ALPARC: Artificial Languages with Phonological and Acoustic Rhythmicity Controls

Lorenzo Titone[1*], Nikola Milosevic[2], Lars Meyer[1,3]

[1] Max Planck Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

[2] Methods and Development Group Neural Data Science and Statistical Computing, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.

[3] Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, 48149 Münster, Germany.

[*] Corresponding author: titone@cbs.mpg.de

Abstract

Infants and adults show the remarkable ability to learn from statistical regularities in the environment. Seminal studies in language acquisition suggested that transitional probabilities between syllables are decisive for language learning. Yet, recent work cautioned that acoustic and phonological regularities can confound transitional probabilities, compromising interpretability. Furthermore, prior linguistic background can impact the learning of a new (artificial) language. To control for such confounds, we developed an open-source Python toolbox that generates *Artificial Languages with Phonological and Acoustic Rhythmicity Controls* (APLARC). First, we explain all functionalities of ALPARC through a step-by-step guide. Then, we demonstrate how ALPARC generates syllable streams encompassing pseudowords that are tailored to critical statistics of real languages. Our results show that ALPARC streams attain more stationary transitional probability distributions and minimize phonological and acoustic confounds relative to stimuli used in prior studies. We conclude that ALPARC would greatly boost the interpretability of future language learning research.

Introduction

Statistical learning (SL) is the ability to extract regularities from the environment. Among its role for several cognitive functions, SL is pivotal for speech segmentation and language acquisition (Bogaerts et al., 2020; Frost et al., 2019). In fact, while lexical boundaries are rarely marked acoustically, they can be inferred from the transitional probabilities (TPs) between syllables, which are typically higher within than across words. Seminal research on SL in language acquisition utilized artificial language learning paradigms to demonstrate that infants and adults can acquire a lexicon of artificial nonsense words marked by low-TP boundaries (e.g., Saffran et al., 1996). After exposure to streams of syllables presenting statistical regularities, learning can be assessed behaviorally by reflection-based measures, such as two-alternative forced choice tasks, where pseudowords should be endorsed as more familiar than chunks that violate the TPs (see also Batterink et al., 2015; Christiansen, 2019, for discussion).

In parallel to this classical behavioral work, recent electrophysiological studies have assessed the neural tracking of statistical regularities at the pseudoword rate by employing so-called frequency-tagging paradigms (Assaneo et al., 2019; Batterink & Paller, 2017; Buiatti et al., 2009; Cunillera et al., 2006, 2009; Getz et al., 2018; Henin et al., 2021; McNealy et al., 2006; Ordin et al., 2020; see also Sjuls et al., 2024 for a review). In this paradigm, syllable streams are presented at one fixed rate (e.g., 3.3 Hz) and exhibit controlled distributions of TPs, such that *TP-structured* streams present hidden rhythmic patterns of low-TP events marking pseudoword boundaries at a different rate (e.g., 1.1 Hz), whereas *TP-uniform* streams present uniform TPs throughout the stream. These studies found increased brain activity at the syllable rate (physically present in the stimulus) in both conditions, but also stronger activity at the pseudoword rate (putatively absent from the stimulus) in the TP-structured stream, suggesting sensitivity to TPs, beyond the acoustics, or perceptual chunking of pseudowords (see Batterink & Paller, 2017; Perruchet & Pacton, 2006; Saffran & Kirkham, 2018; Siegelman et al., 2019; Thiessen, 2017 for discussion). However, recent work raised concerns about the control of the stimuli used in these studies, specifically on the levels of acoustic regularities and phonotactics, which may confound the periodicity at the pseudoword rate (Pinto et al., 2022; van der Wulp, 2021).

One caveat in linguistic SL is that learners may infer lexical boundaries not only based on TPs, but also on acoustic–phonetic–phonological cues (Onnis et al., 2005; Pinto et al., 2022; Seidenberg et al., 2002) or phonotactic rules (Boll-Avetisyan & Kager, 2014; van der Wulp, 2021). For instance, cues such as the repetition of the same phonological feature at regular positions within a stream may accidentally accompany TPs and confound supposed effects of TPs on learning, thus compromising scientific interpretability. In addition to these alternative segmentation cues, linguistic entrenchment with known natural languages and speech statistics may bias artificial language learning (Siegelman et al., 2018). Moreover, precise TP manipulations should be attained to prevent additional sources of variability from further biasing cue diagnosticity and impact SL performance (Hasson, 2017). Hence, both behavioral and electrophysiological studies on SL of artificial languages would benefit from carefully controlling for confounding phonological and acoustic regularities, as well as prior linguistic background, and statistical context (Elazar et al., 2022; Siegelman et al., 2018; Varela et al., 2024). In particular, frequency-tagging studies require rigorous rhythmicity controls to ensure that only the variable of interest (e.g., TPs) and no other factor (e.g., phonological patterns) occurs periodically. In fact, without disentangling statistical and acoustic regularities, it cannot be excluded that tracking at the pseudoword rate would stem, at least partially, from acoustic rhythms, rather than statistical ones.

To account for these confounding factors in artificial language learning studies, we present here an open-source Python toolbox to generate and diagnose *Artificial Languages with Phonological and Acoustic Rhythmicity Controls* (ALPARC). We first provide a step-by-step guide to generate artificial lexicons of variable size that can be tailored to the statistics of existing languages (e.g., phoneme n-gram probabilities and phonotactics), while also implementing phonological and acoustic rhythmicity controls and exact TP manipulations within syllable streams of arbitrary length. Then, we show that ALPARC suppresses acoustic and phonological rhythmicity at the rate of interest and enhances TP stationarity relative to benchmarking artificial language streams. Decisively, we argue that due to its flexibility and robustness, ALPARC could make a useful tool to increase the systematicity in stimulus development and overcome current uncertainties in SL studies.

Methods

ALPARC implements a custom type system of linguistic elements to enable a fast search over large combinatorial spaces and to support the creation of syllable streams with a number of user-specified parameters (Figure 1). Phonological rhythmicity controls are based on a universal matrix of binary phonological features for all phonemes. *Phonemes* are the atomic linguistic units of ALPARC, which other elements, such as *Syllables*, *Pseudowords*, and *Lexicons* are based on. ALPARC can be used both as a dataset, to download our pre-computed artificial languages, and as a library, to generate new artificial languages in a step-by-step fashion. Furthermore, ALPARC can be used to diagnose external artificial languages in terms of phonological rhythmicity confounds.

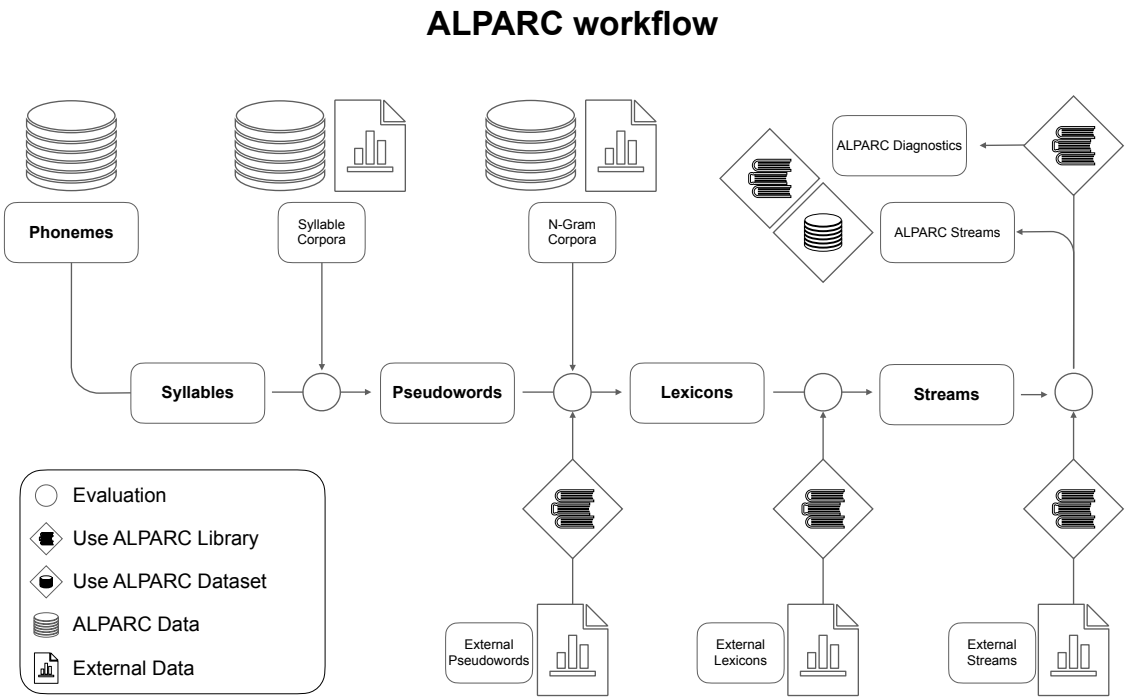

Figure 1. ALPARC workflow. The pipeline highlights all the steps for the generation of artificial language streams. Phonemes annotated with binary vectors of phonological features constitute the atomic linguistic unit of ALPARC. Language-specific corpora can be used to filter the *Syllable* and *Pseudoword Register* based on syllable frequency of use and phoneme n-gram probabilities in real languages. Pseudowords can be generated either internally (using our linguistic constraints) or externally. Then, ALPARC constrains the assembly of a *Lexicon Register* to minimize phonological feature regularities among pseudoword pairs that would confound TPs. Finally, lexicons are converted into TP-controlled streams of syllables. All steps for stream generation are wrapped inside the "generate" command that allows several configurations of parameters. External streams can be diagnosed with the "diagnose" command.

Step 1—from phonemes to syllables: frequency control

Prior linguistic knowledge impacts SL (Elazar et al., 2022; Siegelman et al., 2018; Varela et al., 2024). To address this issue, previous work relied on experts' knowledge or linguistic corpora. For instance, the similarity between artificial and real languages can be controlled by excluding syllables associated with high frequency of use (Kiai & Melloni, 2021), or by avoiding infrequent syllable-level positional similarity relative to existing words (Fernandes et al., 2009; Varela et al., 2024). Expanding on this work, ALPARC offers a systematic method to control for linguistic similarity with natural languages at the level of phonemes and syllables, as well as at the level of pseudowords (see also *Step 2*).

First, ALPARC creates a *Phoneme Register* based on a binary matrix of phonological features for all phonemes, which is provided out-of-the-box. All phonemes are transcribed in the International Phonetic Alphabet notation and annotated with binary *class features* (*syllabic* [*syl*], *sonorant* [*son*], *consonantal* [*cons*]); *manner features* (*continuant* [*cont*], *delayed release* [*delrel*], *lateral* [*lat*], *nasal* [*nas*], *strident* [*strid*]); *laryngeal features* (*voice* [*voi*], *spread glottis* [*sg*], *constricted glottis* [*cg*]); and *place features* (i.e., *anterior* [*ant*], *coronal* [*cor*], *distributed* [*distr*], *labial* [*lab*], *high* [*hi*], *low* [*lo*], *back* [*back*], *round* [*rnd*], *tense* [*tense*], *long* [*long*]). Then, ALPARC creates a *Syllable Register* with specific combinations of phonological features based on this *Phoneme Register*.

To compose the *Syllable Register*, users may specify only one desired syllable structure (default = consonant-vowel syllables with long vowels; alternatives comprise consonant-vowel syllables with short and/or long vowels, or consonant-vowel-consonant syllables). The default is recommended for experimental designs with isochronous input (e.g., frequency-tagging), since it eradicates variability in vowel duration across syllables. Additionally, ALPARC can read linguistic corpora and skim such pre-existing inventories to select syllables with the specified structure, rather than building a *Syllable Register* from scratch, with no guarantee that the resulting syllables belong to a real language. Out-of-the-box, ALPARC comprises preprocessed syllable corpora annotated with frequency of use in German (Schiel, 2010) and English (Baayen et al., 1996), as well as phoneme n-gram and positional probabilities within real words based on another German corpus (Arnold & Tomaschek, 2016).

In addition to the choice of syllable structure, ALPARC can also filter out those syllables whose frequency of use significantly deviate from the uniform distribution (see *Statistical Analyses*). This filter produces more homogeneity in frequency of use than a naturally Zipf-distributed register (Zipf, 1932). In general, corpora-based filters can be useful for two reasons. First, using real syllables from the native language of the learners is essential to ensure intelligibility, since non-distinctive phonemes are ambiguous. Second, eliminating highly (in)frequent syllables would reduce the potential tendency to endorse an artificial pseudoword due to the presence of a particularly (un)familiar syllable. Yet, diverse artificial language learning outcomes may ensue due to varying levels of linguistic similarity with already known language (Siegelman et al., 2018), making it difficult to express what an unbiased baseline would look like. In this context, ALPARC aims to provide an interface to moderate linguistic entrenchment with real languages, but compromises ought to be made on a case-by-case basis. For instance, using only open consonant-vowel syllables make for a German-based *Syllable Register* of adequate size to then apply further phonotactic controls, but these settings are too strict for English, as only a syllable structure unconstrained on vowel duration results compatible with further control settings. The output *Syllable Register* serves as a basis to create a *Pseudoword Register*, where similar filters can be applied at the level of phonemes n-grams and positional probabilities in real words.

Step 2—from syllables to pseudowords: phonotactic control

Phonotactic rules in a given language can bias word segmentation. For instance, in languages such as Dutch, neighboring syllables with the same place of articulation occur more frequently across word boundaries than within words. The adherence to this phonotactic rule—also known as the *obligatory contour principle*—was shown to influence word segmentation (Boll-Avetisyan & Kager, 2014) and to partially account for neural tracking at the pseudoword rate in a recent reanalysis of an artificial language learning study (Batterink & Paller, 2017; van der Wulp, 2021). This finding suggests that rhythmicity in place of articulation may confound TP manipulations. Thus, researchers highlighted the need to better control phonotactics in future studies (Pinto et al., 2022; van der Wulp, 2021).

ALPARC offers a systematic method to generate controlled artificial pseudowords with respect to phonotactics by using the binary matrix of phonological features and controlling for the adherence to specific phonotactic rules during the process of pseudoword generation from the *Syllable Register*. To that end, ALPARC combines simple phonological features into major phono-articulatory classes (Ramers, 1998), that are subdivided into balanced categories according to their place of articulation (*labial*, *apical*, *other*), manner of articulation (*sonorant*, *plosive*, *fricative*), and vowel identity. These classes are then used to restrict the syllable combinations that are allowed to form pseudowords based on the obligatory contour principle.

To compose the *Pseudoword Register*, users can specify the desired number of syllables per pseudoword (default = 3) and set the degree of phonotactic control. By default, ALPARC generates trisyllabic pseudowords according to the obligatory contour principle with constraints on the place and manner of articulation, and vowel identity. Therefore, the resulting pseudowords will comprise three syllables, each of which will have a different place and manner of articulation of the consonants and different vowels. Alternatively, users can import pseudowords generated with external tools, such as Wuggy (Keuleers & Brysbaert, 2010), which might be suitable to control for the phonotactic rules of more languages than the ones that we provide out-of-the-box. Externally-generated pseudowords must be akin to pseudowords that ALPARC would generate (i.e., all pseudowords must have a fixed syllable structure and a fixed number of syllables/phonemes) and should be imported as strings with a single vertical line ("|") splitting syllables and a double vertical line ("||") splitting pseudowords.

To further extend the usability of ALPARC to different languages, we also allow the upload of corpus data including syllable frequency of use and phoneme n-gram probabilities. Such corpora can then be used to apply similar frequency filters as those implemented in the previous step and to ensure that ALPARC produces pseudowords that do not comprise too (in)frequent phoneme combinations (see also *Statistical Analyses*). The output *Pseudoword Register* serves as a basis to create a *Lexicon Register*, where phonological rhythmicity confounds are systematically minimalized.

Step 3—from pseudowords to lexicons: phonological control

In generating artificial lexicons of pseudowords embedded in an isochronous stream of syllables, it is important to consider that phonological regularities provide landmarks that the brain could track instead of (or in addition to) statistical regularities (Di Liberto et al., 2015). Although phonological rhythmicity is likely not an issue for studies that use natural language (due to its high variability), it is important for artificial language learning studies (due to the low variability of the linguistic items), especially if combined with frequency-tagging paradigms (due to the rigid temporal structure of the stimulus). Indeed, phonological information and sub-phonemic cues have been previously argued to be essential for neural tracking of pseudowords (Chen et al., 2020). Thus, if a phonological feature is present at the same rate of TP patterns, statistical and phonological cues cannot be disentangled. In other words, if neighboring pseudowords share a phonological feature at the same phonemic position, but not at all other positions, that feature will be rhythmic and confound the TP rate.

To suppress phonological rhythmicity, we reasoned that limiting the opportunity for a feature to occur at regular intervals will inherently reduce its periodicity in structured streams. In particular, we intend to minimize phonological rhythmicity at the rate of TP patterns reoccurring at a fraction of the syllable rate to delineate pseudowords (e.g., 1/3 for trisyllabic pseudowords). To target this rate, ALPARC builds a matrix of position-wise phonological feature overlap across pseudoword pairs. For each pseudoword pair and each phonological feature, shared binary sequences that mimic the regular TP pattern will be flagged (Figure 2). These flags allow tracking of how many features would repeat identically among syllables that occupy the same relative position between neighboring pseudowords. Next, ALPARC derives a square matrix of feature overlap as the sum of flags across features for each pseudoword pair. Based on this matrix, ALPARC employs a *lazy evaluation strategy* to build sets of pseudowords while ensuring minimal cumulative feature overlap. Here, lazy evaluation means that pseudoword combinations with the least feature overlap are generated first, so that lexicons are only created when they are referenced in subsequent steps. This strategy minimizes feature overlaps while reducing the computation time relative to generating all possible combinations of pseudowords.

To compose the *Lexicon Register*, users could specify the desired number of pseudowords per lexicon (default = 4), the number of lexicons to be generated simultaneously (default = 5), and if such lexicons should be comprised of all unique words (default = true). Users may also specify the subsets of phonological features to control for (default = *sonorant* [*son*], *continuant* [*cont*], *lateral* [*lat*], *nasal* [*nas*], *voice* [*voi*], *coronal* [*cor*], *labial* [*lab*], *high* [*hi*], *back* [*back*] for consonants; and *labial* [*lab*], *high* [*hi*], *low* [*low*], *back* [*back*], *tense* [*tense*], *long* [*long*] for vowels). The output *Lexicon Register* contains sets of pseudowords with minimal cumulative feature overlap.

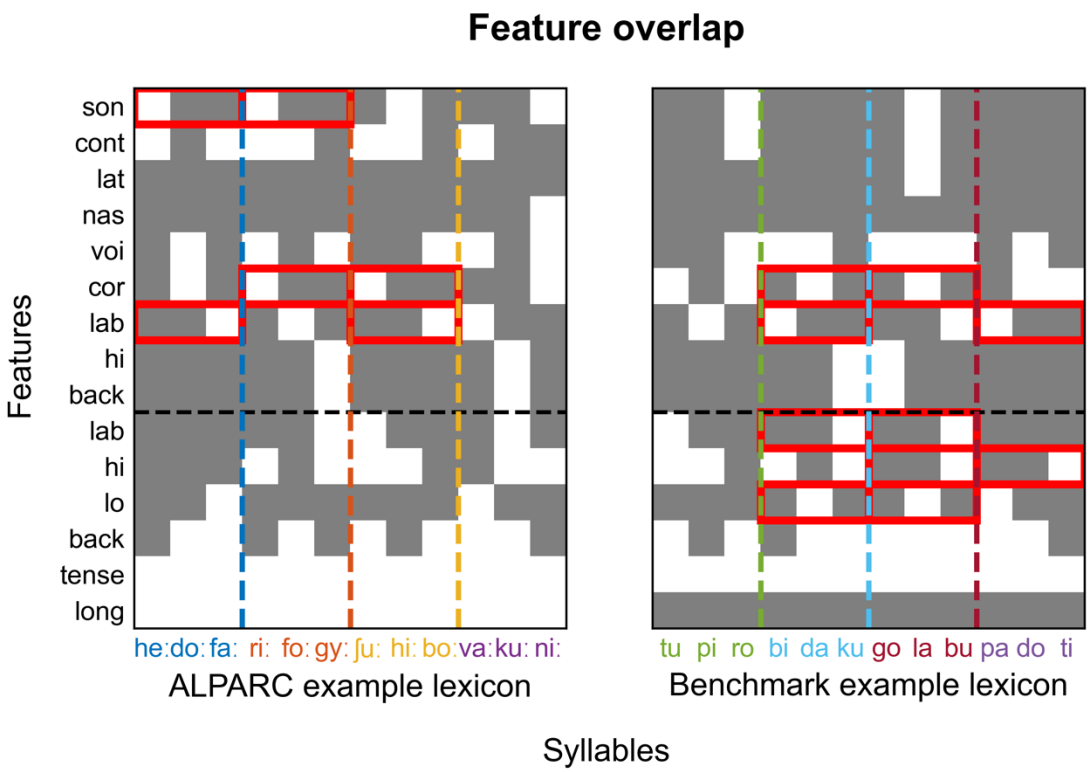

Figure 2. Binary matrices of phonological feature overlap across pseudowords pairs in two example lexicons. Left: feature matrix of one example lexicon generated with ALPARC. Right: feature matrix of one reference lexicon from the literature. Pseudowords are color-coded on the abscissae; vertical dashed lines at their boundaries cut the matrices vertically. The horizontal dashed line separates features associated with the consonant (top) or the vowel (below) phoneme of each syllable. Grey squares represent a minus (or a zero) in the binary feature vector of a given syllable (i.e., the feature is absent); whereas white squares represent a plus (or a one) for that feature (i.e., the feature is present). Red boxes flag repetitive patterns of phonological features across minimal pseudowords pairs that would confound the TP patterns at the pseudoword rate (e.g., [1 0 0, 1 0 0]). To avoid the confound, ALPARC uses a feature matrix for all possible pseudoword pairs and forms lexicons with the least cumulative feature overlap.

Step 4—from lexicons to streams: TP control

A common strategy to manipulate TPs in a stream is to randomly shuffle a fixed number of syllables or pseudowords with the only constraint that they do not repeat consecutively. However, this process may lead to some transitions being more likely than others due to randomness. As a result, TPs would be biased because tokens present different cue diagnosticity on the microscale of all individual pairs. Furthermore, under loose constraints, TPs may vary throughout the stream, leading to irregularities at the macroscale of the stream, with different TPs for the same pair at different stream locations. Statistical irregularities at the micro and macroscales are not exempt from having an impact on neural responses (Hasson, 2017). In fact, if TPs are variable and time-varying, some statistical boundaries can be perceived as more salient than others, with both behavioral and neural consequences in terms of learning and tracking. To suppress this statistical confound, ALPARC employs a new method that guarantees TP stationarity throughout the stream.

One additional caveat concerns the TP-uniform baseline condition. Recent studies pointed out that systematic acoustic differences at the pseudoword rate are actually visible in the acoustic spectra of TP-structured streams, due to the fact that syllables re-occur at fixed positions in the sequence, but they are absent in TP-uniform streams with randomly shuffled syllables (Luo & Ding, 2020; Pinto et al., 2022). Pinto et al. (2022) addressed this issue by using a position-fixed baseline, where syllables in the TP-uniform condition are also constrained to occupy the same position within a pseudoword-size chunk. In doing so, the inevitable pseudoword-peak is introduced also in the baseline condition, thereby guaranteeing that both the TP-structured and the TP-uniform streams have the same acoustic profile. In that study, frequency-domain analyses of brain responses revealed prominent peaks at the pseudoword rate (or its harmonics) for both streams, but higher peaks in the TP-structured relative to the TP-uniform stream (Pinto et al., 2022). These findings indicate that acoustic regularities partially account for neural tracking at the pseudoword rate, but also that SL effects are reliable if compared to an acoustically-matched baseline condition. Hence, this acoustic control is necessary for a safer interpretation of the neural responses at the pseudoword rate in terms of SL.

To precisely manipulate TPs between pseudowords in TP-structured streams and TPs between syllables in TP-uniform streams, ALPARC uses a *pseudo-random-walk* approach, which guarantees that each token follows every other token equally often. In TP-structured streams, each pseudoword is assigned to a token (from 0 to $n - 1$, where n is the number of pseudowords) and repeated a number of times (default = $n \times 15$). To ensure that TPs remain constant throughout the stream, new tokens are sampled iteratively based on a memory matrix of TPs. More specifically, the sampling of new tokens is restrained among those that have the lowest TP to the previous token based on the memory matrix. Moreover, every token must be used once before any could be repeated more than all others and consecutive repetitions of the same token are not allowed. Thus, ALPARC produces streams in which TPs at pseudoword boundaries tend to exactly $1/(n - 1)$. As pseudowords are flattened into plain sequences of syllables, TPs between syllables within any pseudoword are always 1.

ALPARC adopts a similar approach to control TPs in the TP-uniform streams as well. However, in this condition each token maps directly to a syllable (from 0 to $n \times m - 1$, where m is the number of syllables per pseudoword). To replicate recent evidence that repetitions of sounds at fixed positions induce acoustic regularities regardless the statistical structure (Pinto et al., 2022), we constructed two different functions for the TP-uniform position-random and the TP-uniform position-fixed streams. In position-random streams, consecutive repetitions of the same syllables are not allowed; syllables must transition to each other homogenously, and spread evenly across all possible positions within a pseudoword-size chunk. Under these constraints, the TPs across all syllables always tend to exactly $1/(n \times m - 1)$ throughout the stream. On the other hand, position-fixed baseline streams constrain syllable-tokens to occupy only one fixed position within a pseudoword-token (e.g., [0, 1, 2], [3, 4, 5], [6, 7, 8], [9, 10, 11], for $n = 4$, and $m = 3$) and to follow one another respecting this positional rule. As a result, syllable-tokens in position-fixed streams are still constrained to always "walk" the viable path they have walked the least, based on the memory matrix, but syllable TPs always tend to exactly $1/n$ throughout the position-fixed stream. For additional control analyses, we also built streams with different number of syllables ($m = \{2, 3, 4\}$) or pseudowords ($n = \{3, 4, 5\}$).

Finally, to quantify rhythmicity of any given phonological feature in a stream, we developed a *phonological rhythmicity index* (PRI). Each token is mapped either to a pseudoword (in TP-structured streams) or to a syllable (in TP-uniform streams). TP-structured streams are flattened into a connected stream of syllables, such that the phonological feature information can accompany each syllable. This information can be used to convert each stream into a binary matrix representing the presence/absence of any given feature over all syllable locations along the stream. The PRI is computed individually on each of these feature vectors as the correlation between a moving window of phonological feature information and binary convolutional kernels that represent two cycles of a periodic TP pattern (e.g., [1 0 0 1 0 0]) at the rate of interest (see *Statistical analyses*). To obtain streams with minimal PRI, the maximum PRI among all features may be set to be lower than an arbitrary threshold (default = none).

Step 5—from streams to audios: acoustic control

Pinto et al. (2022) adopted a position-fixed baseline to ensure that isochronous TP-structured and TP-uniform streams present identical peaks at the pseudoword rate in the envelope spectra. To show that position-random and position-fixed streams present different acoustic peaks at the pseudoword rate, we synthesized each individual syllable into separate audio files using Wavenet (Oord et al., 2016), and the Speech-Synthesis-Markup-Language (Taylor & Isard, 1997). Then, we edited each syllable using Praat (Boersma & Weenink, 2007) to equalize their duration to 300 ms and their amplitude to 65 dB. Afterwards, we concatenated all sounds into streams using the "wave" Python module. Finally, we conducted spectral analyses on the auditory streams in all TP modalities for one example lexicon (see *Statistical analyses*). Our custom Praat scripts for audio editing are also available at the ALPARC repository (see *Code availability*). These scripts generalize to the case of human recorded speech, as long as syllables are recorded individually, meaning that coarticulation is disregarded. While these acoustic controls are intended mostly for frequency-tagging paradigms, all other types of controls that are performed on the level of syllables, pseudowords, lexicons, and TP-controlled streams are more directly generalizable to the case of human recorded speech.

Statistical analyses

In *Step 1* and *Step 2*, we introduced a method to filter a *Syllable Register* or a *Pseudoword Register* based on corpus statistics. To that end, ALPARC requires a corpus of syllable tokens associated with their natural frequency of use, or a corpus from which phoneme statistics, such as bigram, trigram, or positional probabilities, can be extracted. If such corpus is provided, then frequencies or probabilities are log-transformed, normalized (z-scored), and tested against a uniform distribution. The p-value distributions stemming from the uniformity test on syllable frequency of use can be used to filter the *Syllable Register* based on one arbitrary threshold (default unigram α threshold = 0.05, two-tailed). Likewise, the *Pseudoword Register* can be filtered based on whether all phoneme bigrams or trigrams within each pseudoword are absent in the corpus (i.e., default α threshold = none), or if any of those significantly deviates from the uniform distribution (default α threshold = 0.05, two-tailed). ALPARC may also filter the *Pseudoword Register* based on positional statistics at the phoneme level. This filter operates by removing artificial pseudowords if they present one phoneme that rarely occupies one certain position in real words from the corpus (position-based default α threshold = 0.05, one-tailed).

In *Step 3*, we presented a strategy to build a *Lexicon Register* that minimizes the feature overlap across syllables that occupy the same position among all pseudowords in the lexicon. In *Step 4*, we introduced the PRI as a method to assess feature rhythmicity. To evaluate ALPARC performance, we generated 21 artificial lexicons using German (ALPARC-DEU) or English (ALPARC-ENG) corpora, 21 artificial lexicons as random syllable combinations (ALPARC-RND), and 21 artificial lexicons (BENCHMARK) extracted from the prior literature (Assaneo et al., 2019; Batterink & Paller, 2017; Cunillera et al., 2006; Kiai & Melloni, 2021; McNealy et al., 2006; Saffran et al., 1996). For each lexicon, we simulated 10 streams per TP mode (TP-uniform position-random, TP-uniform position-fixed, and TP-structured streams). We computed Pearson's correlations to assess if the feature overlap of an artificial lexicon predicts the corresponding maximum PRI across features at the stream level. Moreover, we compared the PRI distributions across the different lexicon types for each stream TP mode using two-sided Welch's independent sample t-tests to evaluate if ALPARC reduces the PRIs.

To demonstrate how the PRI is calculated, let us consider one TP-structured stream for one lexicon of four trisyllabic pseudowords. In this stream, TP patterns repeat periodically every three syllables, with TPs of 1 within pseudowords and .33 at pseudoword boundaries. Such a sequence can be recoded into a binary vector of "TP decays" marking low-TP events as 1 and high-TP events as 0 (i.e., [0 0 1 0 0 1, …]). Based on this TP pattern, we define kernels as vectors that represent two cycles of TP decays at all possible lags (i.e., [1 0 0 1 0 0], [0 1 0 0 1 0], [0 0 1 0 0 1]). This definition rests on the assumption that a sequence must repeat consecutively at least twice, for it to be classified as a pattern. Now, let us assume one vector $\boldsymbol{k}$ = [1 0 0 1 0 1 0 0 1 0 0 1 1 0 0], representing the presence (1) or absence (0) of one feature throughout a stream. As we convolve the three kernels through the feature vector $\boldsymbol{k}$, we assess the match between any of the kernels and the current portion of the feature vector. For instance, at the first iteration—$\boldsymbol{k}_{(1,6)}$ = [1 0 0 1 0 1]—we find no exact match with any of the kernels; hence, we note a non-match as a zero. Then, one step ahead, we find $\boldsymbol{k}_{(2,7)}$ = [0 0 1 0 1 0]. Again, no match. The fifth iteration, $\boldsymbol{k}_{(5,10)}$ = [0 1 0 0 1 0], presents an exact match to one of the kernels; so, we note a one. By the end of the stream—until $\boldsymbol{k}_{(10,15)}$—we obtain a total of 3 matches over 10 iterations. The PRI is the ratio between the sum of all such matches and the total number of comparisons, in this case: .33. This value indicates that 33% of the pseudoword tracking effect that might be attributed to TP patterns may instead be due to phonological regularities of the feature under examination. Thus, a lower PRI denotes a reduction of the phonological rhythmicity confound. This confound is further eased if corresponding TP-structured and TP-uniform streams have the same PRI.

In *Step 4*, we also introduced the pseudo-random-walk to generate TP-controlled streams. To assess whether the pseudo-random-walk improves TP stationarity, we compared the TP distribution across stream locations for streams simulated using the pseudo-random-walk approach or the classical pseudo-randomization used in the prior literature, which entails shuffling of the tokens with the only constraint that no consecutive repetitions of the same token are allowed. For each stream mode, we also compared the overall TP variance at each TP boundary (pseudoword boundaries in TP-structured and syllable boundaries in TP-uniform streams) among the two strategies, using Levene's tests.

In *Step 5*, we presented a method to convert the streams of syllables into isochronous auditory sequences. In order to assess whether the amplitude modulation spectra of speech streams in different TP modalities exhibit different peaks at the pseudoword rate, we conducted spectral analyses using MATLAB (The MathWorks, Inc.). First, we synthesized one stream per TP modality based on one example lexicon generated with ALPARC. Then, we computed the amplitude modulation spectra for each stream by squaring the absolute value of the Fourier transform of the artificial speech signals.

Results

Here, we present the output of ALPARC for the generation of *Lexicon Registers* (and streams thereof) under several settings. We compare artificial lexicons of pseudowords tailored to the statistics from German and English corpora (ALPARC-DEU and ALPARC-ENG), against random lexicons that do not implement any control (ALPARC-RND), as well as artificial lexicons used in the prior literature (BENCHMARK). Our results illustrate the effectiveness of using linguistic corpora for normalizing the probabilities of (co-)occurrence of syllables (and phonemes) in spoken languages. Moreover, our analyses demonstrate the robustness in reducing PRIs, and the precision of the pseudo-random-walk approach for manipulating syllable TPs. Crucially, we show that ALPARC produces lexicons with lower PRIs and lower TP variance not only relative to the random lexicons, but also to the benchmark.

Normalization of syllable frequency of use and n-gram phoneme statistics from linguistic corpora

A list of 6,397 syllable tokens paired to their frequency of use was acquired from a large dataset on conversational German (Schiel, 2010). To preserve phoneme timing constant across all syllables, we used ALPARC to skim the corpus and retain only open consonant-vowel syllables, that is, with long vowels based on the binary matrix of phonemes. To further demonstrate the robustness of this method on a different language, we used an English corpus (Baayen et al., 1996) that contained 7935 syllables, annotated with their respective frequency of use and retained 76 consonant-vowel syllables. Then, frequencies were log-transformed, normalized, and tested for non-uniformity (Figure 3).

Syllables that significantly deviated from uniformity were excluded from the German-based *Syllable Register* (α = .05, two-tailed), but not from the English-based *Syllable Register* (α = none) in order to attain an adequate number of tokens to generate *Pseudoword Registers* that also obey the obligatory contour principle with constraints on place and manner of articulation, and vowel identity (see *Step 2* in *Methods*). We filtered the German-based *Pseudoword Register* using bigram, trigram, and positional phoneme statistics from a different German corpus (Arnold & Tomaschek, 2016) to ensure that pseudowords do not contains too (in)frequent phoneme bigrams (α = .05, two-tailed) or non-existing trigrams (α = none). These filters allow for normalizing the frequency distributions of syllables and phoneme n-grams in our *Syllable* and *Pseudoword Registers*. For each *Pseudoword Register*, we sampled a subset of pseudowords (default sample size = 200) for further computations.

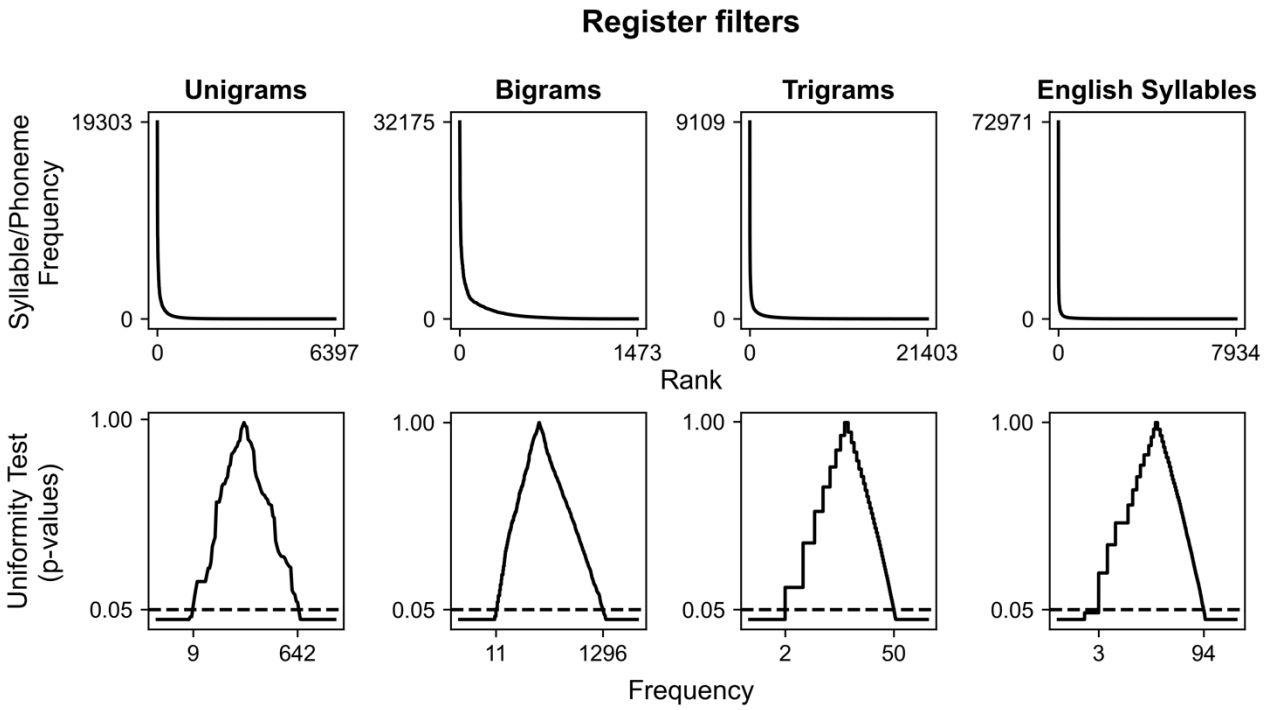

Figure 3. Corpora-based distributions of syllable frequency of use and phoneme n-grams probabilities. Top: Frequency of use of syllables and phoneme sequences are plotted as a function of frequency-ranked tokens. From left to right: German syllable frequency, phoneme bigrams, and trigrams, and English syllable frequency. Bottom: Distributions of p-values derived from uniformity tests based on the z-transformed frequency distributions above are plotted as a function of token frequency. The dashed horizontal line represents our default alpha threshold (.05).

Lexicon generation with minimal phonological rhythmicity index

To control for confounding phonological patterns at the pseudoword rate, we used ALPARC to make *Lexicon Registers* with the smallest possible feature overlap (see *Step 3* in *Methods*). This overlap is quantified at the stream level by the PRI (see *Step 4* in *Methods*). To show that constraints on lexicon generation influence the maximum PRI across features at the stream level, we correlated the feature overlap from 84 lexicons with the maximum PRI averaged across 10 streams (Figure S1). We found high Pearson's correlations in TP-uniform position-fixed (r = .81) and TP-structured (r = .87), suggesting that the feature overlap at the lexicon level influences the PRI at the stream level.

To assess whether ALPARC systematically reduces the PRI, we compared the distribution of PRI across phonological features for streams derived from the 21 ALPARC lexicons controlled for German (ALPARC-DEU) or English (ALPARC-ENG) corpus statistics, the 21 baseline lexicons with no linguistic nor phonological control (ALPARC-RND), and the 21 lexicons from the prior literature (BENCHMARK). Here, we considered the maximum PRI across all features as the most informative metric for assessing the confounding of phonological regularities, as it reflects the worst-case scenario of periodicity across features. For each stream TP mode (TP-uniform position-random, or position-fixed; and TP-structured), we calculated the maximum PRI for each sample per lexicon type (Figure 4) and compared the PRI distributions with two-sided independent sample Welch's t-tests (Table 1).

ALPARC-DEU and ALPARC-ENG produced a lower PRI than both the baseline lexicons and the benchmark for both TP-uniform position-fixed and TP-structured streams, but the benchmark outperformed the baseline for all stream modes. These results indicate that, on average, benchmarking lexicons reduced phonological rhythmicity compared to the baseline, but ALPARC further reduced this confound. To demonstrate that the PRI reduction is consistent across features, we also depict the distributions of PRI for all individual phonological features (Figure S2). To further demonstrate the robustness of our results, we simulated streams based on ALPARC-DEU with 2 or 4 syllables per pseudoword and with 3 or 5 pseudowords per lexicon (Figure S3). Overall, ALPARC performance under these settings is in line with the result that feature controls consistently reduce PRI.

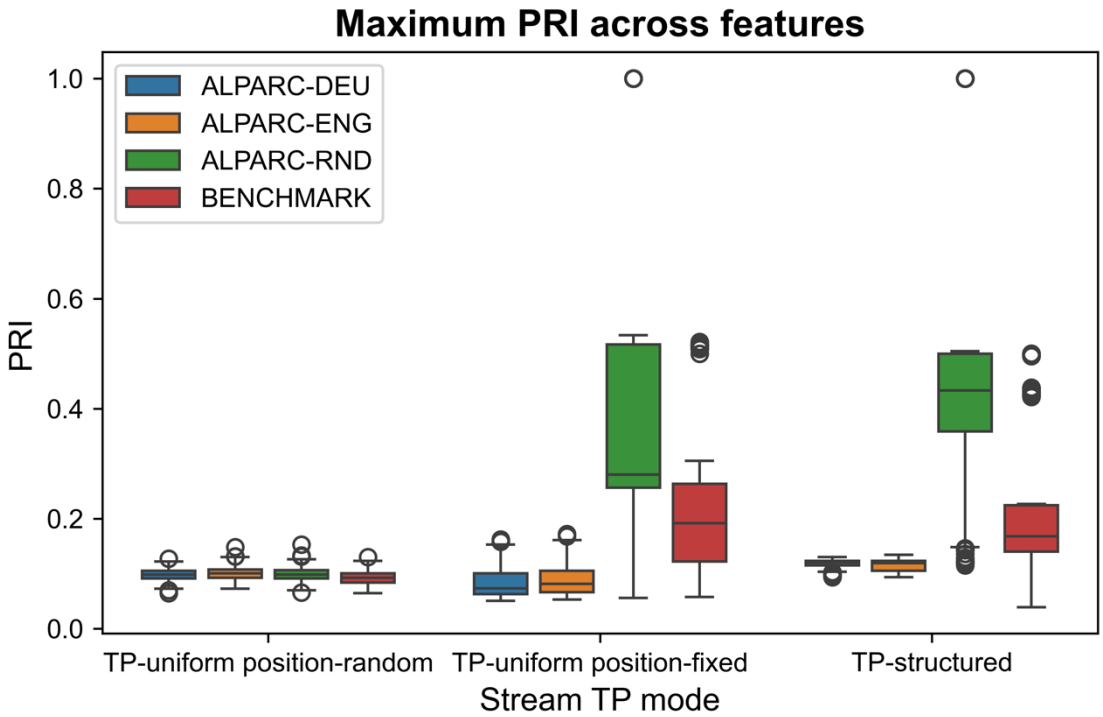

Figure 4. Phonological rhythmicity index. Boxplots of the distributions of maximum PRI across features for each stream TP mode (n = 10 streams per mode per lexicon) and lexicon type (n = 21 per type). Circles represent outliers.

Table 1

Independent Pairwise Welch's t-test results of maximum PRIs for all stream generation modes.

TP mode	Lexicon type	t	df	p-value	CI 95%	Cohen-d	power
TP-uniform position-random	ALPARC-DEU vs. ALPARC-ENG	-2.45	418	.015	[-0. -0.]	0.24	.69
TP-uniform position-random	ALPARC-DEU vs. ALPARC-RND	-1.15	450	.251	[-0. 0.]	0.09	.20
TP-uniform position-random	ALPARC-DEU vs. BENCHMARK	4.32	418	1.96 e-05	[0. 0.01]	0.42	.99
TP-uniform position-random	ALPARC-ENG vs. ALPARC-RND	1.65	400	.099	[-0. 0.]	0.14	.38
TP-uniform position-random	ALPARC-ENG vs. BENCHMARK	6.39	418	4.32 e-10	[0.01 0.01]	0.62	.99
TP-uniform position-random	ALPARC-RND vs. BENCHMARK	5.85	408	9.84 e-09	[0. 0.01]	0.50	.99
TP-uniform position-fixed	ALPARC-DEU vs. ALPARC-ENG	-2.14	418	.033	[-0.01 -0.]	0.21	.57

TP mode	Lexicon type	t	df	p-value	CI 95%	Cohen-d	power
TP-uniform position-fixed	ALPARC-DEU vs. ALPARC-RND	-22.3	430	8.96 e-74	[-0.36 -0.31]	1.34	1.0
TP-uniform position-fixed	ALPARC-DEU vs. BENCHMARK	-16.9	418	3.17 e-49	[-0.13 -0.1]	1.65	1.0
TP-uniform position-fixed	ALPARC-ENG vs. ALPARC-RND	-21.9	431	3.64 e-72	[-0.36 -0.3]	1.32	1.0
TP-uniform position-fixed	ALPARC-ENG vs. BENCHMARK	-16.1	418	1.33 e-45	[-0.13 -0.1]	1.57	1.0
TP-uniform position-fixed	ALPARC-RND vs. BENCHMARK	13.2	562	5.77 e-35	[0.18 0.25]	0.85	.99
TP-structured	ALPARC-DEU vs. ALPARC-ENG	3.12	385	.002	[0. 0.]	0.31	.87
TP-structured	ALPARC-DEU vs. ALPARC-RND	-27.6	408	2.46 e-95	[-0.39 -0.34]	1.67	1.0
TP-structured	ALPARC-DEU vs. BENCHMARK	-10.8	203	9.69 e-22	[-0.12 -0.08]	1.07	1.0
TP-structured	ALPARC-ENG vs. ALPARC-RND	-27.8	408	3.10 e-96	[-0.39 -0.34]	1.68	1.0
TP-structured	ALPARC-ENG vs. BENCHMARK	-11.1	203	1.26 e-22	[-0.12 -0.09]	1.10	1.0
TP-structured	ALPARC-RND vs. BENCHMARK	16.2	607	1.78 e-49	[0.23 0.29]	1.14	1.0

Stream generation with minimal TP variance

ALPARC implements a pseudo-random-walk approach to stabilize the TPs throughout the streams (see *Step 4* in *Methods*). Here, we show that ALPARC improves TP stationarity relative to a random shuffling method (Figure 5). Levene's tests revealed that ALPARC reduces the overall TP variance for both TP-structured ($F_{(1, 476)} = 35.99$, $p = 3.92 \times 10^{-9}$), and TP-uniform position-random streams ($F_{(1, 1436)} = 19.81$, $p = 9.21 \times 10^{-6}$). These results suggest that ALPARC pseudo-random-walk reduces the TP variance compared to randomly shuffled streams, and indicate that ALPARC produces more stable TPs within each stream.

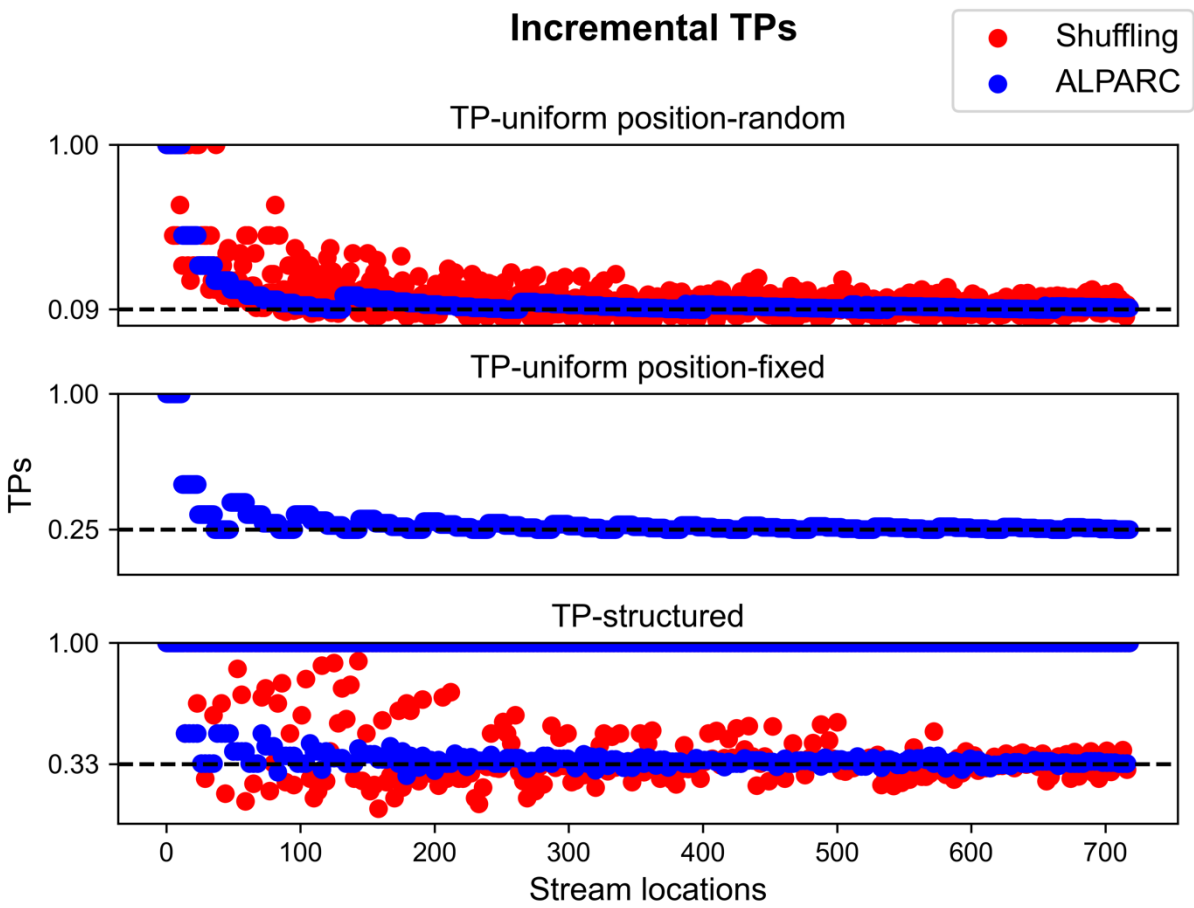

Figure 5. TP variance. Scatter plots of the observed TP values as a function of stream location, based on different pseudo-randomization strategies (colored dots). Dotted lines represent the theoretical value of TPs.

Acoustically-matched amplitude modulation spectra

To investigate if the re-occurrence of syllables at the same position within a pseudoword constitutes a systematic acoustic difference across TP modes that is visible in the frequency-domain (see *Step 5* in *Methods*), we compared the amplitude modulation spectra extracted from TP-uniform position-random, TP-uniform position-fixed, and TP-structured streams from one example ALPARC lexicon. The spectrum of the TP-uniform position-random stream displays no peak at the pseudoword rate (1.11 Hz); whereas the TP-uniform position-fixed spectrum shows a peak at that rate that is equivalent to the peak in the TP-structured stream (Figure 6). Our result corroborates previous findings that syllable positional constraints in isochronous streams induce systematic acoustic differences that are reflected in the amplitude modulation spectra (Pinto et al., 2022).

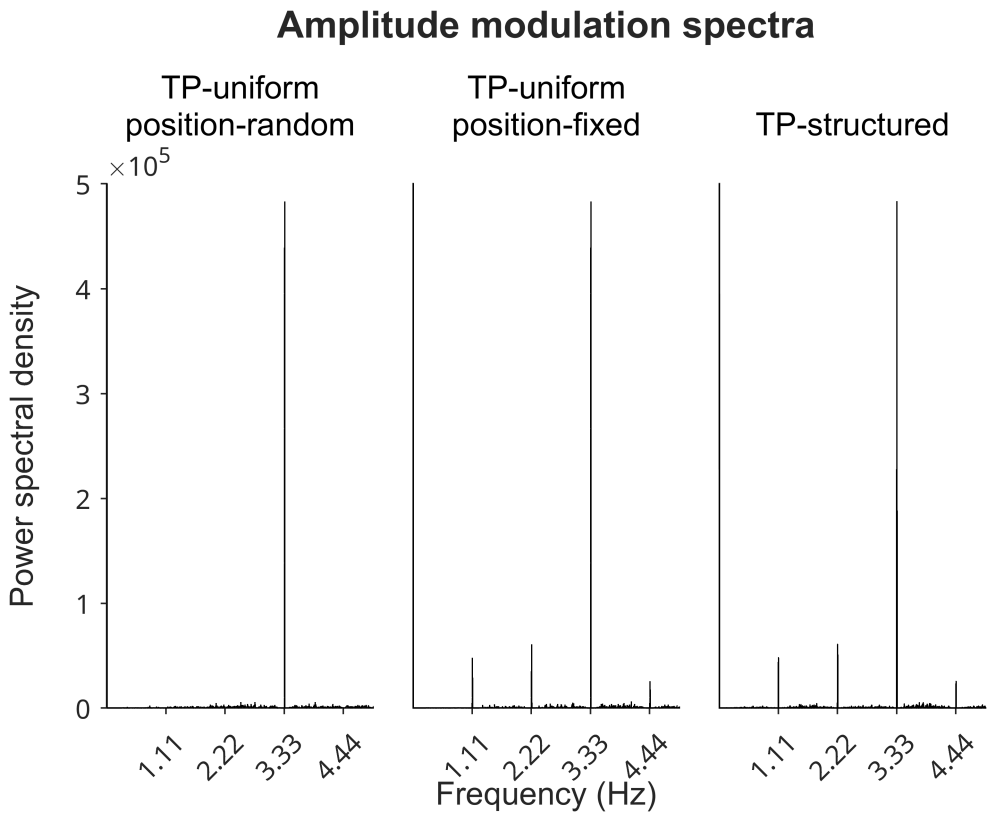

Figure 6. Speech modulation spectra. Power spectra from one example ALPARC lexicon in different stream TP modes. Each syllable had a fixed duration of 300 ms, leading to a peak at the syllable rate (3.33 Hz) in all modes. Syllables in the position-random stream occupy any possible location in the stream, leading to no peak at the rate of pseudowords (1.11 Hz). Syllables in the position-fixed and TP-structured streams are enforced to occur at fixed locations that are multiples of three, leading to a peak at the pseudoword rate (1.11 Hz), and its harmonic (2.22 Hz).

Discussion

We presented a new open-source toolbox (ALPARC) that controls for several confounding factors in linguistic SL studies, namely: prior knowledge of baseline frequency statistics in existing languages; phonotactics; rhythmic phonological patterns; TP precision and stationarity; and acoustic regularities. We described the functionalities of ALPARC step-by-step (see *Methods*). More detailed instructions on how to use ALPARC are provided in the "help" function of our library (see *Code Availability*). ALPARC generates artificial lexicons (of arbitrary size) comprised of pseudowords (of fixed length and syllable structure) that are arranged or scrambled in TP-structured or TP-uniform streams. Our results show that ALPARC eradicates all confounds that limit the interpretability of SL findings.

ALPARC tailors pseudoword generation to the statistics of existing languages

ALPARC provides a framework to model nested linguistic objects (e.g., pseudowords annotated with phonemic features). Thus, it can be flexibly adapted for multiple purposes that are optimized for, but not limited to, linguistic SL designs. In particular, ALPARC allows flexible stimulus design in SL studies by generating *Pseudoword* and *Lexicon Registers* of arbitrary size that are based on one fixed syllable structure (e.g., [open] consonant-vowel or consonant-vowel-consonant syllables). Crucially, ALPARC offers the possibility to tailor the generation of artificial languages to corpus-based statistics of real languages (e.g., German and English). To that end, ALPARC uses linguistic corpora to filter *Syllable* and *Pseudoword Registers* based on the frequency of use of syllables, as well as the phoneme bigram, trigram, and positional probabilities. Importantly, these filters moderate the interference that prior linguistic background may exert on SL performance (Elazar et al., 2022; Siegelman et al., 2018; Varela et al., 2024). Our linguistic controls not only alleviate linguistic entrenchment in behavioral SL studies (Siegelman et al., 2018), but are also essential in frequency-tagging. While computational models held that frequency and transitional constraints are dissociable (Erickson & Thiessen, 2015), both of these sources of information are tracked by the brain (Tobia et al., 2012). Therefore, neural tracking studies may also benefit from frequency controls that would mitigate this confound.

Moreover, ALPARC implements specific phonotactic controls to ensure that pseudowords obey the obligatory contour principle, a rule that has been shown to partially account for neural tracking at the rate of statistical patterns (van der Wulp, 2021). Phonotactics can also bias speech segmentation behaviorally (Onnis et al., 2005; Onnis & Thiessen, 2013). Therefore, ALPARC phonotactic controls may also be relevant for behavioral SL studies. While ALPARC linguistic and phonotactic controls are limited to German and English, and to the obligatory contour principle, users can also upload their own linguistic corpora or add custom code blocks to control for different phonotactic rules based on the native language of their target population. Alternatively, users can import an externally-generated *Pseudoword Register* and still use ALPARC to generate lexicons with minimal PRI and streams with precise TPs, while using own pseudowords that comply with phonotactic rules in any other language.

ALPARC suppresses phonological and acoustic confounds

Phonological and acoustic confounds pose a serious challenge for linguistic SL research (Luo & Ding, 2020; Pinto et al., 2022; van der Wulp, 2021; Varela et al., 2024). ALPARC addresses such confounds by computing a dedicated PRI that quantifies the periodicity of all phonological features at the rate of interest in the syllable streams; and by adopting a position-fixed baseline that matches the acoustic spectra of TP-structured and TP-uniform streams. Our results show that ALPARC lexicons reduce the maximum PRI relative to random lexicons of pseudowords that disregard feature controls, but also to benchmarking lexicons used in prior studies on linguistic SL. While the latter certainly attained a better phonological rhythmicity control than our baseline, which must be due to the expert conduct of the prior studies, ALPARC outperformed this benchmark. Lower PRIs denote a lower impact that the periodicity of phonological features may impose on the syllable streams. Thus, ALPARC lexicons and streams allow for safer interpretation of artificial language learning findings in terms of SL.

Neurons in the human superior temporal gyrus display selective tuning to different phonemes and phonological features (Mesgarani et al., 2014). In line with this finding, prior studies suggested that cortical tracking of speech might also reflect phoneme-level processing (Di Liberto et al., 2015; but see Daube et al., 2019). In addition, one recent perspective on language acquisition proposed that the infants' brain may track sustained phonological features in speech, as the permanency of such features through longer stretches of time matches the processing timescales of brain activity that they display at that age (Menn et al., 2023). Based on this work, it cannot be excluded that the mapping of phonetic features to phonological categories can result in periodic activity at the rate of phonological features that are regularly distributed in a sequence. In this context, our PRI might be useful in future studies to assess neural tracking of specific phonological features. For instance, minimal edits to the function that generates the artificial lexicons could lead to the creation of speech streams—artificially synthesized or recorded naturally—that maximize the PRI of one, or more, phonological features, which could be decisive for the systematic assessment of neural tracking of phonological features.

ALPARC generates sequences with precise TP distributions

TP stationarity has implications for the interpretability of SL findings (Hasson, 2017). To precisely manipulate TPs in a sequence of abstract tokens, ALPARC uses a new pseudo-random-walk method. Contrary to classical random-walks, our pseudo-random-walk does not select new tokens at random based on a pre-specified TP distribution; rather, it constrains next-token selection to always explore all possible paths evenly based on a memory matrix. Based on the Fundamental Theorem of Markov Chains, TPs derived from classical random-walks will approach the stationary distribution slowly, as the number of elements in the sequence grows. Instead, our pseudo-random-walk stabilizes TPs already for short sequences, and it does so far more precisely than random shuffling approaches. Our method can be applied for stimulus development in any SL design even beyond the linguistic domain.

Conclusion

Linguistic entrenchment, phonotactics, and acoustic/phonological regularities are serious confounds that limit the interpretability of behavioral and neurobiological findings on linguistic SL. To address such confounds, we developed ALPARC: a new open-source Python toolbox that generates *Artificial Languages with Phonological and Acoustic Rhythmicity Controls*. ALPARC produces TP-structured and TP-uniform streams of syllables with precise TP distributions. Streams are based on pseudowords that can be tailored to the statistics of existing languages. ALPARC lexicons and streams minimize the confounding rhythmicity of phonological features at the rate of TP patterns, making it especially suitable for study designs that require strict rhythmicity controls (e.g., frequency-tagging). ALPARC can be used as a dataset to download the artificial lexicons and streams that we generated in this paper based on some of the most conventional settings. Alternatively, ALPARC can be used as a library to generate new artificial lexicons and streams based on custom settings. Moreover, ALPARC can be used to diagnose stimuli used in prior studies in terms of phonological rhythmicity confounds, which can be useful for meta-analyses purposes. We recommend the use of ALPARC in future studies to rule out potential lurking variables from the implications of empirical findings on linguistic SL.

Acknowledgments

The study was supported by the Max Planck Society through the award of the Independent Max Planck Research Group Language Cycles to LM. NM was supported by BMBF (Federal Ministry of Education and Research) through ACONITE (01IS22065), and the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI.) Leipzig, and the Max Planck IMPRS CoNI Doctoral Program. The authors would like to thank Dimitra Kiakou and Shashank Shetty Kalavara for insights into the implementation of TP controls in the early stages of development, and Louisa Bogaerts and her lab for insightful comments into potential future developments of the toolbox. LT and LM conceptualized the research. LT, NM, and LM conducted the formal analyses. LT and NM developed the software. LT and NM created the visualizations. LM supervised the research. LM acquired funding. LT wrote the manuscript. LT, NM, and LM edited the manuscript. The authors declare no competing interests.

Code availability

All materials, data, and code are available at: https://github.com/milosen/alparc.

References

Arnold, D., & Tomaschek, F. (2016). *The Karl Eberhards Corpus of spontaneously spoken Southern German in dialogues-audio and articulatory recordings*.

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*, 627–632. https://doi.org/10.1038/s41593-019-0353-z

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The CELEX lexical database (cd-rom)*.

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Boersma, P., & Weenink, D. (2007). *PRAAT: Doing phonetics by computer (Version 5.3.51)*.

Bogaerts, L., Richter, C. G., Landau, A. N., & Frost, R. (2020). Beta-Band Activity Is a Signature of Statistical Learning. *The Journal of Neuroscience*, *40*(39), 7523–7530. https://doi.org/10.1523/JNEUROSCI.0771-20.2020

Boll-Avetisyan, N., & Kager, R. (2014). OCP-PLACE in Speech Segmentation. *Language and Speech*, *57*(3), 394–421. https://doi.org/10.1177/0023830913508074

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Chen, Y., Jin, P., & Ding, N. (2020). The influence of linguistic information on cortical tracking of words. *Neuropsychologia*, *148*, 107640. https://doi.org/10.1016/j.neuropsychologia.2020.107640

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., & 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

Cunillera, T., Toro, J. M., Sebastián-Galles, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*(1), 168–178. https://doi.org/10.1016/j.brainres.2006.09.046

Daube, C., Ince, R. A. A., & Gross, J. (2019). Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. *Current Biology*, *29*(12), 1924-1937.e9. https://doi.org/10.1016/j.cub.2019.04.067

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

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), 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

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output:

Lexicalization during artificial language learning. *Cognition*, *112*(3), 349–366. https://doi.org/10.1016/j.cognition.2009.05.002

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, *145*(12), 1128–1153. https://doi.org/10.1037/bul0000210

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

Hasson, U. (2017). The neurobiology of uncertainty: Implications for statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160048. https://doi.org/10.1098/rstb.2016.0048

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

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Kiai, A., & Melloni, L. (2021). What canonical online and offline measures of statistical learning can and cannot tell us. *bioRxiv*. https://doi.org/10.1101/2021.04.19.440449

Luo, C., & Ding, N. (2020). Cortical encoding of acoustic and linguistic rhythms in spoken narratives. *eLife*, *9*. https://doi.org/10.7554/eLife.60433

MATLAB. (2024). *Version R20240a*. The MathWorks Inc.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *The Journal of Neuroscience*, *26*(29), 7629–7639. https://doi.org/10.1523/JNEUROSCI.5501-05.2006

Menn, K. H., Männel, C., & Meyer, L. (2023). Does Electrophysiological Maturation Shape Language Acquisition? *Perspectives on Psychological Science*, *18*(6), 1271–1281. https://doi.org/10.1177/17456916231151584

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science (New York, N.Y.)*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. https://doi.org/10.1016/j.jml.2005.02.011

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–284. https://doi.org/10.1016/j.cognition.2012.10.008

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *Arxiv*. https://arxiv.org/abs/1609.03499

Ordin, M., Polyanskaya, L., Soto, D., & Molinaro, N. (2020). Electrophysiology of statistical learning: Exploring the online learning process and offline learning product. *European Journal of Neuroscience*, *51*(9), 2008–2022. https://doi.org/10.1111/ejn.14657

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Pinto, D., Prior, A., & Zion Golumbic, E. (2022). Assessing the Sensitivity of EEG-Based Frequency-Tagging as a Metric for Statistical Learning. *Neurobiology of Language*, *3*(2), 214–234. https://doi.org/10.1162/nol_a_00061

Ramers, K. H. (1998). *Einführung in die Phonologie*. Fink. https://books.google.de/books?id=fpLewAEACAAJ

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Schiel, F. (2010, May). BAStat: New Statistical Resources at the Bavarian Archive for Speech Signals. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*(5593), 553–554.

Siegelman, N., Bogaerts, L., Armstrong, B. C., & Frost, R. (2019). What exactly is learned in visual statistical learning? Insights from Bayesian modeling. *Cognition*, *192*, 104002. https://doi.org/10.1016/j.cognition.2019.06.014

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

Sjuls, G. S., Harvei, N. N., & Vulchanova, M. D. (2024). The relationship between neural phase entrainment and statistical word-learning: A scoping review. *Psychonomic Bulletin & Review*, *31*(4), 1399–1419. https://doi.org/10.3758/s13423-023-02425-9

Taylor, P., & Isard, A. (1997). SSML: A speech synthesis markup language. *Speech Communication*, *21*(1), 123–

133. https://doi.org/10.1016/S0167-6393(96)00068-4

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0056

Tobia, M. J., Iacovella, V., Davis, B., & Hasson, U. (2012). Neural systems mediating recognition of changes in statistical regularities. *NeuroImage*, *63*(3), 1730–1742. https://doi.org/10.1016/j.neuroimage.2012.08.017

van der Wulp, I. (2021). *Word segmentation: TP or OCP? A re-analysis of Batterink & Paller (2017)* [Master's Thesis].

Varela, I. G., Orpella, J., Poeppel, D., Ripolles, P., & Assaneo, M. F. (2024). Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning. *Cognition*, *245*, 105737. https://doi.org/10.1016/j.cognition.2024.105737

Zipf, G. K. (1932). *Selected Studies of the Principle of Relative Frequency in Language.* Harvard University Press.