were included in the final SSS-test analysis. Participants that did not clearly belong to one of the two synchronization groups were not included in subsequent analyses (see Section: "Participants categorization according to the SSS-test outcome"). As a result, 15 extra participants were excluded from subgroup 1 (subgroup 1: $N = 99$, 54 women, mean age = 28.7, $SD = 7.0$) and 12 from subgroup 2 (subgroup 2: $N = 86$, 51 women, mean age = 28.8 years, $SD = 8.9$).

An additional sample of 60 participants with the same demographic characteristics as the main cohort completed a control experiment (the data from one participant was not recorded; $N = 59$, 32 women, mean age = 26.6 years, $SD = 7.3$). Participants were recruited from an existing database of subjects that previously participated in experiments using the SSS-test. Accordingly, they were already categorized as high or low synchronizers.

All participants read and signed an informed consent and were compensated for their participation with an Amazon gift card. The protocol was designed to be completed online, and the applied procedures were approved by the XXX ethics committee of the XXXX (protocol 096.H).

2.2. Stimuli: phonotactics and synthesis

We created four different pseudo-languages (henceforth, languages L1 to L4), each consisting of 4 trisyllabic pseudowords (henceforth, *words*). We selected 48 different consonant-vowel (CV) syllables to construct 16 triplets (4 trisyllabic words x 4 languages). The triplets were not randomly generated. Instead, we controlled that the syllables were not assigned to a position within the pseudoword that is highly uncommon in the Spanish language. For example, if a given syllable rarely occurs at the beginning of a word, it would not be assigned as the first syllable of a pseudoword. We used *Syllabarium, Complete Statistics for Basque and Spanish Syllables* online application (Duñabeitia, Cholin, Corral, Perea, & Carreiras, 2010) to compute the token positional frequencies (i.e., the summed lexical frequency of the words containing the syllable in the given position) for positions 1, 2, and 3 for each of the syllables. We did not assign syllables to positions with a token frequency below 700. This value has been selected considering the positional frequency distribution for all plausible CV combinations and positions 1, 2, and 3 (see Fig. 1). With this procedure, we guaranteed a minimum alignment between the syllabic structure of our languages and the syllable statistics of the Spanish language. See Supplementary Table 1 for a complete description of the syllables composing the languages.

Each language was then scored according to how well it adjusted to the syllable distribution in the Spanish language by the following equation:

$$SSI(Lang) = \sum_{i=1}^{12} \frac{Freq(Syll_i, Pos_i)}{\sum_{j=1}^{3} Freq(Syll_i, j)}$$

with i running on the 12 syllables of the language, *Freq(x,y)* representing the token frequency for a syllable x in position y, and Pos_i is the position to which $Syll_i$ was assigned. SSI stands for Spanish Similarity Index. According to this formula, the SSI represents the summation, across all syllables composing a language, of the relative frequency of each syllable in the assigned position within the language's words, given its overall frequency of occurrence in positions one through three (accounting for the number of syllables of the pseudowords). Thus, SSI values reflect how much each artificial language resembles/departs from the Spanish syllabic structure. The distribution of this index for 1000 randomly selected languages (i.e., randomly selecting 4 trisyllabic words without syllable repetition) is a normal distribution centered at 4 with a standard deviation of 0.8. Given that we constrained syllable selection for our languages to token frequencies above 700, the four languages we selected are characterized by average to high SSI values ($SSI_{Lang1} = 4.6$, $SSI_{Lang2} = 5.6$, $SSI_{Lang3} = 4.8$, $SSI_{Lang4} = 5.8$). Based on preliminary data, we deemed this variability sufficient to assess how this factor interacts with speech rhythmicity and auditory-motor synchronization skills, while ensuring significant learning.

For each language, a familiarization pseudoword stream was generated by randomly combining the 4 words with no gap between them and no consecutive repetitions. In addition, all words, and part-words (all possible combinations of the last syllable of one word and the first two syllables of all others, that is, 12 part-words per language) were independently synthesized. All audio files were synthesized using the MBROLA text-to-speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996) with the Spanish Male Voice "es2" at 16 kHz. All phonemes were equal in pitch (200 Hz), pitch rise and fall (with the maximum at 50% of the phoneme), and duration was set as half of the syllable length. Part-words and words were generated with a syllable duration of 250 ms. Three different versions of each language stream were synthesized according to three different rhythmic structures created by manipulating the duration of the syllables as described in the next *sec*tion.

2.3. *Stimuli: rhythmic structure*

We generated three different rhythmic structures to synthesize the word streams. In each case, the duration of each syllable in the stream was randomly selected from a different probability density function (Fig. 2):

Isochronous:

$$Prob(dur) = \begin{cases} 1 \text{ } if \text{ } dur = 0.25 \text{ } sec \\ 0 \text{ } if \text{ } dur \neq 0.25 \text{ } sec \end{cases}$$

Semi-rhythmic:

$$Prob(dur) = \begin{cases} 0 & if \text{ } dur < 0.125 sec \text{ } or \text{ } dur > 0.5 \text{ } sec \\ e^{-10dur} & if \text{ } 0.125 \text{ } sec \leq dur \leq 0.5 sec \end{cases}$$

Arhythmic:

$$Prob(dur) = \begin{cases} 0 & if \text{ } dur < 0.125 sec \text{ } or \text{ } dur > 0.5 \text{ } sec \\ 1 & if \text{ } 0.125 \text{ } sec \leq dur \leq 0.5 sec \end{cases}$$

It has been shown that the natural range for syllable duration is between 125 and 500 ms across a variety of languages studied (Poeppel & Assaneo, 2020). Accordingly, we decided to test three distributions with considerably different functional forms, while ensuring that all syllable durations remain within this natural range. The Isochronous speech condition represents the rhythmic structure typically used in the phSL literature, whereby all syllables have the same duration. Syllable duration was fixed to 250 ms for all syllables (Fig. 2, left panel). For the Semi-rhythmic speech condition, we chose a distribution of durations still within the natural range but markedly different from that producing isochronous speech. Specifically, the distribution was strongly shifted to the lower boundary (shorter syllables more probable; Fig. 2, middle panel). This allowed us to construct an anisochronous speech stream with some syllable durations still more highly represented than others (i.e., a semi-rhythmic structure). The choice of shifting the distribution to the lower (faster syllables) rather than the higher (slower syllables) boundary was arbitrary. We predict a similar outcome for a distribution with a bias for slower syllables. Finally, for the Arhythmic speech condition we selected a uniform distribution whereby every duration within the natural range (125–500 ms) has the same probability of occurrence (Fig. 2, right panel).

All familiarization streams were 2-min long, what resulted in slightly different number of pseudowords on each audio file. Familiarization streams with a rhythmic structure comprised 160 pseudowords, the ones with a semi-rhythmic structure comprised between 170 and 177, and arrhythmic streams comprised 135 pseudowords.

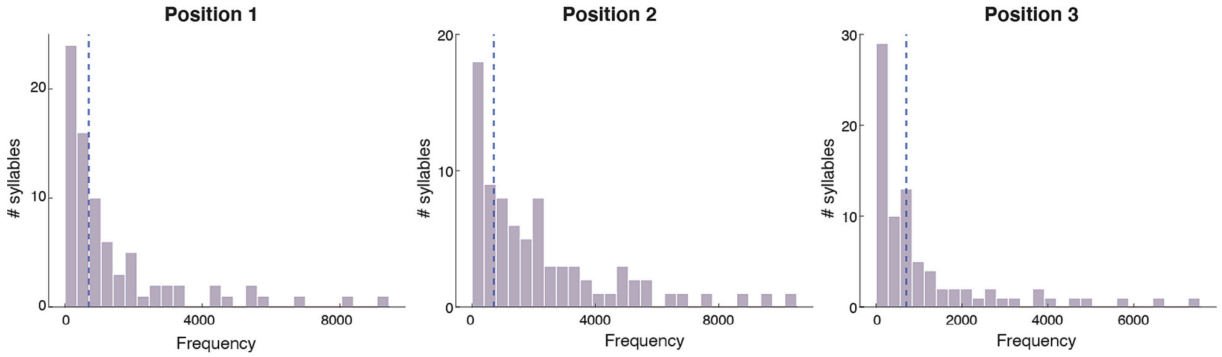


Fig. 1. Distribution of the token positional frequencies for all Spanish CV syllables. The 4 languages were generated by choosing 16 CV syllables for each position (four trisyllabic words per language). If the frequency of appearance of a given syllable in position i was below 700 (dashed line), it was assigned to a different position.

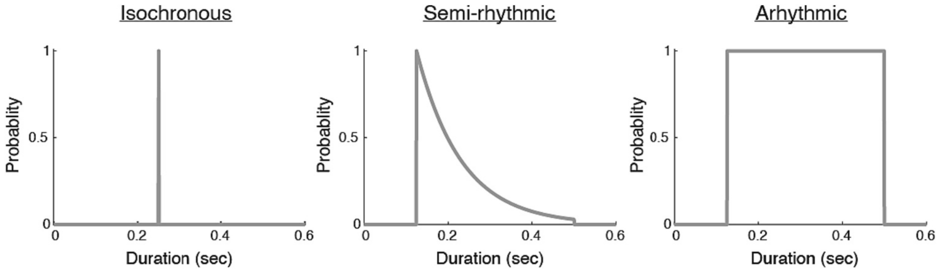


Fig. 2. Probability density functions defining the different rhythmic structures. Normalized probability of occurrence for the syllable durations. Each panel corresponds to a different rhythmic structure.

2.4. Phonological SL task

Participants passively listened to the familiarization stream and subsequently completed 8 two-alternative forced choice (2AFC) trials. On each test trial, a word and a part-word were acoustically and visually presented (i.e., the audio file was played while the written form of the stimulus appeared on the screen). The written presentation of the test items aimed to reduce participants' working memory load. Participants were instructed to select the more familiar stimulus. To construct the 8 test trials, 4 part-words were randomly selected from the pool of 12 and were paired twice with a different word. Participants' learning score was computed as the percentage of correct responses (i.e., choosing the word vs. the part-word).

2.5. Control experiment

An additional audio stimulus was synthesized for a control experiment. A 2-min syllable stream was generated by the *random* concatenation of the syllables composing Language 4 (Supplementary Table 1), with no gap between them and no consecutive repetitions as before. The syllable stream was synthesized as detailed above. The duration of all syllables was 250 ms as in the Isochronous speech condition of the main experiment. The test tokens for the control experiment were the same as those used for Language 4 in the main experiment.

Participants passively listened to the random syllable streams and subsequently completed the 8 two-alternative forced-choice trials corresponding to Language 4. Given that the sequence of syllables in the familiarization streams for this control experiment was entirely random, no learning of the words constitutive of Language 4 was possible during the familiarization phase. Thus, above-chance performance in the post-familiarization test for the control experiment (identifying Language 4 words as more familiar than part-words) can be taken to reflect participants' prior knowledge regarding plausible syllable order in Spanish irrespective of phSL. Conversely, comparing test performance for the control vs. the main experiment's conditions using Language 4 allows us to assess whether phSL actually occurred during these conditions. Because this comparison is particularly relevant in the case of Language 4 (the language with the highest SSI), we limited the control experiment to this artificial language.

2.6. Speech-to-speech synchronization task (SSS-test)

The explicit accelerated version of the SSS-test was conducted following established methodologies (Lizcano-Cortés et al., 2022). In this test, participants are explicitly instructed to continuously and repeatedly whisper the syllable "tah" in synchrony with an external auditory stimulus. The stimulus comprises a continuous random repetition of 16 different syllables and lasts 1 min. The presentation rate of the syllables increases from 4.3 to 4.7 syllables per second, in steps of 0.1 syll/s every 10 s. There is no overlap between these rates and the one of the Isochronous rhythmic structure (i.e., 4 Hz, see previous Section). Participants wear earplugs to diminish their own auditory feedback and whisper close to a microphone to register their vocalizations. For further details about this test refer to (Lizcano-Cortés et al., 2022).

For each participant, the degree of speech auditory-motor synchronization was determined by the phase locking value (PLV) between the envelope of the produced and perceived speech signals within a frequency band of [3.3 5.7] Hz. For this purpose, the following formula was applied:

$$PLV = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i(\theta_1(t)-\theta_2(t))}\right|$$

where t is the discretized time, T is the total number of time points, and θ_1 and θ_2 the phase of the envelope of the perceived and produced speech signals, respectively. The PLV was computed for windows of 5 s in length with an overlap of 2 s. The results for all time windows were averaged providing a single synchronization score per participant. Envelopes were computed as the absolute value of the Hilbert transform of the signals, resampled at 100 Hz, and filtered between 3.3 and 5.7 Hz. The Hilbert transform was then applied over the envelopes to extract the time evolution of their phase.

2.7. Experimental design

The whole experiment was conducted online using the cloud-based research platform Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). Each participant completed four blocks of phSL followed by a 2AFC test, each block with a different language. This was followed by the SSS-test. Two languages were presented with the isochronous structure and the two others with an anisochronous structure. Participants were assigned to one of two subgroups: for subgroup 1, the anisochronous rhythmic structure was the Semi-rhythmic speech condition; for subgroup 2, the anisochronous rhythmic structure was the Arhythmic speech condition (see Section "Stimuli: Rhythmic Structure"). Isochronous and anisochronous rhythmic structures were counterbalanced between participants and interleaved (i.e., isoani-isoani or ani-isoani-iso). Language order was randomized for each participant.

2.8. Participants categorization according to the SSS-test outcome

Previous work shows that the synchronization scores obtained by the SSS-test follow a bimodal distribution, implying that the general population can be segregated into two groups: high and low synchronizers (Assaneo et al., 2021, 2019; Lizcano-Cortés et al., 2022; Orpella et al., 2022; Rimmele et al., 2022). While high synchronizers synchronize their vocalizations to the external stream of syllables, synchrony is impaired for the low synchrony group (Mares et al., 2023). Before proceeding to categorize our participants as high or low synchronizers, we tested the bimodality of the obtained distribution of synchronization measurements. We adjusted two different Gaussian mixture distribution models (McLachlan & Peel, 2000), with 1 and 2 components, and computed their Bayesian Information Criterion. In line with previous studies, we found that the model that better fits our data distribution is the one with

2 components (BIC_1 = -54.3 and $BIC_2 = -95.9$). After confirming the bimodal nature of the data, we used the adjusted Gaussian mixture distribution with two components to label each participant as a high or low synchronizer. From the model, we extracted two critical PLV values: the lower boundary (the value below which participants have more than a 75% chance of belonging to the low synchrony group) and the higher boundary (the value above which participants have more than a 75% chance of belonging to the high synchrony group)(see Fig. 3). Participants below/above the lower/higher boundary were classified as low/high synchronizers. Participants above the lower boundary and below the higher one were excluded from subsequent analyses (see Participants section).

2.9. Linear mixed-effects model analysis

Two generalized linear mixed-effects model analyses were performed to predict participants' responses, one for each subgroup of participants (subgroup 1: Isochronous + Semi-rhythmic conditions; subgroups 2: Isochronous + Arhythmic conditions). We used the *buildmer* library (Voeten, 2019) in R (R Core Team, 2020). This library allowed us to identify the largest converging general linear mixed-effects model and, from there, perform a stepwise elimination to find the model that better explains participants' responses based on the change in Akaike's Information Criterion (AIC). The initial model included three fixed-effects predictors: *SSI* (a continuous variable comprising the z-scored Spanish Similarity Index representing the similarity of each language to the syllable order distribution in Spanish), *Group* (a categorical variable indicating whether the participant is a low or a high synchronizer according to the SSS-test), and *Rhythmic Structure* (a categorical variable indicating whether the rhythmic structure of the language was isochronous or anisochronous; subgroup 1: Isochronous vs Semi-rhythmic; subgroup 2: Isochronous vs Arhythmic). All interactions between these three variables were included in this model. Intercepts, but not slopes, were allowed to vary per participant. For the optimal models obtained, we assessed the effects of the predictors on learning performance by means of likelihood ratio tests based on Type 3 sums of squares using the *afex* library (Singmann et al., 2024). Estimated marginal means and trends were computed using the *emmeans* R package. All reported p values are Bonferroni corrected for multiple comparisons.

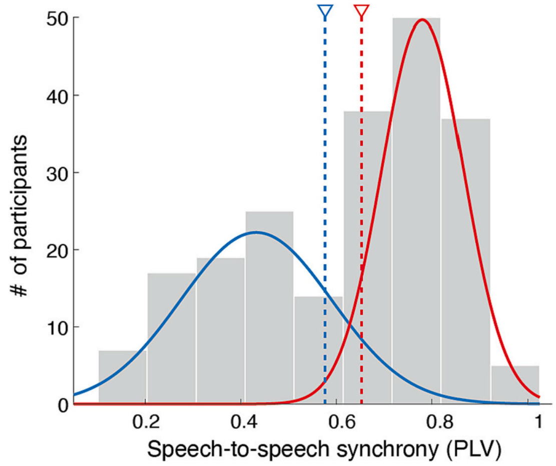


Fig. 3. SSS-test outcome. Histogram of the synchronization measurements obtained with the SSS-test for all participants evaluated in this study (gray bars). Superimposed in filled colored lines are the two distributions obtained by the Gaussian mixture distribution model fitting procedure. Dashed blue line: lower boundary. Participants with PLVs below this value have >75% probability of belonging to the low synchrony group. Dashed red line: higher boundary. Participants with PLVs above this value have >75% probability of belonging to the high synchrony group.

3. Results

Participants, classified as low or high synchronizers according to the SSS-test, completed four phonological SL blocks. Each block comprised a different language with a given *SSI score* assessing how well the syllabic structure of the words composing that language aligned with the syllables' statistics of the Spanish language (see Materials and Methods). Two languages were presented with the Isochronous rhythmic structure. The remaining two languages were presented with an anisochronous rhythmic structure (Semi-rhythmic for subgroup 1; Arhythmic for subgroup 2).

As a first exploration of the results, we tested for a general effect of rhythmic structure over learning, regardless of individual differences and the specific syllables composing the different languages. By pooling together high and low synchronizers and averaging across languages, we found no significant differences between the Isochronous rhythmic structure and neither of the two other anisochronous conditions (see Supplementary Fig. 1). Next, we performed a linear mixed-effects analysis to assess for an interaction between rhythmic structure and the other controlled variables. More precisely, for each subgroup of participants (subgroup 1: Isochronous + Semi-rhythmic conditions; subgroup 2: Isochronous + Arhythmic conditions), we assessed whether Group (auditory-motor synchrony status: high vs low synchronizer), Rhythmic Structure (isochronous vs anisochronous speech) and SSI score modulate participants' phSL performance. We estimated the optimal converging linear mixed-effects model through backwards stepwise elimination. For subgroup 1 we found that the model that better accounts for participants' responses included the interaction between Group and SSI, but not Rhythmic Structure (see Table 1). Following this significant interaction, we computed the estimated marginal mean trends for the relationship between learning and SSI for each Group. In line with previous work (Elazar et al., 2022), we found that SSI had a significant effect on learning, regardless of synchrony Group ($trend_{HIGHS} = 0.38$, zratio = 7.27, $p < 0.001$; $trend_{LOWS} = 0.22$, zratio = 3.55, $p < 0.001$). However, high synchronizers showed a steeper trend than low synchronizers (see Fig. 4a, zratio = 1.98 $p = 0.048$). To explore the average learning across languages, we computed the estimated marginal means. Results showed that both groups of participants (high and low synchronizers) performed above chance ($mean_{HIGHS} = 0.72$, zratio = 11.09, $p < 0.001$; $mean_{LOWS} = 0.68$, zratio = 7.56, $p < 0.001$).

Regarding subgroup 2, we found that the best model for our data included a triple interaction between Group, SSI, and Rhythmic Structure (see Table 2). As before, we computed the estimated marginal mean trends of the learning as a function of SSI score for each Group. This time, given the significant effect of Rhythmic Structure, we performed the analysis on Isochronous and Arhythmic speech conditions separately (see Fig. 4b). For the Isochronous speech condition we found, as for subgroup 1, that high synchronizers showed a steeper trend than lows ($trend_{HIGHS} = 0.51$ and $trend_{LOWS} = 0.16$, zratio = 2.81, $p = 0.005$). Interestingly, we found that this difference between groups was not present in the Arhythmic speech condition ($trend_{HIGHS} = 0.22$ and

Table 1
Linear mixed-effects model results for Subgroup 1. *SSI*: z-scored Spanish Similarity Index score, Group: high or low synchrony group according to the SSS-test outcome, Sub: Participants and * stands for an interaction. Significant results are marked in bold.

Subgroup 1: Isochronous + Semi-rhythmic		
Best Model	*Learn ~ SSI + Group + SSI*Group + (1|Sub)*	
Analysis of Deviance (Type III χ^2 Test)		
	χ^2	**p**
Intercept	123.02	**<0.001**
Group	1.50	0.22
SSI	52.89	**<0.001**
*SSI*Group*	3.90	**0.048**

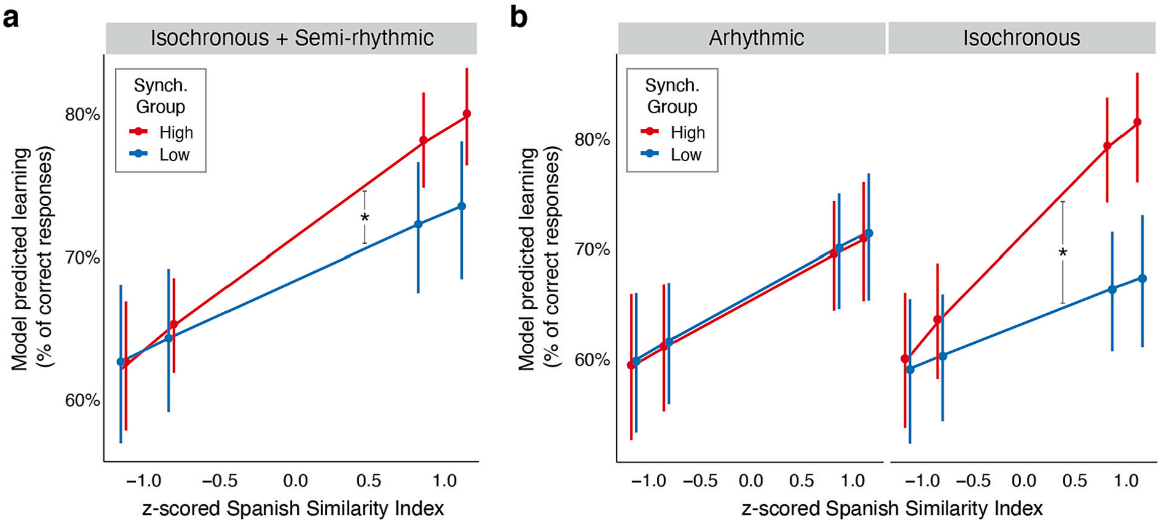


Fig. 4. Linear mixed model results: percentage of learning as a function of the Spanish Similarity Index. a. Results obtained for the subgroup of participants presented with the Isochronous and Semi-rhythmic speech conditions. b. Results obtained for the subgroup of participants presented with the Isochronous and Arhythmic speech conditions. Dots: model predicted group means. Bars: 95% confidence interval. Red: high synchronizers. Blue: low synchronizers. * $p < 0.05$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Linear mixed-effects model results for Subgroup 2. *SSI*: z-scored Spanish Similarity Index score, Group: high or low synchrony group according to the SSS-test outcome, Sub: Participants, Rhy: stimulus rhythmic structure. * stands for an interaction. Significant results are marked in bold.

Subgroup 2: Isochronous + Arhythmic		
Best Model	*Learn ~ SSI + Group + Rhy + SSI*Group + Group*Rhy + SSI*Rhy + SSI*Rhy*Group + (1|Sub)*	
Analysis of Deviance (Type III χ^2 Test)		
	χ^2	p
Intercept	36.64	**<0.001**
Group	0.04	0.83
Rhy	8.73	**0.003**
SSI	6.99	**0.008**
*Group*SSI*	0.01	0.93
*Rhy*SSI*	5.29	**0.021**
*Group*Rhy*	7.97	**0.004**
*Group*Rhy*SSI*	4.23	**0.038**

$\text{trend}_{\text{LOWS}} = 0.22$, zratio $= -0.085$, $p = 0.93$). Moreover, only the high synchrony group showed a significant difference between Isochronous and Arhythmic speech conditions (Highs: $\text{trend}_{\text{RAND}} = 0.22$, $\text{trend}_{\text{ISO}} = 0.51$, zratio $= 2.30$, $p = 0.02$; Lows: $\text{trend}_{\text{RAND}} = 0.22$, $\text{trend}_{\text{ISO}} = 0.16$, zratio $= -0.58$, $p = 0.56$). To further test that learning occurred in the Arhythmic speech condition across languages, we conducted a marginal mean estimation for the mean performance across languages. We found that both groups (high and low synchronizers) performed above chance for both rhythmic structures (Isochronous speech: $\text{mean}_{\text{HIGHS}} = 0.73$, zratio $= 8.89$, $p < 0.001$; $\text{mean}_{\text{LOWS}} = 0.63$, zratio $= 5.246$, $p < 0.001$; Arhythmic speech: $\text{mean}_{\text{HIGHS}} = 0.65$, zratio $= 6.05$, $p < 0.001$; $\text{mean}_{\text{LOWS}} = 0.66$, zratio $= 6.43$, $p < 0.001$).

We conducted a control experiment to assess 1) whether performance in the different conditions of the main experiment can be attributed to phSL rather than simply resulting from participants' prior knowledge of plausible syllable order in Spanish and 2) whether the difference in performance observed between synchrony groups derives from high synchronizers being more attuned to the statistics of their native language than low synchronizers, rather than from their SL abilities. That is, given that the words composing all our languages were designed to guarantee a minimum alignment with the syllable-level statistics of the Spanish language (average to high SSI scores), it is plausible to perform above chance in the post-familiarization tests without phSL occurring during the familiarization phase by simply relying on prior linguistic knowledge. To explore this possibility, we exposed a new cohort of participants to a random concatenation of the 16 syllables composing Language 4 (see Control Experiment in Materials and Methods). Participants subsequently completed the same test used to evaluate phSL for that language, followed by the SSS-test. Given that all syllables in this new speech stream had equal transitional probabilities, no phSL was possible during the familiarization phase. However, above chance test performance is still possible based on participant's prior knowledge. We found that both groups (i.e., high and low synchronizers) performed above chance (see Fig. 5a, Highs: $N = 33$ and $p = 0.0145$; Lows: $N = 26$ and $p = 0.0123$; two-sided Wilcoxon signed rank test against 50%) with no significant difference between the groups ($p = 0.51$; two-sided Wilcoxon rank sum test). This result indicates that, although both groups can rely on prior knowledge to identify the combination of syllables more likely to occur together in their native language, high synchronizers outperform lows only when phSL is possible.

To further assess whether above-chance performance in the 2AFC tests of the main experiment can be attributed to phSL rather than to participants' preference for items that more closely resemble familiar-sounding words in their native language, we compared test performance for the control experiment to performance for all conditions in the main experiment using Language 4 (i.e, with the different rhythmic structures). Note that this comparison is most critical for this language because of its higher adherence to the Spanish language statistics (higher SSI score). A direct comparison between all conditions using Language 4 showed a significant increment in performance whenever SL was possible (Fig. 5b), that is when the familiarization streams contained the language's words rather than a random concatenation of the language's syllables. This result shows that, even for the language with the closest similarity to the participants native language, SL influences participants' responses and does so across all three different rhythmic structures tested.

4. Discussion

We tested the effect of syllabic rhythmic structure (Isochronous, Semi-rhythmic, Arhythmic) and its interaction with two factors known to influence phSL: auditory-motor synchronization (high vs. low

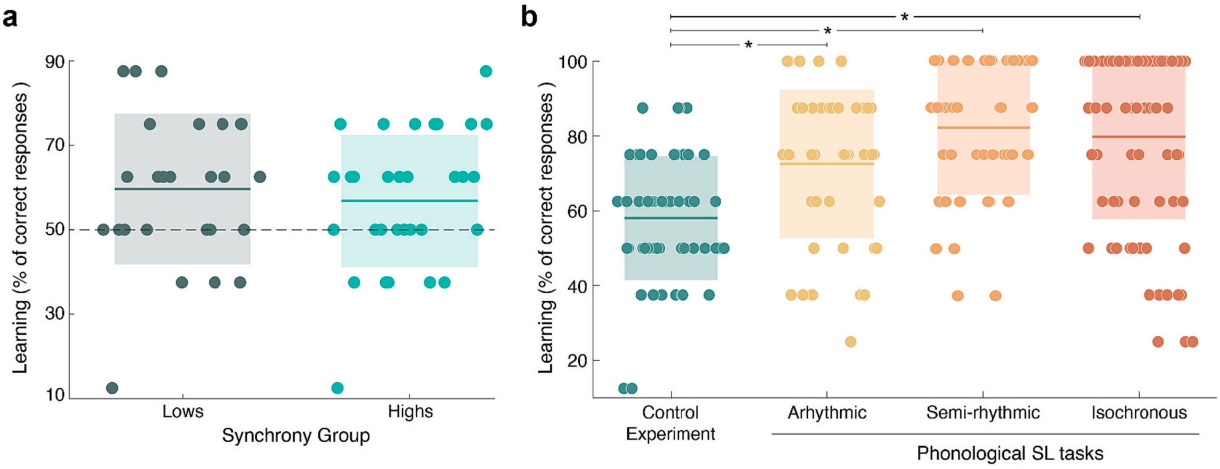


Fig. 5. Control experiment assessing participants' ability to identify words (vs part-words) based solely on the words' similarity to the Spanish language. a. Performance comparison between synchrony groups. Both groups performed above chance (HIGHS: $N = 33$ and $p = 0.0145$; LOWS: $N = 26$ and $p = 0.0123$; two-sided Wilcoxon signed rank test against 50%) with no significant difference between groups ($p = 0.51$; two-sided Wilcoxon rank sum test). Dark and light green indicate high and low synchronizers, respectively. b. Test performance for Language 4 across the different familiarization conditions. Performance was significantly higher for the three SL conditions (Arhythmic, Semi-Rhythmic, Isochronous) compared to the control experiment, in which no SL is possible (Control: $N = 59$; Arhythmic: $N = 46$; Semi-rhythmic: $N = 47$; Isochronous: $N = 91$). * $p < 0.001$ for a two-sided Wilcoxon rank sum test, Bonferroni corrected for multiple comparisons. In all panels, dots represent individual subjects, solid lines mean values, dashed line chance level, and shaded region standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

synchronizers) and prior knowledge regarding syllable order (SSI). Overall, phSL appears robust to rhythmic variability (deviations from isochrony), with significant learning following isochronous as well as anisochronous presentations. Participants showed sensitivity to the plausibility of the (pseudo)words' syllable order with respect to their native language. Interestingly, high auditory-motor synchronizers only exhibited enhanced performance over low synchronizers in the presence of both rhythmically structured input (i.e., Isochronous and Semi-rhythmic but not Arhythmic speech) and prior information (languages with higher SSI) - but neither of these factors in isolation. That is, even when syllables are rhythmically presented, high synchronizers do not outperform lows unless syllable order is highly plausible in their native language.

To our knowledge, this is the first study to test deviations from isochronous syllable presentations systematically. Robustness to deviations from isochrony is a necessary feature for a mechanism that is argued to be critical for natural language acquisition; syllables in naturally occurring speech may be quasi-rhythmic but are not isochronous. Previous phSL research showed successful learning using naturalistic stimuli (e.g., Hay et al., 2011; Pelucchi et al., 2009) with controlled yet not perfectly isochronous syllable durations, suggesting some robustness to rhythmic variability akin to that in the Semi-rhythmic condition in the current study. While we only tested a distribution of syllable durations favoring faster syllables under this condition, we conjecture that deviations from isochrony favoring shorter syllables (within the natural range of durations) will yield a similar performance. More importantly, however, we show that phSL is also possible in the face of artificial languages with a completely arhythmic syllabic structure.

Our results also demonstrate that participants exploit or rely on their prior knowledge regarding syllable order: languages with words more similar to their native language were better learned. Critically, we showed that this knowledge is used *for* learning during the task's familiarization phase. That is, although a high SSI was sufficient to bias participants' responses in a test following a random syllable stream, learning was significantly greater following the phSL familiarization phase. This aligns with the findings of Elazar et al. (2022) and Siegelman et al. (2018). However, it is worth noting that the metric we used to characterize the languages (the SSI) is not the same as the syllable transitional probability used in previous works. This suggests that prior knowledge is organized in terms of general attributes of the statistical regularities of the participants native language (i.e., not restricted to syllable transitional probability). Follow up studies could explore how the SSI interacts with the transitional probabilities between syllables to modulate phSL.

A revealing finding was that the only phSL conditions that distinguished high from low synchronizers involved both prior knowledge *and* rhythmic structure. This was true for Semi-rhythmic as well as Isochronous presentations. Previous studies reported significant differences between high and low synchronizers, with high synchronizers consistently outperforming lows (Assaneo et al., 2019; Orpella et al., 2022). In those experiments, high synchronizers were also shown to differ neuroanatomically, neurophysiologically, and functionally. For example, Orpella et al. (2022) showed that, while learning across synchrony groups correlated with the engagement of an auditory network comprising auditory and superior temporal cortex, only high synchronizers additionally engaged the dorsal language stream, including inferior frontal and parietal cortex. Moreover, the engagement of the dorsal language stream correlated with participants' behavioral auditory-motor synchronization abilities and boosted the phSL of high auditory-motor synchronizers. Furthermore, high synchronizers lost their learning advantage when the use of the dorsal language stream for learning was compromised via articulatory suppression (i.e., when repeating a nonsense syllable during the familiarization phase). Together with previous findings (Assaneo et al., 2019), this pattern of results led us to hypothesize the existence of two distinct mechanisms for phSL: (i) a default mechanism engaging bilateral auditory and superior temporal cortex that is independent of auditory-motor synchronization and (ii) an additional mechanism engaging the dorsal language stream that leverages the rhythmic structure of the auditory input and boosts learning. Regarding this additional mechanism, a possibility rooted in other experimental findings (Assaneo et al., 2021; Park et al., 2015; Rimmele et al., 2018) is that the dorsal language stream affords temporal predictions that help align the auditory cortex (i.e., entrain its activity) to the input stream resulting in a better phonological encoding of the syllables and subsequent SL. However, evidence from the current study does not support this possibility: high synchronizers, who we predict based on previous work (Assaneo et al., 2019; Orpella et al.,

2022) differentially recruited the dorsal language stream during learning, did not show better phSL than low synchronizers when the speech input simply increased in rhythmic predictability (i.e., change from Arhythmic to Semi-rhythmic or Isochronous in conditions of low SSI).

An alternative hypothesis is that the engagement of the dorsal language stream shown by high synchronizers during phSL relates to the use of predictions in which both temporal and content information about the upcoming speech elements go hand in hand (Orpella et al., 2021). The fact that high synchronizers show significantly better phSL in the presence of both predictable rhythmic structure and prior information aligns well with this hypothesis. This is also consistent with a recently reported overlap in left parietal regions for phSL and the integration of temporal and content predictive cues (Orpella et al., 2020). Data from several phSL studies (e.g., Cunillera et al., 2009; Karuza et al., 2013; Orpella et al., 2021) suggest that this putative prediction mechanism leverages the same dorsal stream architecture used to generate predictions for speech perception (Auksztulewicz et al., 2018; Rimmele et al., 2018) and motor control (Guenther, 2015), including speech motor regions and the basal ganglia. Orpella et al., for example, used behavior, computational modeling, and fMRI to show that trial-by-trial phSL responds to prediction-based learning that correlates with activity in left fronto-temporal cortical regions as well as bilateral basal ganglia (Orpella et al., 2021). Future research could investigate whether conditions besides input predictability (e.g., challenging listening situations, aging) drive the engagement of this prediction-based mechanism. In addition, whether the learning advantage conferred by this additional mechanism is truly quantitative (i.e., producing more or more robust rather than simply quicker learning) given longer familiarization time remains an open question.

In sum, we show that phSL is modulated by the consistency of the 'new' language with the statistics of the learners' first language, but not by syllable rhythmicity alone. However, high auditory-motor synchronizers are better statistical learners when the speech input contains both kinds of cues (temporal and content). Thus, the picture that emerges from the current and previous data is that of (1) a default mechanism for phSL that is robust to syllabic rhythmic variability in the input and leverages prior knowledge and (2) an additional mechanism, used by a subset of the population (high auditory-motor synchronizers), that leverages prior knowledge and input rhythmicity concurrently for learning.

CRediT authorship contribution statement

Ireri Gómez Varela: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Joan Orpella:** Writing – review & editing, Formal analysis, Conceptualization. **David Poeppel:** Funding acquisition, Conceptualization. **Pablo Ripolles:** Writing – review & editing, Investigation, Formal analysis, Funding acquisition, Conceptualization. **M. Florencia Assaneo:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Authors report no competing interests.

Data availability

The data supporting the findings of this study are available as Supplementary Data. All other data are available from the corresponding author.

Acknowledgments

This work was supported by UNAM-DGAPA-PAPIIT IA202921/IA200223 (M.F.A.), National Science Foundation (https://www.nsf.gov/) grant 2043717 (D·P, P.R., J.O. and M.F.A.), I.G.V. received CONACYT funding (CVU: 1045881) from the Mexican government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2024.105737.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388–407.

Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. *Nature Human Behaviour, 5*(1). https://doi.org/10.1038/s41562-020-00962-0. Article 1.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience, 22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Auksztulewicz, R., Schwiedrzik, C. M., Thesen, T., Doyle, W., Devinsky, O., Nobre, A. C., … Melloni, L. (2018). Not all predictions are equal: "What" and "when" predictions modulate activity in auditory cortex through different mechanisms. *Journal of Neuroscience, 38*(40), 8680–8693. https://doi.org/10.1523/JNEUROSCI.0369-18.2018

Bhaya-Grossman, I., & Chang, E. F. (2022). Speech computations of the human superior temporal gyrus. *Annual Review of Psychology, 73*(1), 79–102. https://doi.org/10.1146/annurev-psych-022321-035256

Coull, J. T., Cheng, R.-K., & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology, 36*(1). https://doi.org/10.1038/npp.2010.113. Article 1.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., … Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage, 48*(3), 541–553. https://doi.org/10.1016/j.neuroimage.2009.06.069

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

Duñabeitia, J. A., Cholin, J., Corral, J., Perea, M., & Carreiras, M. (2010). SYLLABARIUM: An online application for deriving complete statistics for Basque and Spanish orthographic syllables. *Behavior Research Methods, 42*(1), 118–125.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). *October). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes, ICSLP'96 (Vol. 3*, pp. 1393–1396). IEEE.

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "tabula" is anything but "rasa:" What determines performance in the auditory statistical learning task? *Cognitive Science, 46*(2), Article e13102. https://doi.org/10.1111/cogs.13102

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Getz, H., Ding, N., Newport, E. L., & Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition, 181*, 135–140. https://doi.org/10.1016/j.cognition.2018.08.019

Guenther, F. H. (2015). *Neural control of speech*. The MIT Press.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63*(2), 93–106. https://doi.org/10.1016/j.cogpsych.2011.06.002

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., … Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances, 7*(8), eabc4530. https://doi.org/10.1126/sciadv.abc4530

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*(1), 46–54. https://doi.org/10.1016/j.bandl.2012.11.007

Lizcano-Cortés, F., Gómez-Varela, I., Mares, C., Wallisch, P., Orpella, J., Poeppel, D., … Assaneo, M. F. (2022). Speech-to-speech synchronization protocol to classify human participants as high or low auditory-motor synchronizers. *STAR Protocols, 3*(2), Article 101248. https://doi.org/10.1016/j.xpro.2022.101248

Lopez-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T. F., & Rodriguez-Fornells, A. (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cerebral Cortex, 21*(12), 2742–2750. https://doi.org/10.1093/cercor/bhr064

López-Barroso, D., Ripollés, P., Marco-Pallarés, J., Mohammadi, B., Münte, T. F., Bachoud-Lévi, A.-C., … de Diego-Balaguer, R. (2015). Multiple brain networks underpinning word learning from fluent speech revealed by independent component analysis. *NeuroImage, 110*, 182–193. https://doi.org/10.1016/j.neuroimage.2014.12.085

Mares, C., Echavarría Solana, R., & Assaneo, M. F. (2023). Auditory-motor synchronization varies among individuals and is critically shaped by acoustic features. *Communications Biology, 6*(1). https://doi.org/10.1038/s42003-023-04976-y. Article 1.

Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition, 224*, Article 105054. https://doi.org/10.1016/j.cognition.2022.105054

McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, Inc.. https://www.wiley.com/en-mx/Finite+Mixture+Models-p-9780471006268

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38*(1–3), 84–105. https://doi.org/10.1159/000260016

Orpella, J., Assaneo, M. F., Ripollés, P., Noejovich, L., López-Barroso, D., de Diego-Balaguer, R., & Poeppel, D. (2022). Differential activation of a frontoparietal network explains population-level differences in statistical learning from speech. *PLoS Biology, 20*(7), Article e3001712. https://doi.org/10.1371/journal.pbio.3001712

Orpella, J., Mas-Herrero, E., Ripollés, P., Marco-Pallarés, J., & de Diego-Balaguer, R. (2021). Language statistical learning responds to reinforcement learning principles rooted in the striatum. *PLoS Biology, 19*(9), Article e3001119. https://doi.org/10.1371/journal.pbio.3001119

Orpella, J., Ripollés, P., Ruzzoli, M., Amengual, J. L., Callejas, A., Martinez-Alvarez, A., … de Diego-Balaguer, R. (2020). Integrating when and what information in the left parietal lobe allows language rule generalization. *PLoS Biology, 18*(11), Article e3000895. https://doi.org/10.1371/journal.pbio.3000895

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25*(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

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

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rebuschat, P., & Williams, J. N. (2012). *Statistical learning and language acquisition.* Walter de Gruyter.

Rimmele, J. M., Kern, P., Lubinus, C., Frieler, K., Poeppel, D., & Assaneo, M. F. (2022). Musical sophistication and speech auditory-motor coupling: Easy tests for quick answers. *Frontiers in Neuroscience, 15*. https://doi.org/10.3389/fnins.2021.764342

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences, 22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

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., 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

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition, 177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Singmann, H., Bolker, B., Westfall, S., & Aust, F. (2024). afex: Analysis of factorial experiments. https://afex.singmann.science/.

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

Voeten, C. C. (2019). Using `buildmer' to automatically find & compare maximal (mixed) models. https://cran.r-project.org/web/packages/buildmer/vignettes/buildmer.html.