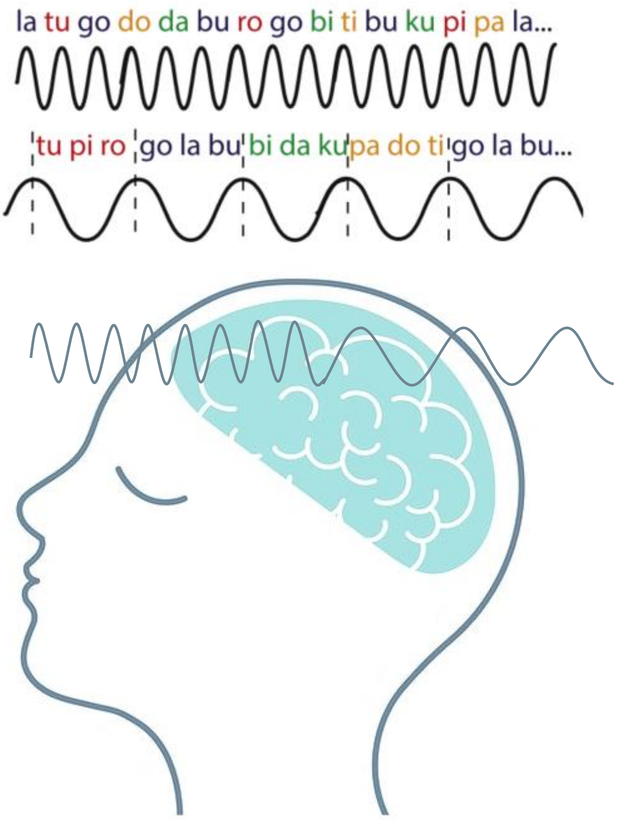
Word segmentation: TP or OCP?

A re-analysis of Batterink & Paller (2017)



Iris van der Wulp

Student number: 4295609

RMA Linguistics, Utrecht University

Research Master Thesis

Supervisors: Prof. Dr. Frank Wijnen & Dr. Marijn Struiksma

Second Reader: Dr. Karin Wanrooij

11-03-2021

Utrecht University

Acknowledgements

I would like to thank my supervisors Prof. Dr. Frank Wijnen and Dr. Marijn Struiksma for supporting me throughout the process of conducting this research and writing this thesis with great flexibility, inventiveness and patience. Whenever there was another unexpected hiccup, or when I needed help figuring out another technical issue, you were always willing to take a look and give advice. I always felt supported and at ease during our (online) meetings. Your expertise and guidance have shaped my research interests, and my plans for the future. I hope we can continue to work together through PhD in the Humanities.

I would also like to offer my sincere thanks Dr. Laura Batterink for sending me her EEG data and analysis script, as well as answering all my questions remarkably fast, all the way from the other side of the globe.

I want to thank my boyfriend, Floris van Kooten, for always supporting me, helping me relax, and being the best sparring partner for academic discussions.

I finally want to thank my parents Marga van Meulenaarsgraf and Pieter van der Wulp for their support and love.

Abstract

Research on statistical learning suggests that to segment speech into words, infants keep track of transitional probabilities (TPs) between syllables: the likelihood that syllable *X* occurs given syllable *Y*. TPs between neighboring syllables within words are higher than TPs at word boundaries. Batterink and Paller (2017) measured neural oscillations with EEG during statistical learning, which are known to phase-lock to the rhythm of an auditory stimulus. In the study of Batterink and Paller (2017), participants listened to a structured stream, consisting of four tri-syllabic words (TPs within words: 1.0, between: 0.33), and a random stream (TPs 0.09). Exposure to the structured stream but not the random stream led to an increase of phase-locking to the word frequency (1.1 Hz), compared to the syllable frequency (3.3 Hz).

However, some participants unexpectedly segmented the random stream into tri-syllabic units as well. The current study provides an alternative explanation for the findings of Batterink and Paller (2017) through the Obligatory Contour Principle (OCP) with a constraint on place of articulation (OCP-PLACE). Boll-Avetisyan and Kager (2014) showed that OCP-PLACE can influence word segmentation in Dutch. We performed a data re-analysis of Batterink and Paller (2017), replicating their analysis with Linear Mixed Modelling (LMM) and investigating the OCP-PLACE constraint as a possible alternative explanation of the data, including participants' triplet segmentation in the random stream.

We confirmed the statistical robustness of the results found by B&P2017, reporting the same results with our LMM approach as their ANOVA. Furthermore, we found a significant effect of OCP that is parallel to the effect of condition in the data of B&P2017. Further research should investigate the independent effects of OCP-PLACE on word segmentation in English and consider OCP-PLACE as a possible confounder that should be controlled for in further statistical language learning experiments.

Table of Contents

1. Introduction

The newborn infant faces an enormous challenge when it comes to learning its first words. While there are spaces between words in written language, these are not apparent in the continuous stream of soundwaves that is natural speech (as illustrated in figure 1). The infant needs to find out where the word boundaries are, in order to segment the speech and acquire a vocabulary. This is referred to as the problem of *word* or *speech segmentation*. A critical mechanism underlying word segmentation is hypothesized to be *statistical learning* (Saffran, 2003; Saffran, Aslin, et al., 1996).

1.1. What is statistical learning?

Statistical learning refers to the process by which organisms detect and internalize the statistical structure of (sequential or spatial) stimulus arrays. In particular, it has been shown that learners are sensitive to *transitional probabilities* (TPs) between (subsequent) units of stimuli. In natural language, these units can be syllables (Bertoncini & Mehler, 1981). The TP between syllables refers to the likelihood that a syllable *X* directly follows a given syllable *Y*. In natural language, the TPs of syllables that are part of the same word are usually higher than TPs of syllables at word boundaries, as is visualized in (1) (Saffran, 2003, p. 111).

(1) a. **pre-tty** ba-by
TP: 0.80
b. pre-**tty ba**-by
TP: 0.0003

In language acquisition, tracking TPs between syllables is hypothesized to be an underlying learning mechanism that aids the infant in solving the challenge of speech segmentation. Research has shown that both adults and infants are sensitive to TPs between neighboring syllables and can use this statistical information to segment word-like units from a stream of continuously spoken nonsense syllables (Batterink & Paller, 2017, 2019; Choi et al., 2020; Saffran, Aslin, et al., 1996, 1996; Saffran, Newport, et al., 1996).

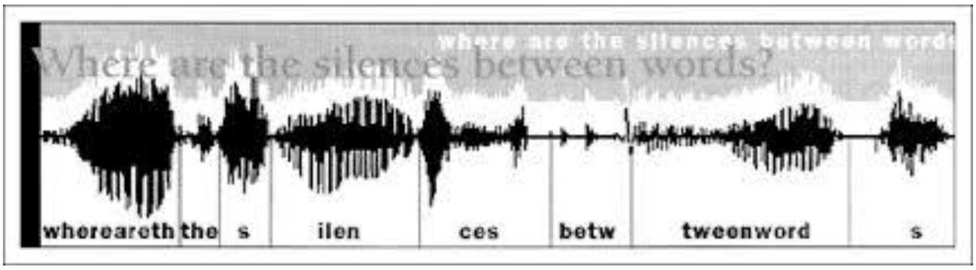


Figure 1. Waveform of the sentence "Where are the silences between words?" The y-axis shows the loudness, the x-axis the time. This example illustrates that there are no clear silences at word boundaries in the acoustic structure of a spoken sentence (Saffran, 2003, p. 111).

Typical statistical learning experiments (cf. Saffran, Aslin, et al., 1996; Siegelman & Frost, 2015) consist of two phases: a familiarization phase and a test phase. In the familiarization phase, the participant listens to an artificial language that contains a few multi-syllabic pseudo-words (from now on ‘words’), as illustrated in (2a). These *speech streams* are usually generated by a speech synthesizer, and controlled for any acoustic information that could cue word boundaries such as prosody, pauses, and stress differences. The TPs for syllables within each word are 1.0 (they always occur together), whereas TPs for adjacent syllables at word boundaries is lower (~ 0.33 in (2a); words are presented in a pseudo-random order where the same word cannot repeat consecutively).

(2) a. **Familiarization phase**

...tupirogolabubidakupadotigolabubidakutupiropadoti...

b. **Test phase**

Which is a word from the language you just heard?

tupiro	or	godapi?
(word)		(non-word)

In the subsequent test phase, participants are tested on their knowledge of the words from the artificial language. For adult participants, this is usually done with a *two-alternative-forced-choice* (2AFC) task containing a word from the artificial language and a non-word that was not present in the artificial language (2b). Participants then have to choose which word was present in the language they just listened to, and which was not. If the task is performed with infants, the 2AFC task is presented auditorily, where significant differences in listening times to the words and non-words indicate that the infant can detect a difference in what is familiar and what is new (Gómez & Gerken, 2000; Saffran, Aslin, et al., 1996).

Batterink and Paller (2017) point out that statistical learning comprises two stages: *identification* and *memorization*. In the identification stage, TPs between syllables are taken into account by the language learner, who unconsciously shifts their perception from individual syllables to word-like units. Thus, this is where the statistical structure of the perceived speech stream is detected by the learner and short-term representations of words are formed. In the memorization stage, these words are stored in long-term memory. Classical statistical learning experiments as illustrated in (2) cannot distinguish the identification from the memory storage component, since they only measure the word memorization outcome in the test phase. Batterink and Paller (2017) demonstrated that the identification component of statistical learning can be measured via participants’ brainwaves.

1.1.1. Measuring neural oscillations during statistical learning

Batterink and Paller (2017) used *electroencephalography* (EEG) as a means to probe the identification component *during* statistical learning, by measuring the *neural oscillations* of their participants. This paragraph will discuss in more detail what neural oscillations are and how they can be measured during statistical learning.

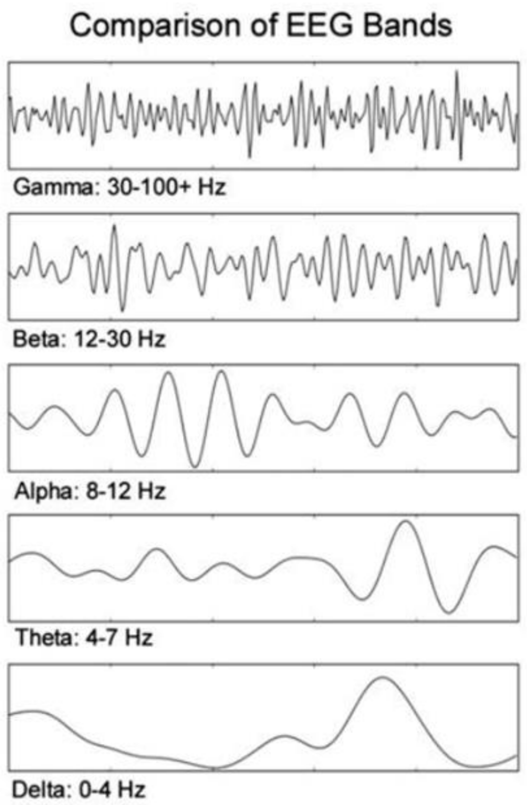


Figure 2. Oscillatory frequency bands and their frequencies in Hz (Nacy et al., 2016, p. 141).

Rhythmic brain activity is always present. A neural oscillation consists of successive ‘waves’ in the EEG output, showing this activity. Different neuron populations in the brain fire in patterns of a certain frequency, shifting in their excitability from relatively depolarized to relatively hyperpolarized (Peelle & Davis, 2012). As the quote below illustrates, neural oscillations can be characterized by their frequency range, called *frequency bands*. The most common ones can be seen in figure 2.

“***Neuronal oscillation***: the periodic shifting of a neuron or neuronal ensemble between high and low excitability states (phases), at some frequency in cycles per second or Hertz (Hz). Neuronal oscillations are often characterized by the frequency range (band) they occupy in the spectrum” (Schroeder et al., 2008, p. 106).

Neural oscillations have previously been shown to *phase-lock* (also referred to as *entrainment* or *synchronization*) to the rhythm of an auditory stimulus, in order to optimally process such stimuli (cf. Buiatti et al., 2009; Kabdebon et al., 2015; Peelle & Davis, 2012). The phase of the neural oscillations is adjusted to match the (quasi-)rhythmic phase of the auditory signal. If this alignment is such that inputs arrive at the time of highest neuronal excitability (hyperpolarization), the auditory inputs can be processed with maximal efficiency. Peelle and Davis (2012) argue that oscillations phase-locked in this optimal way can be thought of as making a prediction about the timing of upcoming critical information.

A crucial acoustic cue for the temporal characteristics of speech is found in the low-frequency amplitude fluctuations in the speech signal, which correspond to the approximate duration of a spoken syllable (~3-8 Hz; Delta and Theta frequency bands). These low-amplitude fluctuations resulting from the speaker's jaw movements corresponding to syllables (the jaw opens for the vowel), provide a foundation for the other temporal and hierarchic characteristics of speech (Ladányi et al., 2020; Peelle & Davis, 2012). Subsequently, *nested* oscillations in other frequency domains can be derived from oscillations phase-locked to syllables. High-frequency oscillations are proposed to track the phonemic information in the speech signal, while low-frequency oscillations process bigger units such as words and phrases (Peelle & Davis, 2012).

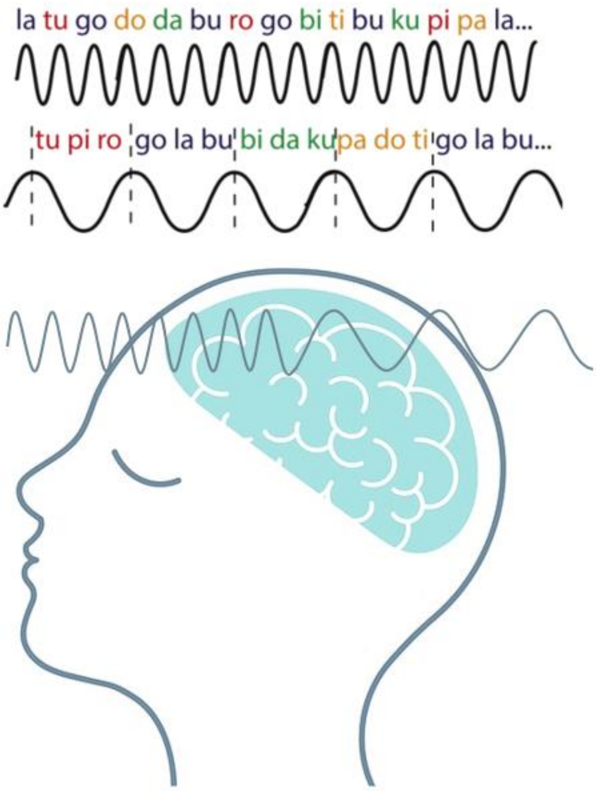


Figure 3. Example of statistical learning, visualized by neural oscillations. Neural phase-locking shifts from the syllable frequency (faster waves) to the word frequency (slow waves).

Research on statistical learning employing EEG provides valuable insights in the speech segmentation process, complementing the traditional offline (word learning) approaches. Figure 3 illustrates phase-locking in statistical learning. In participants who are sensitive to the TPs in the signal, neural phase-locking shifts from the frequency of individual syllables to the frequency of multi-syllabic words. These participants also perform better on offline (word recognition) tests, compared to participants who continuously showed more phase-locking to the syllable frequency (Batterink & Paller, 2017; Choi et al., 2020).

1.1.2. Batterink and Paller (2017)

Batterink and Paller (2017, henceforth B&P2017) presented participants with two speech streams; a *structured* stream and a *random* stream. The structured stream consisted of four tri-syllabic nonsense words. TPs between syllables within each word were 1.0, whereas the TPs of syllables at word boundaries were 0.33, as the same word did not repeat consecutively.

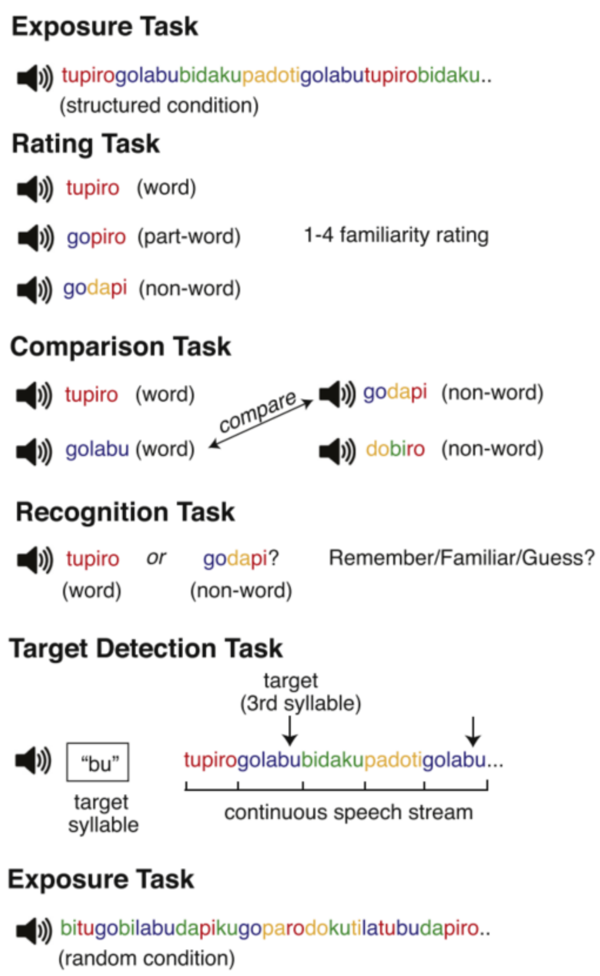


Figure 4. Summary of B&P2017's experimental design (p. 35)

The random stream consisted of an unstructured sequence of 12 syllables, with TPs being 0.09 overall, as the same syllable did not repeat consecutively. There were two syllable inventories, yielding two structured streams and two random streams in total. Every participant was exposed to a structured stream from one syllable inventory and a random stream from the other syllable inventory. The duration of each syllable was 300 msec. and each stream lasted for twelve minutes, divided into three blocks of four minutes each. During the exposure to the speech streams, EEG was recorded with 64 active electrodes. After exposure to the structured stream, participants completed four post-exposure tasks: a rating task, a comparison task, a recognition task, and a target detection task (figure 4). The analysis included only the rating task for explicit memory of the words from the structured stream, and the target detection task for implicit memory of these words.

To determine the amount of phase-locking to the syllables and words, B&P2017 used the Inter-Trial Coherence (ITC) formula to determine the amount of phase-locking to the syllable frequency (3.3 Hz, as the length of one syllable was 300 msec.) and the word or tri-syllabic frequency (1.1 Hz) in both the structured and random speech stream conditions. The ITC measure ranges from 0 (no phase-locked activity to a given frequency) to 1 (completely phase-locked activity to a given frequency). From these ITC scores, B&P2017 calculated a Word Learning Index (WLI), which can be viewed in (3).

(3) $$\text{WLI} = \frac{ITC_{word\ frequency}}{ITC_{syllable\ frequency}}$$

The WLI increases when there is more phase-locking to the word frequency than to the syllable frequency, and decreases when there is more phase-locking to the syllable frequency than to the word frequency. B&P2017 expected to find higher WLI scores which would increase over the duration of exposure when participants were exposed to the structured stream, versus lower WLI scores when participants were exposed to the random stream, which would be similar throughout exposure.

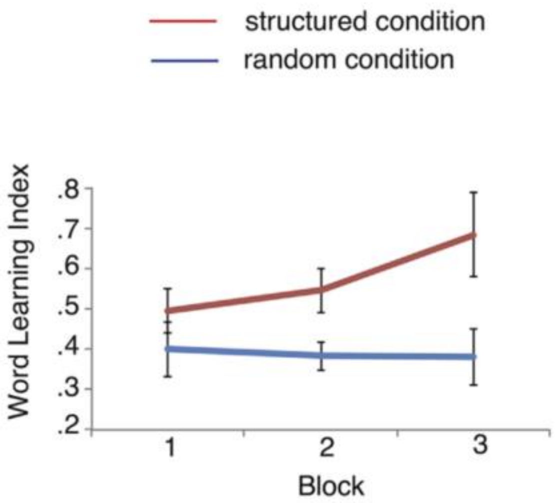


Figure 5. Word Learning Index (WLI) results per block and condition (B&P2017, p. 40). The WLI increased for the structured condition as a function of exposure per block, while the WLI for the random condition remained relatively constant.

The results obtained with an ANOVA indicated that the WLI in the structured condition was indeed higher than in the random condition (F (1, 44) = 17.3, $p < .001$; p. 39). The structured WLI also increased over the blocks of exposure, yielding higher WLI scores as exposure to the structured stream lasted longer and no increase in the WLI score during exposure to the random stream (F (2, 88) = 3.72, p = .029; p. 39). This is also illustrated in figure 5. With respect to the behavioral results, word rating accuracy on the post-exposure rating task was significantly above chance: 62.1% (SD = 14.3%; t (23) = 21.3, p <.001; p. 38). Words presented in the structured condition were rated as most familiar, followed by part-words, followed by non-words that were not present in the structured stream as least familiar. Reaction times on the target detection task became faster as the target syllable occurred later in a word from the structured condition. When the target syllable was the last syllable of a word, participants responded faster than when the target syllable was the middle or first syllable of the word. This indicated that participants who had acquired (implicit) knowledge of the TPs between syllables of a word, could predict its last syllable and thus reacted faster than for syllables in other positions.

B&P2017 used correlations to determine whether the WLI influenced subsequent post-exposure task performance. The structured WLI significantly correlated with the reaction times for the target detection task (r = .42, p = .039, WLI log-transformed; p. 41), indicating that participants with a higher WLI during exposure to the structured stream showed faster reaction times on the target detection task. However, the WLI in the random exposure condition unexpectedly also significantly correlated with the reaction times on the target detection task (*r* = .59, *p* = .003; p. 41). Correlations between the WLI and the rating task were positive but did not reach significance (structured WLI: rating accuracy: *r* = .30, *p* = .16; rating score: *r* = .32, *p* = .12; random WLI: rating accuracy: *r* = .22, *p* = .29; rating score: *r* = .34, *p* = .11; p. 41). To summarize, B&P2017 demonstrated that the EEG signal reflected a perceptual shift in their participants from syllable units to trisyllabic word-units in the structured stream, making the identification stage of statistical learning visible. Their study also confirmed their hypothesis that the word identification component of statistical learning is correlated with performance on the post-exposure target detection reaction time task, reflecting implicit learning.

An interesting point that should not be overlooked is that B&P2017 state in their discussion that the WLI significantly correlated with the reaction times on the target detection task in not only the structured condition (*r* = .42, *p* = .039; p. 41), but also in the random condition (*r* = .59, *p* = .003; p. 41). Moreover, the WLI scores in the structured and random conditions were highly correlated across participants as well (*r* = .63, *p* = .001; pp. 39 & 42). It thus seems that some participants unexpectedly segmented the random speech stream into tri-syllabic units as well, even though the TPs between the syllables used in the random stream were all the same and therefore could not have given rise to such a segmentation. Multiple explanations are possible for this effect. B&P2017 point to a "general tendency of an individual to seek out underlying patterns in the environment, particularly at the triplet level" (p. 42). They hypothesize that some individuals might impose a triplet structure on incoming stimuli, where others might impose a duplet or quadruplet structure and again others would show more bottom-up processing, not imposing any structure at all. In the case of their study, the triplet-imposing individuals would have an advantage, providing high WLI scores and better performance on the post-learning tasks, even though this would not be caused by statistical learning cued by TPs.

The presentation order of the conditions (structured stream first versus random stream first) is also mentioned by B&P2017 as a possible explanation for this effect. The structured and random WLI significantly correlated with respect to the order of stimulus presentation, where the correlation was stronger for participants who received the structured stream first (*r* = .74, *p* < .001; p. 39) than participants who received the random stream first (*r* = .46, *p* = .024; p. 39), although this difference did not reach significance (*z* = 1.43, *p* = .076, one-tailed; p. 39) and did not influence the other main effects such as the effect of condition and the interaction between condition and block (all *p* values > .1; p. 39).

Another possible explanation for the finding that some individuals also segmented the random stream into triplets, not explored by B&P2017, can be sought in the phonological properties of the syllables (and tri-syllabic words) in the speech streams they used. Of specific interest here is the phonotactic Obligatory Contour Principle. The current study will explore this alternative explanation after explaining what the Obligatory Contour Principle entails (paragraph 1.2). We will also re-analyze the data of B&P2017 with Linear Mixed Models (LMM), which have been shown to provide more reliable results for this kind of data than the ANOVA method (Aarts et al., 2014; Boisgontier & Cheval, 2016). The rationale for this methodological consideration will be discussed in section 1.3.

1.2. The Obligatory Contour Principle as a cue for word segmentation

Boll-Avetisyan and Kager (2014) showed that the *Obligatory Contour Principle* (OCP) with a constraint on the feature of place of articulation (OCP-PLACE) in phonology can influence word segmentation. The OCP requires subsequent phonemes to be featurally non-identical within words. OCP-PLACE specifically prefers avoidance of consonants with the shared feature [PLACE]. This effect crosses intervening vowels, thus constraining consonants adjacent when intervening vowels are left out of the equation (Boll-Avetisyan & Kager, 2014). In the case of the artificial language used by B&P2017 and other statistical learning experiments, the syllables comprise of a consonant-vowel (CV) structure. Thus, words in such artificial languages have a CVCVCV structure. The OCP-PLACE constraint would favor a distribution where the three consonants within these tri-syllabic words would have different [PLACE] features. This has been attested to have a psychological reality in multiple languages (Boll-Avetisyan & Kager, 2014; Coetzee, 2010). For instance, native speakers of Dutch reject nonwords that violate OCP-PLACE faster than nonwords that adhere to the OCP-PLACE constraint (Shatzman & Kager, 2007).

Boll-Avetisyan and Kager (2014) searched the Corpus Gesproken Nederlands, a phonetically transcribed corpus of spoken Dutch (Goddijn & Binnenpoorte, 2003) for the adherences and violations of OCP-PLACE in the Dutch language, both within words and across word boundaries. They found that CVC sequences in Dutch where the consonants share the feature [PLACE] are under-represented in the Dutch lexicon and spontaneous speech. Sequences of OCP-PLACE violating labials (P) and dorsals (K) were found to be more underrepresented than OCP-PLACE violations with coronals (T). Example (4) illustrates which Dutch consonants fall into these categories.

(4) labials = /p, b, f, v, w, m/ → P
dorsals = /k, g, x, ŋ/ → K
coronals = /t, d, s, z, ʃ, ʒ, r, l, n/ → T

Boll-Avetisyan and Kager (2014) experimentally confirmed that Dutch participants listening to an artificial language with syllables of a CV structure, and the consonants of those syllables altering between OCP-PLACE violations and adherences such as …PPTPTPPT… preferred segmenting this language into PTP items over OCP-violating PPT and TPP items. Apparently, Dutch listeners tend to assume that a word boundary will fall between two adjacent consonants that share the feature [PLACE], as is consistent with the distribution of OCP-PLACE in the Dutch lexicon.

With respect to English, co-occurrences of dorsal, labial and coronal consonants within syllables and/or words are also underrepresented in the lexicon (Dmitrieva & Anttila, 2008; Frisch, 1996; Monaghan & Zuidema, 2015). This seems to be a gradient effect, similar to Dutch; the larger the number of (other) consonant phonemes separating two consonants, the smaller the likelihood that listeners will assume a word boundary based on the OCP-PLACE constraint (Dmitrieva & Anttila, 2008; Frisch, 1996). It thus seems that English has a similar OCP-PLACE effect as Dutch. A study such as Boll-Avetisyan and Kager (2014), experimentally testing the influence of OCP-PLACE on word segmentation in English has not yet been performed. However, Coetzee (2010) showed that native speakers of English judged nonwords violating OCP-PLACE as less well-formed than nonwords adhering to OCP-PLACE. This is similar to the results found by Shatzman and Kager (2007) for Dutch nonwords violating OCP-PLACE. Therefore, we could expect listeners to prefer the segmentation of a speech stream in such a way that consecutive syllables with similar-place consonants serve as a cue for a word boundary. Participants of B&P2017 were adult native speakers of English. Therefore, the OCP-PLACE constraint could be an alternative explanation for their data, including the unexpected triplet segmentation of the random speech stream by some participants. This leads us to the first research question of the current study:

1. Can the results by B&P2017 be explained by the phonotactic OCP-PLACE constraint? In other words, can it be the listener's sensitivity to the OCP-PLACE constraint that cues word boundaries instead of sensitivity to TPs?

1.3. ANOVA versus Linear Mixed Models

An interesting development in experimental science is the upcoming use of *Linear Mixed Models* (LMM) for statistical analyses. B&P2017 used ANOVA to analyze their data, but also switched to LMM in their later work (Batterink, 2020; Batterink & Paller, 2019; Choi et al., 2020). ANOVA is a statistical model that can test differences between means in more than two conditions. However, ANOVA assumes the independence of observations. B&P2017 cannot accommodate to this assumption, since the study had every subject listen to multiple presentations of the same words/syllables, and data was collected from multiple electrodes, measuring multiple neurons per electrode in each subject (Boisgontier & Cheval, 2016). Aarts et al. (2014) call this a *nested design*: “Nested designs are designs in which multiple observations or measurements are collected in each research object” (p. 491). A violation of the independence of observation assumption necessary for ANOVA (such as a nested design) is associated with an increase of false positives; type I errors (Aarts et al., 2014; Boisgontier & Cheval, 2016; Nieuwenhuis et al., 2011). A common workaround for this problem is to aggregate over electrodes and trials, which makes the ANOVA less prone to type I errors. B&P2017 used this method by aggregating over electrodes and by summarizing the trials of their study into three blocks per condition. However, while repeated measures ANOVA allows assessing the effects of within-subject factors, it cannot deal with several data points for a subject in the same condition. Repeated measures ANOVA can only account for non-independent observations from one subject if each observation is made in a different condition, which in the case of B&P2017 would have led to the impossible task of having a different condition per presented word or block.

Statistical analyses using Linear Mixed Models (LMM) are perfect for data collected with nested designs. LMM treats the data points within a subject (each word/syllable presentation in the speech streams) as level 1 units, which are nested in a level 2 unit: the participant and so on (Aarts et al., 2014). Moreover, LMM can take crossed data structures into account, where the same subjects are observed in multiple conditions (Baayen et al., 2008; Boisgontier & Cheval, 2016). This is also the case in B&P2017, as each subject was exposed to both a random and a structured speech stream (condition). Finally, LMM takes continuous effects that unfold during the course of an experiment into account, while considering potential continuous covariates as well. These qualities of LMM provide us with a perfect statistical test to confirm the robustness of the findings from B&P2017, in line with their later studies (Batterink, 2020; Batterink & Paller, 2019; Choi et al., 2020). Therefore, our second research question is as follows:

2. Can we reproduce the results found by B&P2017, by re-analyzing their data with Linear Mixed Models instead of ANOVA?

1.4. Current study

In the current re-analysis, we used LMM to replicate and extend B&P2017's results, answering our two research questions. With respect to the first research question, we hypothesized that the OCP-PLACE constraint has contributed to participants' segmentation of the input strings into trisyllabic words, both in the structured and random conditions. Participants of B&P2017 were native speakers of English, a language which shows an effect of OCP-PLACE which is hypothesized to cue segmentation as is the case in Dutch. The OCP-PLACE constraint was hypothesized to explain the data of B&P2017 to some extent, or even more accurately than the random and structured conditions. Particularly in the random stream such OCP-influence was expected, because some participants have shown a triplet segmentation of this condition, even though the TPs between syllables did not give rise to such an underlying triplet structure. The effect of the OCP-PLACE constraint was hypothesized to be expressed as more phase-locking to the triplet frequency (a higher WLI) if a speech stream (accidentally) adhered to the OCP-PLACE constraint more, as opposed to a lower WLI if the speech stream adhered less to the OCP-PLACE constraint.

Presentation order of the conditions (structured, random) was also hypothesized to influence the results. As explained under 1.1.2, B&P2017 found that participants who were exposed to the structured stream first yielded a higher correlation between the structured WLI and the random WLI ("structured first": $r = .74$, $p < .001$; "random first": $r = .46$, $p = .024$; p. 39). Thus, we hypothesized that there is a higher WLI in the random condition for subjects who received the structured condition first.

Our second research question addresses the point made in paragraph 1.3, arguing that LMM is a more reliable statistical test for the type of study performed by B&P2017 than ANOVA. We hypothesized that the effect found by B&P2017 would be robust, showing an increase of the tri-syllabic frequency of 1.1 Hz. ($ITC_{word\ frequency}$) and a decrease of the syllable frequency of 3.3 Hz ($ITC_{syllable\ frequency}$) in the structured condition, corresponding to an increase of the WLI. On the other hand, no such increases/decreases were expected in the random condition. The random condition was expected to show a constant rate of $ITC_{syllable\ frequency}$ which should be larger overall than the $ITC_{word\ frequency}$, corresponding to a constant WLI value which is lower than the WLI in the structured condition (figure 5).

We report our re-analysis method and results in the next chapters, replicating the EEG analysis of B&P2017 and applying LMM to obtain our statistical results for this re-analysis. We first describe the dataset we received from B&P2017. We then report how we replicated the EEG analysis method of B&P2017 and created a new variable which measures the adherence to the OCP-PLACE constraint in each speech stream used by B&P2017. We statistically tested with LMM if differences between the speech streams with respect to the amount of adherence to the OCP-PLACE constraint were present. We also replicated the ANOVA by B&P2017, after which we built two models using the LMM approach: one model replicating B&P2017, and one model where our OCP variable took the place of B&P2017's condition variable.

2. Methodology

2.1. Description of the data set

Dr. Laura Batterink kindly sent us the EEG files and MatLab (*MATLAB*, 2019) analysis script of B&P2017, as well as their code of the experiment in Presentation® (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The data set contained raw EEG data of 46 participants, acquired during B&P2017's exposure task in two conditions per participant: structured and random. As described in paragraph 1.1.2 above, the structured condition contained tri-syllabic nonsense words, with TPs within words being 1.0 and between words 0.33. The random condition consisted of 12 nonsense syllables presented pseudo-randomly, with TPs being 0.09 overall. All syllables in both conditions were presented at a rate of 300 msec. per syllable, yielding a frequency of 3.3 Hz for syllables and 1.1 Hz for tri-syllabic (word) units in both conditions. Each condition was presented to all participants in three blocks of approximately four minutes each.

B&P2017 recorded data from a total of 47 participants. They excluded two participants due to technical issues in the EEG data acquisition, as is also reported in B&P2017 (p. 34). We analyzed the same EEG data as B&P2017 of 45 participants. Half of the participants of B&P2017 additionally performed behavioral post-exposure tasks, testing their explicit and implicit memory of the words from the structured stream, as described in paragraph 1.1.2 above. The current study focusses on re-analyzing the EEG data from the exposure task and our data set did not include this behavioral data.

2.2. EEG analysis

The full MatLab[1] analysis script used for this study can be found in our OSF repository[2] (Van der Wulp et al., 2021). This is the script we received from B&P2017, adapted to our re-analysis. With respect to preprocessing and artifact rejection, we used the same data exclusion criteria

[1] For readers without a MatLab license, we recommend to open and read the script with a free text editor such as Notepad++ (https://notepad-plus-plus.org/downloads/).
[2] https://osf.io/gu7xb/?view_only=f6538baf851c4e58918bf193fed193e2

and methods as described in B&P2017. The code we received had already manually identified the bad channels per participant. These bad channels were interpolated by a VEOG channel. We then used a band-pass filter from 0.1 to 30 Hz and timelocked the data to the onset of each tri-syllabic unit in both conditions. The data was then divided into epochs of 12 tri-syllabic units lasting 10.8 seconds, which overlapped 5/6 of their length. After that, an automatic artifact rejection procedure was used, based on a threshold amplitude adjusted individually per participant (ranging from 200 to 350 μV, average = 210 μV). Again, we adhered to the threshold values previously selected by B&P2017. Stereotypical eye movements were retained, as B&P2017 argued that eye artifacts do not affect the phase-locking of neural oscillations.

B&P2017 computed the phase-locking value per block using the Inter-Trial Coherence (ITC) measure. The ITC ranges from 0 to 1, with 1 being perfect phase-locked neural activity to a given stimulus frequency, and 0 being no phase-locking to this given frequency. The ITC was calculated with a continuous Morlet wavelet transformation using the *newtimef* formula in EEGLAB (Delorme & Makeig, 2004). B&P2017 computed the Morlet wavelet transformations in 0.1 Hz steps, with 1 cycle at the lowest frequency (0.2 Hz), increasing by a factor of 0.5, reaching 45 cycles at the highest frequency (20.2 Hz). The ITC was calculated in both conditions (structured/random) for the syllable frequency of 3.3 Hz ($ITC_{\text{syllable frequency}}$) and the tri-syllabic frequency of 1.1 Hz. ($ITC_{\text{word frequency}}$). Subsequently, B&P2017 calculated the Word Learning Index (WLI, repeated in (5) below) per participant per block in both conditions.

(5) $$\text{WLI} = \frac{ITC_{word\ frequency}}{ITC_{syllable\ frequency}}$$

Since one of the goals of the current study was to re-analyze the data from B&P2017 with Linear Mixed Models (LMM), we aimed to calculate the ITC and WLI values per tri-syllabic unit instead of per block. The number of word presentations could then be modeled as a continuous predictor in the LMM. This would show a similar approach as a subsequent paper from the same authors (Batterink & Paller, 2019). However, we found that it was not possible to calculate the ITC per tri-syllabic unit because the ITC formula requires multiple trials (~100+) to be accurate. Calculating the ITC over a small amount of trials results in a lot of noise. Batterink and Paller (2019) also encountered this problem (p. 62). They bundled the epochs together into 12-epoch groups, covering the course of the presented speech stream as a moving window. However, this resulted in relatively noisy data. Therefore, they smoothed the data "by using a moving average filter with a span of 5 data points (i.e., each *n*th data point was averaged with data points *n*-2, *n*-1, *n*+1, and *n*+2)" (p. 62).

Since B&P2017 had divided each speech stream into three exposure blocks, we adhered to this structure instead. After artifact rejection, we divided the epochs into three groups: block 1, block 2, and block 3. Each block contained +/- 230 epochs for each participant. This is enough to result in reliable ITC values. The ITC was subsequently calculated for each block per participant in each condition in the same way as B&P2017, described above. From these ITC values, we also calculated the WLI per participant per block in each condition, which became our dependent variable for the LMM analysis.

B&P2017 performed all further analyses on six central electrodes: FC1, C1, FCz, Cz, FC2, and C2. They chose these six electrodes because these were the ones “where ITC at the word and syllable frequencies showed the strongest values” (B&P2017, p. 37). This is

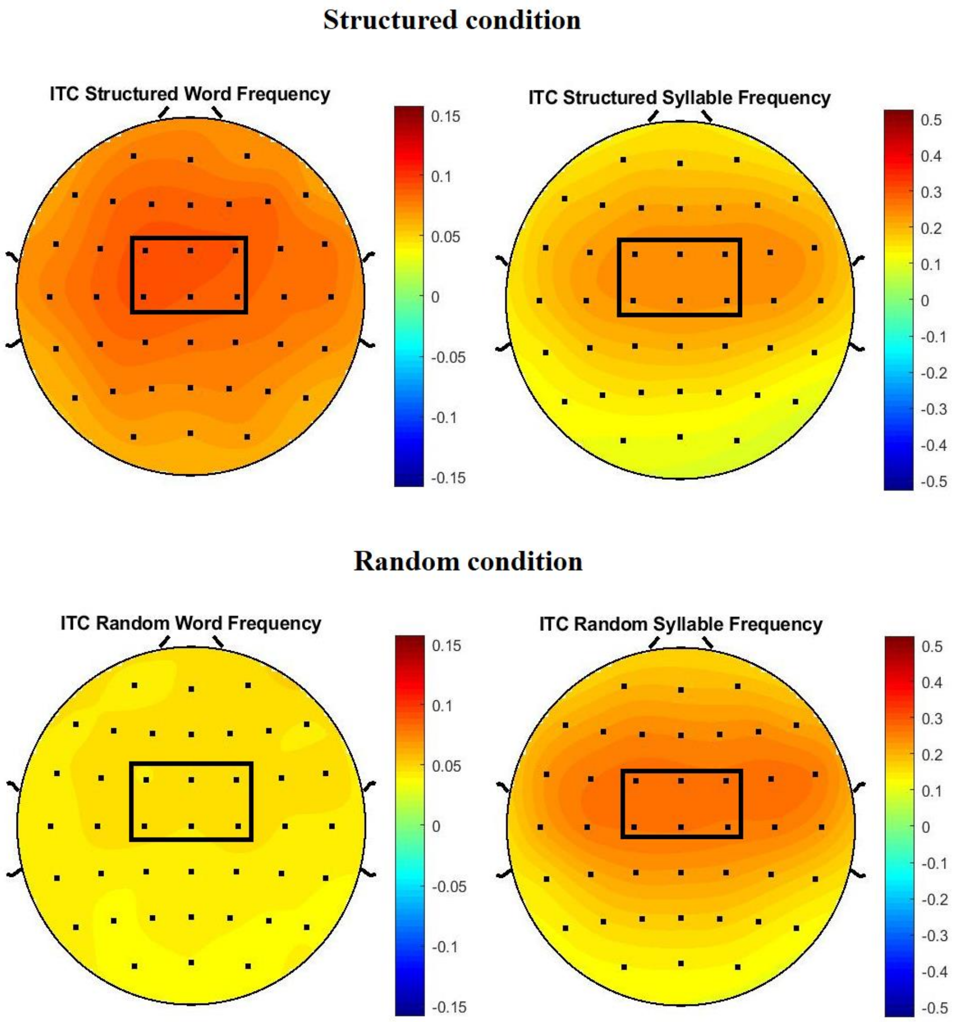


Figure 6, Topographical plots showing the average ITC distribution across the scalp. The black square line surrounds the six electrodes that were used by B&P2017 for further analyses. These six electrodes displayed the highest ITC values in the respective conditions (structured, random) at the respective frequencies (word, syllable). Note that different scales are used for the different frequencies. In the structured condition, there is a higher ITC shown at the word frequency than at the syllable frequency. In the random condition, the reverse is true: a higher ITC is visible at the syllable frequency than at the word frequency.

illustrated in figure 6, which is our replication of figure 4B in B&P2017 (p. 40) and shows the topographic plots of the ITC for the word and syllable frequencies in the two conditions (structured/random). In figure 6, the six electrodes used by B&P2017 for further analyses are marked by a square surrounding them. Figure 6 here and figure 4B in B&P2017 illustrate that these electrodes indeed do show the highest ITC values. To fully replicate B&P2017, we chose to also conduct all further analyses on these six electrodes.

2.3. OCP analysis

B&P2017 pre-defined the syllable presentation order of their speech streams. In total, there were four speech streams made out of two syllable inventories. There was both a structured and a random stream made out of each syllable inventory. Participants would always hear one condition in one syllable inventory, and the other condition in the other syllable inventory. The speech streams were given in numbers, where each syllable sound file corresponded to a unique number. Each stream consisted of 2400 syllable presentations (800 triplets). We used IPython Notebook (Perez & Granger, 2007) to transform these streams of numbers into written versions of the syllables and words presented in B&P2017, and to transform these into the three categories relevant to the OCP-PLACE constraint, based on the place of articulation of the consonant of the syllable. These categories are repeated in (6). All speech streams in these three forms can be found in Appendix A. The full IPython Notebook script[3] can be found in the OSF repository.[4]

(6) labials = /p, b, f, v, w, m/ → P
dorsals = /k, g, x, ŋ/ → K
coronals = /t, d, s, z, ʃ, ʒ, r, l, n/ → T

With regard to quantifying the adherences or violations of the OCP-PLACE constraint in these speech streams, a new variable was created: the OCP variable, illustrated in (7d) for the structured condition in syllable inventory 1 (Appendix A.1.2). The OCP variable refers to the amount of OCP-adherences both within and between tri-syllabic units in the speech streams from B&P2017. For the random streams, we formed triplets starting from the first syllable of the speech streams, so at position 1, 4, 7, etc. Between triplets in both conditions, the first syllable of a triplet with a consonant that has a similar [PLACE] feature as the consonant of the last syllable of the previous triplet is *desired* (e.g. PTK KTP). Within triplets, repetitions of

[3] We included the .ipynb file, as well as an HTML file for readers who do not have Python on their computer.
[4] https://osf.io/gu7xb/?view_only=f6538baf851c4e58918bf193fed193e2

consonants with similar [PLACE] should be *avoided* (e.g. PTT). Example (7a) shows a snippet from words in structured stream 1, (7b) shows the OCP-categorized (as in (6)) versions of these words. The OCP variable ranges from 0 to 3 for each triplet (7d).

(7) a. …tupiro bidaku padoti bidaku golabu tupiro bidaku padoti golabu padoti tupiro…

b.	TPT	PTK	PTT	PTK	KTP	TPT	PTK	PTT	KTP	PTT	TPT
c.	011	011	010	011	111	011	011	010	011	110	111
d.	2	2	1	2	3	2	2	1	2	2	3

The triplet receives 0 or 1 point for the transition between the last triplet and the current triplet (0 points if OCP-PLACE is violated or 1 point for adherence to OCP-PLACE; (7c)), and two points for the two syllable transitions within the word (per syllable transition 0 points if OCP-PLACE is violated or 1 point for adherence to OCP-PLACE; (7c)). The OCP variable then consists of the added scores per triplet (7d).

The IPython Notebook script in the OSF repository[5] also includes the code which calculates the OCP variable for each speech stream. Since the WLI was calculated per block, we also calculated a mean OCP score per block for each speech stream. We found that for the structured streams, the ends of block 1 and 2 and the beginnings of block 2 and 3 were halfway a 'word'. Therefore, we coded the half-words at the beginnings and ends of these blocks as NA values (also for the WLI scores), which were not taken into account for further analyses.

2.4. Statistical analyses

We used RStudio (version R-3.6.3; RStudio Team, 2015) to perform our statistical analyses. The data file was read into R and saved after adaptations with the readxl (Wickham, Bryan, et al., 2019) and writexl (Ooms, 2020) packages. The full Excel data files and R script used for our data analysis can be found in the OSF repository. We calculated descriptive statistics and plots of our data with tidyverse (Wickham, Averick, et al., 2019), ggplot2 (Wickham, 2011) and Plotrix (Lemon, 2006).

For the OCP data, we performed two χ^2 tests to determine if there is a difference in the division of OCP scores of our OCP variable between the structured and random conditions of B&P2017. Participants always received one condition in one syllable inventory and the other condition in the other syllable inventory. Therefore, we performed a total of two χ^2 tests: one for each combination that was presented to the participants (structured1 & random2, or structured2 & random1). We also replicated the ANOVA by B&P2017 to determine if our WLI

[5] https://osf.io/gu7xb/?view_only=f6538baf851c4e58918bf193fed193e2

values corresponded to theirs. We used the package rstatix (Kassambara, 2020) to perform this analysis.

For the LMM analysis, we used the lmerTest package (Kuznetsova et al., 2016). Subsequent pairwise comparisons were computed with the package emmeans (Lenth et al., 2020). We iteratively added predictors and used the −2LL χ^2 test of the model's fit to the data to determine if an added factor improved the model (p < .05). Appendix B contains the summaries of this process. The estimates of the factors reported in the results (chapter 3) belong to the final models. We built two models in total. The first model is a LMM approach to the analysis of B&P2017, which will be referred to as the TP model. The second model included our (centered) OCP variable as a predictor, instead of condition. This model will be referred to as the OCP model. We centered our OCP variable to make it easier to interpret lower and higher OCP scores which became scores under and above 0, respectively.

3. Results

3.1. EEG results

We calculated the Inter-Trial Coherence (ITC) and the Word Learning Index (WLI) of the EEG data from B&P2017. Figure 7 is our replication of figure 4A in B&P2017 (p. 40). Figure 7 displays the ITC as a function of frequency per condition and block. In the structured condition, there is an increasing ITC per block to the word frequency (1.1 Hz) and a decreasing ITC to the

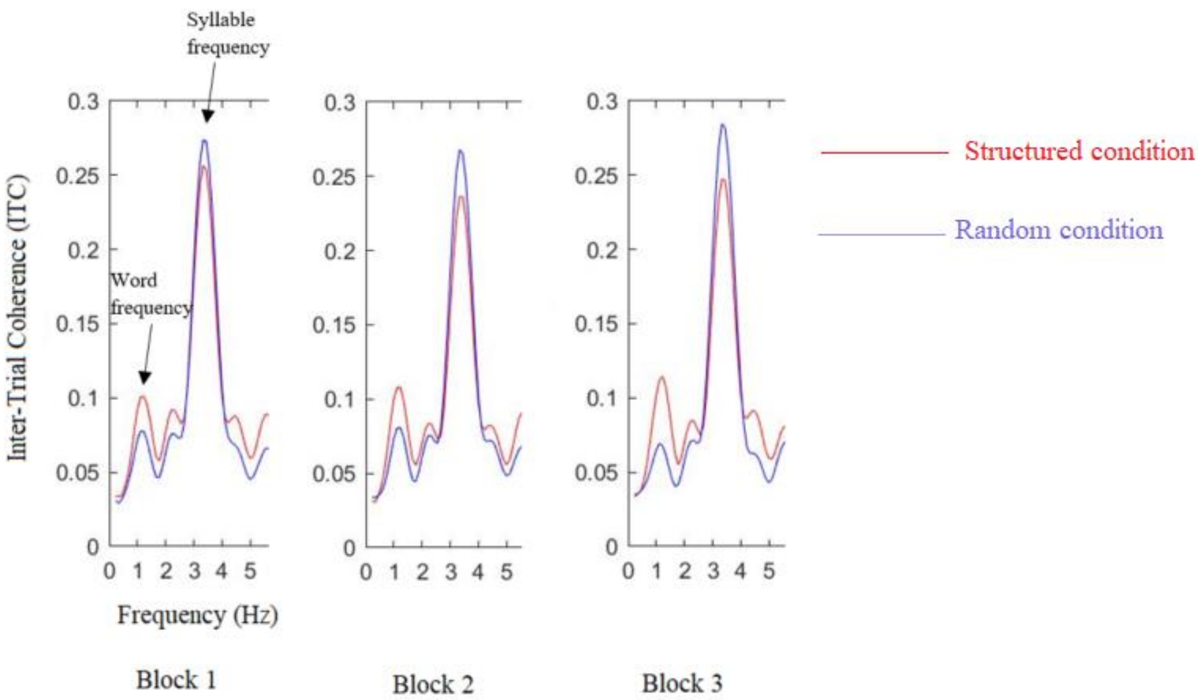


Figure 7. Inter-Trial Coherence as a function of condition and frequency per block. This image is our replication of Figure 4A in B&P2017 (p. 40).

syllable frequency (3.3 Hz). In the random condition, there is an increasing ITC per block to the syllable frequency, but no increasing ITC to the word frequency. This result replicates the findings by B&P2017. The full graphs with ITC values for all frequencies (0.1 – 20 Hz) per block can be found in Appendix C.

The ITC scores for the word and syllable frequencies were used to calculate the WLI. With this WLI as the dependent variable, we replicated the ANOVA of B&P2017 before we performed our LMM analysis. Surprisingly, our ANOVA initially did not yield the same results. B&P2017 reported a general effect of condition across blocks (F (1, 44) = 17.3, p < .001; p. 39). Our replication also revealed a significant effect of condition, but with a different F-value (F (1, 44) = 14.6, p < .001). Moreover, B&P2017 reported an interaction of condition and block, indicating that in the structured condition the WLI rises as exposure progresses, while this does not happen in the random condition (F (2, 88) = 3.72, p = .029; p. 39). However, our replication of this interaction did approach but not reach significance (F (2, 88) = 2.99, p = 0.056). As to the source of these differences in results, we found that we had generated different WLIs than B&P2017 due to our method of dividing the EEG data into the three blocks. Since our initial plan to generate the WLI per triplet changed during the analysis (as described in paragraph 2.2), we divided the data into the blocks *after* preprocessing and artifact rejection by dividing the remaining data into three equal parts. B&P2017 divided the data manually *before* processing the data, as the analysis script did not include code on dividing the data into the blocks. Therefore, we corresponded with Dr. Batterink again, who then sent us their SPSS (IBM Corp., 2017) data file, including their ITC and WLI values. When we replicated the ANOVA with their WLI values, the results were identical to the results reported in B&P2017. The descriptive statistics of the WLI calculated for this re-analysis and the WLI we received from B&P2017 can be seen in table 1 below, as well as the differences between the scores.

Table 1.
Descriptive statistics of our WLI and the WLI calculated by B&P2017 (N = 45 per condition).

	Structured				**Random**			
			Range				**Range**	
	Mean	***SD***	**Lower**	**Upper**	**Mean**	***SD***	**Lower**	**Upper**
WLI current analysis	0.60	0.56	0.12	4.76	0.42	0.47	0.07	3.56
WLI B&P2017	0.58	0.51	0.11	4.08	0.39	0.40	0.07	2.96
Difference	0.02	0.05	0.01	0.68	0.03	0.07	0	0.60

The differences between both WLIs were quite small. To determine this statistically, we performed a Pearson's correlation analysis on our WLI and B&P2017's WLI, which was highly positive and significant (r (268) = 0.96, p < .001). This correlation is also shown in figure 8 below. Nonetheless, we aimed for a replication of B&P2017 and therefore decided to use the WLI scores from B&P2017 for all further analyses.

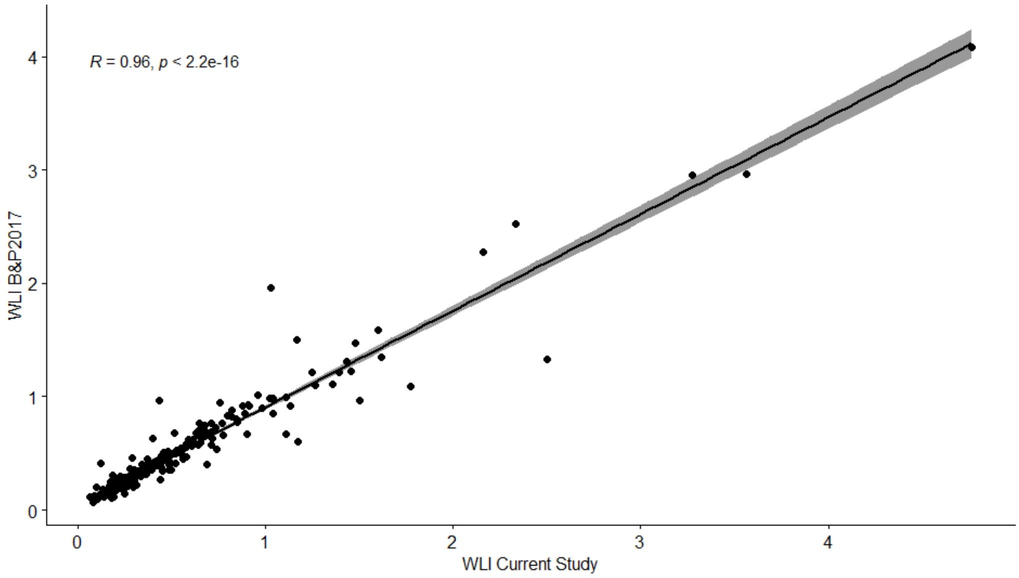


Figure 8. Correlation between the WLI of B&P2017 and the WLI calculated in the current study. The correlation is positive and highly significant (r = 0.96, p < .001).

3.2. OCP-PLACE results

Our OCP variable quantified the adherence of each speech stream used by B&P2017 to the OCP-PLACE constraint. Each word or triplet received between 0 and 3 points. Higher scores indicate more adherence to the OCP-PLACE constraint. Table 2 shows the mean OCP scores and standard deviations for each speech stream. Table 2 illustrates that the structured speech streams show a higher adherence to the OCP-PLACE constraint than the random speech streams. Nonetheless, there is also some degree of adherence to OCP-PLACE in the random speech streams. If there would be no adherence to OCP-PLACE, the OCP score would be (close to) zero.

Table 2.
Mean OCP scores and standard deviations (SD) for each speech stream of B&P2017 (N = 800 per stream).

Stream	Mean OCP-score	*SD*
Structured (syll inventory 1)	2.09	0.65
Structured (syll inventory 2)	2.07	0.65
Random (syll inventory 1)	1.65	0.84
Random (syll inventory 2)	1.69	0.79

We performed two χ^2 tests; one for each list presented to the participants (i.e. syllable inventory 1 for the structured stream and syllable inventory 2 for the random stream), to determine if these differences in the division of the OCP scores between the structured and random conditions were significant. Tables 3 and 4 show the observed values, standardized residuals[6], and *p*-values for the different levels of the OCP score per list. For both lists, there was a significant difference overall between the distribution of the OCP scores over the conditions (table 3: χ^2 (3, N = 1596) = 148.27, p < .001; table 4: χ^2 (3, N = 1596) = 117.16, p < .001). Both structured streams do not contain words which completely violate OCP-PLACE and thus receive 0 points on our OCP variable, whereas this does happen for the random streams (see tables 3 and 4). The structured streams also show less instances of words receiving 1 point, and more instances of words receiving 2 or 3 points than the random streams. Table 3 illustrates that there are significant differences in the distribution of the OCP scores for all categories in that list. Table 4 demonstrates that the other list yields significant differences between the conditions for the categories 0, 1, and 3, but not for the category 2 of the OCP variable, which approaches significance.

Table 3.
Division of OCP scores, standardized residuals, and p-values for the χ2 test over structured1 and random2.

	Structured (syll inventory 1)		Random (syll inventory 2)		
OCP score	**Count**	**St. residual**	**Count**	**St. residual**	***p*-value**
0	0	-5.83	68	5.83	< .001
1	134	-4.61	264	4.61	< .001
2	456	2.83	343	-2.83	.005
3	208	3.30	123	-3.30	< .001

Table 4.
Division of OCP scores, standardized residuals, and p-values for the χ2 test over structured2 and random1.

	Structured (syll inventory 2)		Random (syll inventory 1)		
OCP score	**Count**	**St. residual**	**Count**	**St. residual**	***p*-value**
0	0	-5.20	54	5.20	< .001
1	139	-3.95	249	3.95	< .001
2	461	1.85	385	-1.85	.064
3	198	3.55	110	-3.55	< .001

[6] Standardized residual = *z*-score = (observed – model) / √ model

3.3. Linear Mixed Models

For our LMM analysis, we iteratively added predictors and used the −2LL χ^2 test of the model's fit to the data to determine if a factor improved the model (Appendix B). We built two models in total. The first model is a LMM approach to the analysis of B&P2017, which will be referred to as the TP model. The second model included our centered OCP variable as a predictor, instead of condition, which will be referred to as the OCP model. A direct comparison between the TP model and the OCP model, including the estimates of all main effects and interactions can be found in Appendix D.

3.3.1. The TP model

The TP model[7] included a random intercept for both participant and word (triplet-item) and the WLI as the dependent variable. Following B&P2017, we added a fixed factor of condition first, followed by a fixed factor of block. We then added an interaction of condition and block, following B&P2017 as well. This interaction improved the model significantly, compared to a model with only the individual fixed factors ($p < .001$; Appendix B). We also added the order of condition presentation (structured first or random first) as a fixed factor, but this did not improve the model ($p = .96$; Appendix B). Therefore, our final model included main effects and an interaction between condition and block, and random intercepts for participant and word.

Our final TP model indicated a significant main effect of condition, where the structured condition yielded higher WLI scores than the random condition ($b = 0.10$, t (291.40) = 15.79, $p < .001$, 95% CI [0.09, 0.11]). Moreover, the interaction between condition and block was significant. The WLI rises in block 2 and 3 for the structured condition, but not for the random condition (structured*block 2: $b = 0.07$, t (56150) = 11.59, $p < .001$, 95% CI [0.06, 0.08]; structured*block 3: $b = 0.21$, t (57150) = 35.57, $p < .001$, 95% CI [0.20, 0.22]).

Table 5.
*Pairwise Comparisons Condition*Block in the TP model[a]*

(I) reference Random Block 1	Mean difference (I-J)	Std. Error	df	*t*	Sig.[b]
Random 1 – Structured 1	-0.10	0.006	291	-15.79	< .001
Random 2 – Structured 2	-0.17	0.006	292	-26.64	< .001
Random 3 – Structured 3	-0.31	0.006	292	-49.05	< .001

a. Dependent Variable: WLI (B&P2017)
b. Adjustment for multiple comparisons: Tukey

[7] The formula of the model: WLI_B ~ Condition*Block + (1|Participant)+ (1|Word)

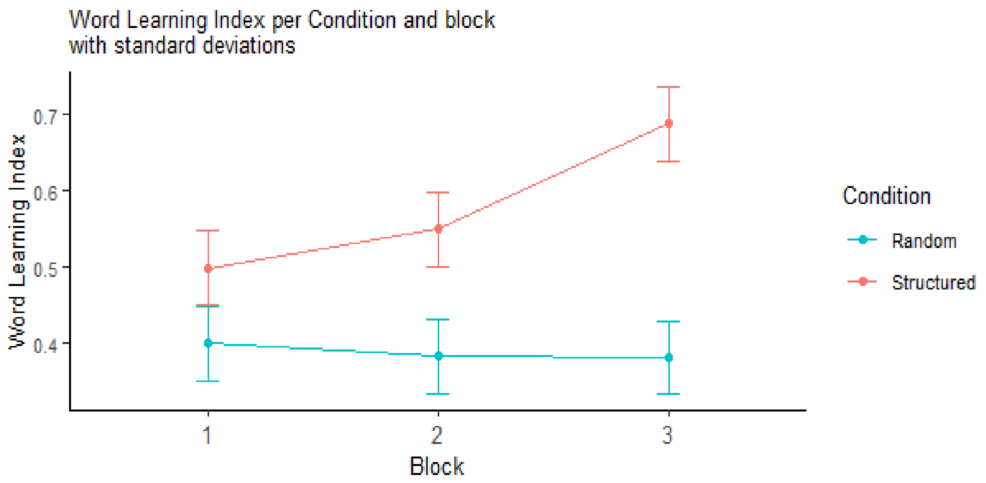


Figure 9. WLI from the TP model as a function of block (SDs as error bars). The WLI in the structured condition (red line) rises as a function of block. This does not happen in the random condition (blue line). This figure is our replication of figure 4C in B&P2017 (p. 40), also shown as figure 5 above.

Table 5 displays the pairwise comparisons for the conditions per block, illustrating that the difference between the conditions grows as exposure progresses. The full table with all pairwise comparisons can be found in Appendix E. All pairwise comparisons shown in in Appendix E are significant, except for the comparison between random block 2 and random block 3. This illustrates that the WLI in the random condition does not rise as a function of block, while this is the case in the structured condition. Figure 9 illustrates this result as well. This means that participants show increasingly more phase-locking to the word-frequency than to the syllable frequency in the structured condition. Thus, participants acquired triplet word-units in the structured condition, but not in the random condition. This result replicates the findings by B&P2017.

3.3.2. The OCP model

The OCP model[8] included the same random intercepts for participant and word (triplet) as the TP model, as well as the WLI being the dependent variable. We included our centered OCP variable as a fixed factor, as well as block. Thus, the OCP variable takes the place of condition, in comparison to the TP model. We then also added an interaction of our OCP variable and block, which improved the model significantly, compared to a model with just the individual fixed factors ($p < .001$; Appendix B). Again, we added the order of condition presentation (structured first or random first) as a fixed factor which did not significantly improve this model either ($p = 0.95$; Appendix B) and was therefore left out of the model. The final model thus

[8] The formula of the model: WLI_B ~ Centered_OCP*Block + (1|Participant)+ (1|Word)

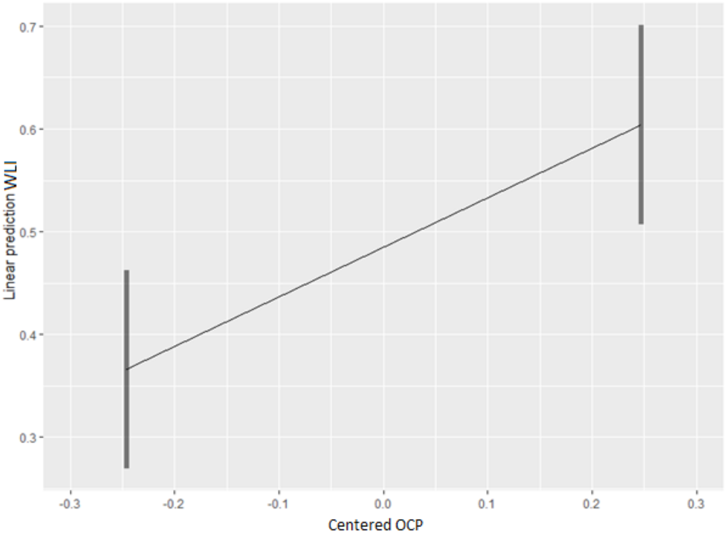
included main effects and an interaction of OCP and block, with random intercepts for participant and word.

Our final OCP model revealed a significant main effect of OCP ($b = 0.23$, t (304.10) = 18.71, $p < .001$, 95% CI [0.20, 0.25]), increasing the WLI as the OCP variable increases. The model also included a significant interaction of OCP and block (OCP*block2: $b = 0.17$, t (59190) = 12.48, $p < .001$, 95% CI [0.14, 0.20]; OCP*block3: $b = 0.60$, t (50190) = 40.98, $p <$ .001, 95% CI [0.57, 0.62]). In contrast to the condition variable used in the TP model, which is nominal with two levels (structured, random), the OCP variable is a continuous ratio variable ranging from 0 to 3 points. Therefore, this interaction should be interpreted differently than the condition and block interaction in the TP model. Table 6 displays the estimates of the centered OCP and block interaction trend. In block 2, the mean increase in the WLI for an increase of one point on the OCP variable is 0.40, while in block 3 this increase in the WLI for a one-point increase of OCP is 0.82. Thus, the increase in WLI for a one-point increase in the OCP variable becomes larger over time.

Table 6.
*Estimates of Fixed Factors Centered OCP*Block in the OCP model[a]*

Parameter				95% Confidence Interval	
Block	**Centered OCP trend**	**Std. Error**	**df**	**Lower Bound**	**Upper Bound**
1	0.23	0.01	304	0.20	0.25
2	0.40	0.01	302	0.37	0.42
3	0.82	0.01	303	0.80	0.85

a. Dependent Variable: WLI (B&P2017)



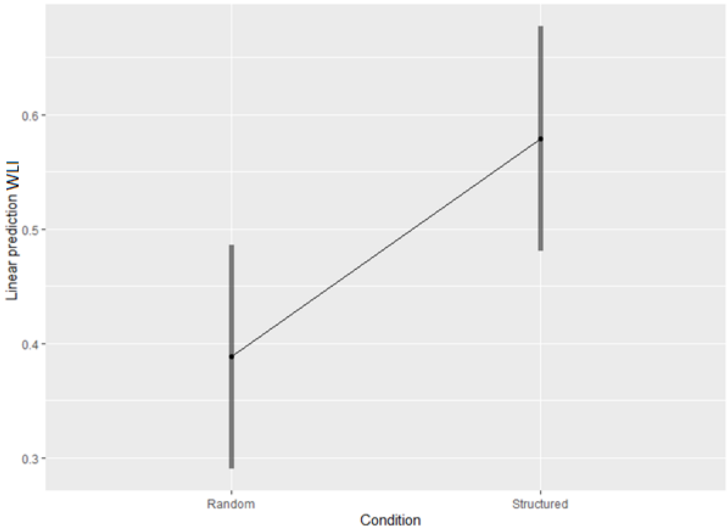


Figure 10. Linear predictions for the WLI in the OCP model (left) and the TP model (right), showing similar linear patterns. The WLI in the structured condition is higher than the WLI in the random condition. At the same time, the WLI rises as the OCP score rises. The OCP score in the structured condition is higher than in the random condition.

Figure 10 demonstrates a comparison of the linear predictions for the WLI in the TP model and the OCP model. It illustrates a similar effect: lower OCP scores elicit lower WLIs, while higher OCP scores elicit higher WLIs. This is similar to the condition variable in the TP model, where the structured condition elicits higher WLIs than the random condition. This similarity is likely caused by the fact that the OCP in the structured condition is always higher than in the random condition (section 3.2).

4. Discussion and conclusion

The current study aimed to re-analyze the EEG data initially collected, analyzed and reported by B&P2017, answering two additional research questions.

1. Can the results by B&P2017 be explained by the phonotactic OCP-PLACE constraint? In other words, can it be the listener's sensitivity to the OCP-PLACE constraint that cues word boundaries instead of sensitivity to TPs?

We calculated a new OCP variable based on the speech streams used by B&P2017, yielding a score of OCP-adherence ranging from 0 to 3 per triplet in each speech stream. We conclude that adherence to the OCP-PLACE constraint for a triplet segmentation is present in all speech streams and both conditions. Structured conditions yielded higher OCP scores than random conditions, in parallel with the TP structure of the conditions. Our LMM analysis then provided us with both a TP model and an OCP model. Both models indicated a significant interaction of either condition and block or OCP and block. In the TP model, the WLI increases as the number of items perceived is increasing when a participant is exposed to the structured condition, while this does not happen in the random condition (figure 9). In the OCP model, increments of the WLI as a function of OCP score increase as exposure progresses (irrespective of TP condition; table 6). Thus, we could say that the OCP variable could be a substitute for the TP-based condition variable by B&P2017 (figure 10). The difference in the WLI increases between the structured and random condition over the blocks, and in parallel a higher OCP score exerts a larger influence on the WLI as exposure progresses. An important difference between the structured and random streams is that the structured streams yield significantly higher OCP scores. Therefore, the OCP variable could explain the data as well, similar to the TP conditions variable.

Why the OCP variable is higher in the structured than the random condition can be explained by the fact that the structured speech streams contained four repeating words, while the random streams contained randomly concatenated syllables. OCP-PLACE adherence and thus scores on our OCP variable in the structured condition for these four words would therefore repeat. Moreover, there were no words in the two structured streams that yielded an OCP score of 0. On the other hand, in the random streams the same triplet rarely occurred (more than) twice and there were multiple triplets yielding an OCP score of 0 in both random streams. Our $\chi 2$ tests showed that this difference between the structured and random conditions with respect to OCP-adherence is significant. Thus, we cannot disentangle the OCP effect from the effect of condition, because there is always a higher OCP score in the structured than the random conditions.

Therefore, further experimental research is needed where OCP is held constant in both speech streams, or where OCP is manipulated explicitly while keeping the TP structure under control, to further investigate the OCP's effects on word segmentation. Moreover, as mentioned in 1.2, research similar to Boll-Avetisyan and Kager (2014), experimentally testing the influence of OCP-PLACE on word segmentation *in English* (and other languages) has not yet been performed. Their study found an effect of OCP-PLACE as a cue for word segmentation in Dutch, and based on that result we assumed that OCP-PLACE could perhaps explain the finding by B&P2017 that some participants unexpectedly segmented the random speech streams into triplets as well. Since we found adherences to the OCP-PLACE constraint in the random conditions and since we found a significant effect of OCP in our LMM analysis, this could indeed be the case. However, OCP-adherence was higher in the structured condition than in the random condition, which makes it impossible to fully disentangle the two effects in this re-analysis without further experimental investigation.

2. Can we reproduce the results found by B&P2017, by re-analyzing their data with Linear Mixed Models instead of ANOVA?

Our second research question can be answered more easily, confirming the robustness of B&P2017's previously reported effects. Our LMM replication of their ANOVA yielded the same significant main effect of condition, as well as an interaction of condition and block. This indicates that the WLI increases over time in the structured condition but not in the random condition (figure 9). This means that participants showed increasingly more phase-locking to the word frequency than to the syllable frequency in the structured condition. Thus, participants acquired triplet word-units in the structured condition, but not (as much) in the random condition. This result replicates the findings by B&P2017 and provides a confirming answer to our second research question.

A limitation of our LMM re-analysis is that we were unable to model the exposure time as a continuous variable, by calculating the WLI per presented triplet instead of per block. Follow-up research could aim for such a continuous analysis by employing a moving window of epochs containing ~100 trials, overlapping 99/100 of their length. In this way the problem of noisy data could be avoided, while creating the possibility to calculate the WLI per triplet-item. Moreover, the segmentation of the data into the three blocks per condition appeared to have a significant effect on the interaction between condition and block. Because we calculated the blocks *after* preprocessing and artifact rejection (whereas B&P2017 did so *before* cleaning the data), we found different results while replicating the ANOVA of B&P2017, losing the significance for the condition and block interaction. This kind of differing results could also be avoided if the WLI is calculated per item, making these kind of block calculations redundant.

Finally, B&P2017 and the current re-analysis performed all significance testing on six central electrodes (FC1, C1, FCz, Cz, FC2, and C2) because "ITC at the word and syllable frequencies showed the strongest values [in these electrode locations]" (B&P2017, p. 37). Further research should also perform these analyses with a 64-electrode average, to see if the result is still robust when not only the electrodes with the strongest ITC values are taken into consideration. More research on the localization of statistical learning and language processing in general (performed using methods with a good spatial resolution such as fMRI or MEG) is instrumental to form expectations about the localization of the ITC in future studies.

In conclusion, this re-analysis confirmed the statistical robustness of the results found by B&P2017, re-analyzing their data with a LMM approach instead of employing an ANOVA, and therefore yielding a lower risk of a type I error. Furthermore, the OCP-PLACE constraint provided an alternative explanation of B&P2017's data and could explain their unexpected finding that some participants segmented the random speech streams into triplets as well. We found a significant effect of OCP that is parallel to the effect of condition in the data of B&P2017. Further research should investigate the independent effects of OCP-PLACE on word segmentation in English. The OCP-PLACE constraint must also be considered as a possible confounder that should be controlled for in further statistical language learning experiments.

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491–496. https://doi.org/10.1038/nn.3648

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Batterink, L. J. (2020). Syllables in Sync Form a Link: Neural Phase-locking Reflects Word Knowledge during Language Learning. *Journal of Cognitive Neuroscience*, *32*(9), 1735–1748. https://doi.org/10.1162/jocn_a_01581

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., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247–260. https://doi.org/10.1016/S0163-6383(81)80027-6

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews*, *68*, 1004–1005. https://doi.org/10.1016/j.neubiorev.2016.05.034

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

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal Infants Discover Statistical Word Patterns at Similar Rates as Adults: Evidence From Neural Entrainment. *Psychological Science*, *31*(9), 1161–1173. https://doi.org/10.1177/0956797620933237

Coetzee, A. W. (2010). Grammar is both categorical and gradient. In S. Parker (Ed.), *Phonological Argumentation: Essays on Evidence and Motivation.* (pp. 9–42). Equinox Pub. Ltd. https://doi.org/doi:10.7282/T3D50JZQ

Dmitrieva, O., & Anttila, A. (2008). The gradient phonotactics of English CVC syllables. *Poster Presented at the 11th Laboratory Phonology Conference, Wellington, New Zealand*.

Frisch, S. (1996). *Similarity and frequency in phonology* [PhD Thesis].

Goddijn, S., & Binnenpoorte, D. (2003). Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus. *Proceedings of ICPhS*, 1361–1364.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4

Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25–36. https://doi.org/10.1016/j.bandl.2015.03.005

IBM Corp. (2017). *IBM SPSS Statistics for Windows* (Version 2017) [Computer software].

Kassambara, A. (2020). Rstatix: Pipe-friendly framework for basic statistical tests. *R Package Version 0.6. 0*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in linear mixed effects models [Computer software]*.

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *WIREs Cognitive Science*, *11*(5), e1528. https://doi.org/10.1002/wcs.1528

Lemon, J. (2006). *Plotrix: A package in the red light district of R. R-News 6: 8–12*.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated marginal means. R package version 1.4. 4*.

MATLAB (9.6.0.1072779 (R2019a)). (2019). [Computer software]. The MathWorks Inc. https://www.mathworks.com/products/matlab.html

Monaghan, P., & Zuidema, W. (2015). General purpose cognitive processing constraints and phonotactic properties of the vocabulary. *Workshop on the Evolution and Phonetic Capabilities: Causes, Constraints and Consequences. Retrieved From https://pure.mpg.de/rest/items/item_2351030/component/file_2351029/content#page=24*

Nacy, S. M., Kbah, S. N., Jafer, H. A., & Al-Shaalan, I. (2016). Controlling a servo motor using EEG signals from the primary motor cortex. *American Journal of Biomedical Engineering*, *6*(5), 139–146.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

Ooms, J. (2020). writexl: Export Data Frames to Excel 'xlsx'Format. 2020. *R Package Version*, *1*.

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

Perez, F., & Granger, B. E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, *9*(3), 21–29. https://doi.org/10.1109/MCSE.2007.53

RStudio Team. (2015). *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, Inc.

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

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

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.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113. https://doi.org/10.1016/j.tics.2008.01.002

Shatzman, K., & Kager, R. (2007). A role for phonotactic constraints in speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1409–1412.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Van der Wulp, I., Wijnen, F., & Struiksma, M. E. (2021, March 1). Word segmentation: TP or OCP? A re-analysis of Batterink & Paller (2017). Retrieved from https://osf.io/gu7xb/?view_only=f6538baf851c4e58918bf193fed193e2

Wickham, H. (2011). Ggplot2. *WIREs Computational Statistics*, *3*(2), 180–185. https://doi.org/10.1002/wics.147

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wickham, H., Bryan, J., Kalicinski, M., Valery, K., Leitienne, C., Colbert, B., Hoerl, D., Miller, E., & Bryan, M. J. (2019). *Package 'readxl.'*

Appendix

Appendix A. Speech streams and OCP transcriptions

The appendix below displays the output of the Python code used to calculate the OCP variable for the speech streams of B&P2017. The speech streams were numbers in a .txt file. Python has proven to be a powerful tool to transform these numbers into written versions of the syllables and words presented to the participants by B&P2017, and to calculate our OCP variable.

The Python Notebook file RMAThesis_Appendix.ipynb and HTML file RMAThesis_Appendix.html can be found in the OSF repository: https://osf.io/gu7xb/?view_only=f6538baf851c4e58918bf193fed193e2

A.1. Structured Stream 1

A.1.1. Structured Stream 1 as provided by B&P2017

12 9 10 1 3 6 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 1 3 6 5 7 2 8 4 11 5 7 2 1 3 6 8 4 11 5 7 2 8 4 11
1 3 6 5 7 2 12 9 10 5 7 2 12 9 10 5 7 2 12 9 10 5 7 2 1 3 6 12 9 10 1 3 6 5 7 2 1 3 6 12 9 10 5 7
2 8 4 11 12 9 10 5 7 2 1 3 6 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 12 9 10 1 3 6 8 4 11 12 9 10 5 7 2
12 9 10 1 3 6 8 4 11 1 3 6 5 7 2 1 3 6 12 9 10 5 7 2 1 3 6 12 9 10 5 7 2 8 4 11 12 9 10 8 4 11 5
7 2 1 3 6 8 4 11 12 9 10 8 4 11 1 3 6 12 9 10 5 7 2 1 3 6 5 7 2 1 3 6 8 4 11 1 3 6 12 9 10 5 7 2 1
3 6 5 7 2 1 3 6 8 4 11 12 9 10 8 4 11 1 3 6 8 4 11 12 9 10 1 3 6 12 9 10 5 7 2 1 3 6 12 9 10 8 4
11 12 9 10 1 3 6 8 4 11 12 9 10 5 7 2 8 4 11 12 9 10 5 7 2 8 4 11 12 9 10 5 7 2 12 9 10 1 3 6 5
7 2 8 4 11 12 9 10 8 4 11 5 7 2 8 4 11 1 3 6 8 4 11 1 3 6 12 9 10 5 7 2 8 4 11 1 3 6 12 9 10 8 4
11 5 7 2 8 4 11 5 7 2 12 9 10 1 3 6 8 4 11 1 3 6 8 4 11 5 7 2 8 4 11 1 3 6 8 4 11 5 7 2 8 4 11 12
9 10 1 3 6 8 4 11 12 9 10 8 4 11 12 9 10 5 7 2 1 3 6 8 4 11 5 7 2 1 3 6 8 4 11 12 9 10 1 3 6 12 9
10 1 3 6 12 9 10 8 4 11 12 9 10 5 7 2 12 9 10 1 3 6 5 7 2 1 3 6 8 4 11 1 3 6 5 7 2 1 3 6 12 9 10
8 4 11 12 9 10 8 4 11 12 9 10 8 4 11 12 9 10 1 3 6 5 7 2 8 4 11 5 7 2 8 4 11 5 7 2 8 4 11 12 9
10 5 7 2 12 9 10 5 7 2 12 9 10 1 3 6 5 7 2 1 3 6 5 7 2 8 4 11 5 7 2 8 4 11 12 9 10 8 4 11 12 9 10
5 7 2 12 9 10 8 4 11 1 3 6 8 4 11 1 3 6 5 7 2 8 4 11 5 7 2 1 3 6 8 4 11 1 3 6 8 4 11 5 7 2 1 3 6 8
4 11 1 3 6 8 4 11 5 7 2 1 3 6 8 4 11 5 7 2 12 9 10 8 4 11 12 9 10 8 4 11 1 3 6 5 7 2 8 4 11 1 3 6
12 9 10 8 4 11 5 7 2 8 4 11 5 7 2 1 3 6 8 4 11 5 7 2 12 9 10 5 7 2 12 9 10 8 4 11 1 3 6 12 9 10 1
3 6 8 4 11 12 9 10 5 7 2 8 4 11 5 7 2 8 4 11 1 3 6 12 9 10 8 4 11 1 3 6 5 7 2 8 4 11 1 3 6 5 7 2
12 9 10 8 4 11 12 9 10 8 4 11 12 9 10 5 7 2 1 3 6 8 4 11 1 3 6 5 7 2 1 3 6 5 7 2 8 4 11 12 9 10 5
7 2 1 3 6 5 7 2 12 9 10 5 7 2 8 4 11 5 7 2 12 9 10 5 7 2 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 5 7 2
12 9 10 5 7 2 1 3 6 5 7 2 12 9 10 8 4 11 12 9 10 5 7 2 8 4 11 1 3 6 5 7 2 12 9 10 1 3 6 12 9 10 5
7 2 12 9 10 5 7 2 12 9 10 1 3 6 8 4 11 1 3 6 5 7 2 1 3 6 5 7 2 8 4 11 12 9 10 5 7 2 12 9 10 5 7 2
1 3 6 5 7 2 1 3 6 5 7 2 8 4 11 5 7 2 8 4 11 12 9 10 8 4 11 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 12 9
10 5 7 2 1 3 6 8 4 11 5 7 2 1 3 6 12 9 10 8 4 11 5 7 2 12 9 10 8 4 11 1 3 6 8 4 11 5 7 2 12 9 10
8 4 11 1 3 6 5 7 2 8 4 11 12 9 10 1 3 6 8 4 11 5 7 2 1 3 6 8 4 11 5 7 2 1 3 6 8 4 11 5 7 2 1 3 6
12 9 10 1 3 6 8 4 11 1 3 6 5 7 2 8 4 11 1 3 6 8 4 11 1 3 6 8 4 11 5 7 2 8 4 11 12 9 10 8 4 11 12
9 10 5 7 2 1 3 6 5 7 2 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 12 9 10 1 3 6 5 7 2 12 9 10 8 4 11 5 7 2 8
4 11 1 3 6 8 4 11 12 9 10 5 7 2 12 9 10 8 4 11 1 3 6 5 7 2 8 4 11 5 7 2 12 9 10 1 3 6 12 9 10 8 4
11 5 7 2 1 3 6 5 7 2 8 4 11 5 7 2 1 3 6 12 9 10 5 7 2 12 9 10 8 4 11 12 9 10 8 4 11 12 9 10 1 3 6
8 4 11 1 3 6 8 4 11 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 5 7 2 1 3 6 12 9 10 8 4 11 12 9 10 5 7 2 12
9 10 1 3 6 12 9 10 8 4 11 1 3 6 5 7 2 12 9 10 5 7 2 8 4 11 12 9 10 1 3 6 12 9 10 8 4 11 5 7 2 1 3
6 5 7 2 1 3 6 8 4 11 5 7 2 12 9 10 5 7 2 12 9 10 1 3 6 5 7 2 12 9 10 8 4 11 12 9 10 1 3 6 12 9 10
1 3 6 8 4 11 1 3 6 8 4 11 12 9 10 5 7 2 12 9 10 5 7 2 12 9 10 1 3 6 12 9 10 8 4 11 12 9 10 1 3 6
8 4 11 12 9 10 8 4 11 12 9 10 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 5 7 2 1 3 6 8 4 11 1 3 6 5 7 2 1

3 6 5 7 2 8 4 11 1 3 6 12 9 10 1 3 6 8 4 11 12 9 10 1 3 6 5 7 2 12 9 10 8 4 11 1 3 6 5 7 2 12 9 10 1 3 6 5 7 2 8 4 11 1 3 6 5 7 2 1 3 6 12 9 10 1 3 6 12 9 10 5 7 2 1 3 6 8 4 11 5 7 2 8 4 11 5 7 2 12 9 10 5 7 2 1 3 6 12 9 10 8 4 11 5 7 2 1 3 6 12 9 10 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 1 3 6 8 4 11 12 9 10 8 4 11 12 9 10 1 3 6 5 7 2 12 9 10 5 7 2 12 9 10 5 7 2 8 4 11 12 9 10 1 3 6 8 4 11 12 9 10 5 7 2 1 3 6 12 9 10 1 3 6 5 7 2 12 9 10 8 4 11 1 3 6 12 9 10 1 3 6 8 4 11 1 3 6 5 7 2 12 9 10 1 3 6 12 9 10 8 4 11 1 3 6 12 9 10 5 7 2 12 9 10 8 4 11 12 9 10 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 8 4 11 1 3 6 12 9 10 1 3 6 12 9 10 5 7 2 8 4 11 12 9 10 5 7 2 1 3 6 8 4 11 5 7 2 8 4 11 1 3 6 12 9 10 8 4 11 1 3 6 12 9 10 5 7 2 1 3 6 12 9 10 1 3 6 12 9 10 5 7 2 8 4 11 5 7 2 12 9 10 8 4 11 1 3 6 8 4 11 5 7 2 12 9 10 8 4 11 1 3 6 12 9 10 5 7 2 8 4 11 1 3 6 8 4 11 12 9 10 8 4 11 5 7 2 1 3 6 8 4 11 1 3 6 5 7 2 8 4 11 5 7 2 1 3 6 5 7 2 8 4 11 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 8 4 11 5 7 2 1 3 6 8 4 11 1 3 6 8 4 11 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 1 3 6 8 4 11 12 9 10 5 7 2 8 4 11 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 8 4 11 5 7 2 1 3 6 8 4 11 1 3 6 8 4 11 5 7 2 8 4 11 1 3 6 8 4 11 1 3 6 5 7 2 12 9 10 1 3 6 12 9 10 1 3 6 8 4 11 5 7 2 12 9 10 1 3 6 12 9 10 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 12 9 10 8 4 11 1 3 6 12 9 10 1 3 6 12 9 10 8 4 11 1 3 6 12 9 10 5 7 2 8 4 11 5 7 2 8 4 11 5 7 2 1 3 6 12 9 10 8 4 11 1 3 6 5 7 2 8 4 11 12 9 10 8 4 11 1 3 6 12 9 10 1 3 6 12 9 10 8 4 11 12 9 10 5 7 2 8 4 11 5 7 2 12 9 10 1 3 6 8 4 11 12 9 10 1 3 6 5 7 2 1 3 6 8 4 11 5 7 2 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 1 3 6 5 7 2 12 9 10 8 4 11 1 3 6 12 9 10 8 4 11 1 3 6 12 9 10 8 4 11 1 3 6 8 4 11 5 7 2 12 9 10 5 7 2 12 9 10 5 7 2 12 9 10 1 3 6 12 9 10 5 7 2 1 3 6 8 4 11 1 3 6 12 9 10 1 3 6 5 7 2 12 9 10 1 3 6 5 7 2 1 3 6 5 7 2 8 4 11 12 9 10 1 3 6 8 4 11 12 9 10 1 3 6 12 9 10 8 4 11 5 7 2 8 4 11 5 7 2 8 4 11 12 9 10 1 3 6 8 4 11 12 9 10 5 7 2 8 4 11 12 9 10 5 7 2 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 1 3 6 12 9 10 8 4 11 5 7 2 8 4 11 12 9 10 5 7 2 1 3 6 5 7 2 1 3 6 5 7 2 1 3 6 12 9 10 8 4 11 5 7 2 12 9 10 5 7 2 8 4 11 5 7 2 1 3 6 12 9 10 5 7 2 1 3 6 8 4 11 12 9 10 8 4 11 5 7 2 12 9 10 1 3 6 5 7 2 5 7 2

A.1.2. Structured Stream 1 in words

tupiro bidaku golabu tupiro bidaku tupiro golabu bidaku golabu padoti golabu bidaku padoti golabu padoti bidaku golabu tupiro golabu tupiro golabu tupiro golabu bidaku tupiro bidaku golabu bidaku tupiro golabu padoti tupiro golabu bidaku golabu tupiro bidaku tupiro golabu tupiro bidaku padoti tupiro golabu tupiro bidaku padoti bidaku golabu bidaku tupiro golabu bidaku tupiro golabu padoti tupiro padoti golabu bidaku padoti tupiro padoti bidaku tupiro golabu bidaku golabu bidaku padoti bidaku tupiro golabu bidaku golabu bidaku padoti tupiro padoti bidaku padoti tupiro bidaku tupiro golabu bidaku tupiro padoti tupiro bidaku padoti tupiro golabu padoti tupiro golabu padoti tupiro golabu tupiro bidaku golabu padoti tupiro padoti golabu padoti bidaku padoti bidaku tupiro golabu padoti bidaku tupiro padoti golabu padoti golabu tupiro bidaku padoti bidaku padoti golabu padoti bidaku padoti golabu padoti tupiro bidaku padoti tupiro padoti tupiro golabu bidaku padoti golabu bidaku padoti tupiro bidaku tupiro bidaku tupiro padoti tupiro golabu tupiro bidaku golabu bidaku padoti bidaku golabu bidaku tupiro padoti tupiro padoti tupiro padoti tupiro bidaku golabu padoti golabu padoti golabu padoti tupiro golabu tupiro golabu tupiro bidaku golabu bidaku golabu padoti golabu padoti tupiro padoti tupiro golabu tupiro padoti bidaku padoti bidaku golabu padoti golabu bidaku padoti bidaku padoti golabu bidaku padoti bidaku padoti golabu bidaku padoti golabu tupiro padoti tupiro padoti bidaku golabu padoti bidaku tupiro padoti golabu padoti golabu bidaku padoti golabu tupiro golabu tupiro padoti bidaku tupiro bidaku padoti tupiro golabu padoti golabu padoti bidaku tupiro padoti bidaku golabu padoti bidaku golabu tupiro padoti tupiro padoti tupiro golabu bidaku padoti bidaku golabu bidaku golabu padoti tupiro golabu bidaku golabu tupiro golabu padoti golabu tupiro golabu bidaku golabu padoti tupiro padoti golabu tupiro golabu bidaku golabu tupiro padoti tupiro golabu padoti bidaku golabu tupiro bidaku tupiro golabu tupiro golabu tupiro bidaku padoti bidaku golabu bidaku golabu padoti tupiro golabu tupiro golabu bidaku golabu bidaku golabu padoti golabu padoti tupiro padoti bidaku golabu padoti tupiro padoti tupiro golabu bidaku padoti golabu bidaku tupiro padoti golabu tupiro padoti bidaku padoti golabu tupiro padoti bidaku golabu padoti tupiro bidaku padoti golabu bidaku padoti golabu bidaku padoti golabu bidaku tupiro bidaku padoti bidaku golabu padoti bidaku padoti bidaku padoti golabu padoti tupiro padoti tupiro golabu bidaku golabu bidaku golabu bidaku tupiro padoti tupiro bidaku golabu tupiro padoti golabu padoti bidaku padoti tupiro golabu tupiro padoti bidaku golabu padoti golabu tupiro bidaku tupiro padoti golabu bidaku golabu padoti golabu bidaku tupiro golabu tupiro padoti tupiro padoti tupiro bidaku padoti bidaku padoti bidaku golabu bidaku tupiro padoti golabu bidaku tupiro padoti tupiro golabu tupiro bidaku tupiro padoti bidaku golabu tupiro golabu padoti tupiro bidaku tupiro padoti golabu bidaku golabu bidaku padoti golabu tupiro golabu tupiro bidaku golabu tupiro padoti tupiro bidaku tupiro bidaku padoti bidaku padoti tupiro golabu tupiro golabu tupiro bidaku tupiro padoti tupiro bidaku padoti tupiro padoti tupiro bidaku golabu padoti tupiro padoti golabu bidaku padoti bidaku golabu bidaku golabu padoti bidaku tupiro bidaku padoti tupiro bidaku golabu tupiro padoti bidaku golabu tupiro bidaku golabu padoti bidaku golabu bidaku tupiro bidaku tupiro golabu bidaku padoti golabu padoti golabu tupiro golabu bidaku tupiro padoti golabu bidaku tupiro bidaku golabu bidaku tupiro padoti bidaku padoti tupiro padoti tupiro bidaku golabu tupiro golabu tupiro golabu padoti tupiro bidaku padoti tupiro golabu bidaku tupiro bidaku golabu tupiro padoti bidaku tupiro bidaku padoti bidaku golabu tupiro bidaku tupiro padoti bidaku tupiro golabu tupiro padoti tupiro golabu tupiro bidaku tupiro golabu padoti bidaku tupiro bidaku tupiro golabu padoti tupiro golabu bidaku padoti golabu padoti bidaku tupiro padoti bidaku tupiro golabu bidaku tupiro bidaku tupiro golabu padoti golabu tupiro padoti bidaku padoti golabu tupiro padoti bidaku tupiro golabu padoti bidaku padoti tupiro padoti golabu bidaku padoti bidaku golabu padoti golabu bidaku golabu padoti golabu tupiro bidaku tupiro golabu padoti golabu bidaku padoti bidaku padoti bidaku golabu padoti tupiro padoti bidaku padoti tupiro golabu padoti golabu tupiro bidaku tupiro golabu padoti golabu bidaku padoti bidaku padoti golabu padoti bidaku padoti bidaku golabu tupiro bidaku tupiro bidaku padoti golabu tupiro bidaku tupiro bidaku golabu padoti tupiro padoti tupiro padoti bidaku tupiro bidaku tupiro padoti bidaku tupiro golabu padoti golabu padoti golabu bidaku tupiro padoti bidaku golabu padoti tupiro padoti bidaku tupiro bidaku tupiro padoti tupiro golabu padoti golabu tupiro bidaku padoti tupiro bidaku golabu bidaku padoti golabu bidaku golabu bidaku tupiro padoti bidaku golabu tupiro padoti bidaku tupiro padoti bidaku tupiro padoti bidaku padoti golabu tupiro golabu tupiro golabu tupiro bidaku tupiro golabu bidaku padoti bidaku tupiro bidaku golabu tupiro bidaku golabu bidaku golabu padoti tupiro bidaku padoti tupiro bidaku tupiro padoti golabu padoti golabu padoti tupiro bidaku padoti tupiro golabu padoti tupiro golabu bidaku golabu bidaku tupiro padoti bidaku tupiro padoti golabu padoti tupiro golabu bidaku golabu bidaku golabu bidaku tupiro padoti golabu tupiro golabu padoti golabu bidaku tupiro golabu bidaku padoti tupiro padoti golabu tupiro bidaku golabu golabu

A.1.3. Consonant phoneme order Structured Stream 1

 # P = labial
 # K = dorsal
 # T = coronal

TPT PTK KTP TPT PTK TPT KTP PTK KTP PTT KTP PTK PTT KTP PTT PTK KTP TPT KTP TPT KTP TPT KTP PTK TPT PTK KTP PTK TPT KTP PTT TPT KTP PTK KTP TPT PTK TPT KTP TPT PTK PTT TPT KTP TPT PTK PTT PTK KTP PTK TPT KTP PTK TPT KTP PTT TPT PTT KTP PTK PTT TPT PTT PTK TPT KTP PTK KTP PTK PTT PTK TPT KTP PTK KTP PTK PTT TPT PTT PTK PTT TPT PTK TPT KTP PTK TPT PTT TPT PTK PTT TPT KTP PTT TPT KTP PTT TPT KTP TPT PTK KTP PTT TPT PTT KTP PTT PTK PTT PTK TPT KTP PTT PTK TPT PTT KTP PTT KTP TPT PTK PTT PTK PTT KTP PTT PTK PTT KTP PTT TPT PTK PTT TPT PTT TPT KTP PTK PTT KTP PTK PTT TPT PTK TPT PTK TPT PTT TPT KTP TPT PTK KTP PTK PTT PTK KTP PTK TPT PTT TPT PTT

TPT PTT TPT PTK KTP PTT KTP PTT KTP PTT TPT KTP TPT KTP TPT PTK KTP PTK KTP PTT KTP PTT TPT PTT TPT KTP TPT PTT PTK PTT PTK KTP PTT KTP PTK PTT PTK PTT KTP PTK PTT PTK PTT KTP PTK PTT KTP TPT PTT TPT PTT PTK KTP PTT PTK TPT PTT KTP PTT KTP PTK PTT KTP TPT KTP TPT PTT PTK TPT PTK PTT TPT KTP PTT KTP PTT PTK TPT PTT PTK KTP PTT PTK KTP TPT PTT TPT PTT TPT KTP PTK PTT PTK KTP PTK KTP PTT TPT KTP PTK KTP TPT KTP PTT KTP TPT KTP PTK KTP PTT TPT PTT KTP TPT KTP PTK KTP TPT PTT TPT KTP PTT PTK KTP TPT PTK TPT KTP TPT KTP TPT PTK PTT PTK KTP PTK KTP PTT TPT KTP TPT KTP PTK KTP PTK KTP PTT KTP PTT TPT PTT PTK KTP PTT TPT PTT TPT KTP PTK PTT KTP PTK TPT PTT KTP TPT PTT PTK PTT KTP TPT PTT PTK KTP PTT TPT PTK PTT KTP PTK PTT KTP PTK PTT KTP PTK TPT PTK PTT PTK KTP PTT PTK PTT PTK PTT KTP PTT TPT PTT TPT KTP PTK KTP PTK KTP PTK TPT PTT TPT PTK KTP TPT PTT KTP PTT PTK PTT TPT KTP TPT PTT PTK KTP PTT KTP TPT PTK TPT PTT KTP PTK KTP PTT KTP PTK TPT KTP TPT PTT TPT PTT TPT PTK PTT PTK PTT PTK KTP PTK TPT PTT KTP PTK TPT PTT TPT KTP TPT PTK TPT PTT PTK KTP TPT KTP PTT TPT PTK TPT PTT KTP PTK KTP PTK PTT KTP TPT KTP TPT PTK KTP TPT PTT TPT PTK TPT PTK PTT PTK PTT TPT KTP TPT KTP TPT PTK TPT PTT TPT PTK PTT TPT PTT TPT PTK KTP PTT TPT PTT KTP PTK PTT PTK KTP PTK KTP PTT PTK TPT PTK PTT TPT PTK KTP TPT PTT PTK KTP TPT PTK KTP PTT PTK KTP PTK TPT PTK TPT KTP PTK PTT KTP PTT KTP TPT KTP PTK TPT PTT KTP PTK TPT PTK KTP PTK TPT PTT PTK PTT TPT PTT TPT PTK KTP TPT KTP TPT KTP PTT TPT PTK PTT TPT KTP PTK TPT PTK KTP TPT PTT PTK TPT PTK PTT PTK KTP TPT PTK TPT PTT PTK TPT KTP TPT PTT TPT KTP TPT PTK TPT KTP PTT PTK TPT PTK TPT KTP PTT TPT KTP PTK PTT KTP PTT PTK TPT PTT PTK TPT KTP PTK TPT PTK TPT KTP PTT KTP TPT PTT PTK PTT KTP TPT PTT PTK TPT KTP PTT PTK PTT TPT PTT KTP PTK PTT PTK KTP PTT KTP PTK KTP PTT KTP TPT PTK TPT KTP PTT KTP PTK PTT PTK PTT PTK KTP PTT TPT PTT PTK PTT TPT KTP PTT KTP TPT PTK TPT KTP PTT KTP PTK PTT PTK PTT KTP PTT PTK PTT PTK KTP TPT PTK TPT PTK PTT KTP TPT PTK TPT PTK KTP PTT TPT PTT TPT PTT PTK TPT PTK TPT PTT PTK TPT KTP PTT KTP PTT KTP PTK TPT PTT PTK KTP PTT TPT PTT PTK TPT PTK TPT PTT TPT KTP PTT KTP TPT PTK PTT TPT PTK KTP PTK PTT KTP PTK KTP PTK TPT PTT PTK KTP TPT PTT PTK TPT PTT PTK TPT PTT PTK PTT KTP TPT KTP TPT KTP TPT PTK TPT KTP PTK PTT PTK TPT PTK KTP TPT PTK KTP PTK KTP PTT TPT PTK PTT TPT PTK TPT PTT KTP PTT KTP PTT TPT PTK PTT TPT KTP PTT TPT KTP PTK KTP PTK TPT PTT PTK TPT PTT KTP PTT TPT KTP PTK KTP PTK KTP PTK TPT PTT KTP TPT KTP PTT KTP PTK TPT KTP PTK PTT TPT PTT KTP TPT PTK KTP KTP

A.2. Structured Stream 2

A.2.1. Structured Stream 2 as provided by B&P2017

5 4 2 11 1 8 10 3 6 11 1 8 12 7 9 11 1 8 10 3 6 12 7 9 11 1 8 12 7 9 10 3 6 12 7 9 11 1 8 5 4 2
10 3 6 5 4 2 10 3 6 5 4 2 12 7 9 10 3 6 11 1 8 10 3 6 12 7 9 11 1 8 5 4 2 11 1 8 10 3 6 11 1 8 5
4 2 10 3 6 12 7 9 11 1 8 10 3 6 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 5 4 2 12 7 9 10 3 6 11 1 8 10 3
6 11 1 8 12 7 9 10 3 6 12 7 9 11 1 8 10 3 6 5 4 2 10 3 6 11 1 8 10 3 6 11 1 8 12 7 9 5 4 2 11 1 8
10 3 6 5 4 2 11 1 8 12 7 9 10 3 6 5 4 2 12 7 9 10 3 6 12 7 9 5 4 2 10 3 6 11 1 8 10 3 6 11 1 8 12
7 9 10 3 6 5 4 2 11 1 8 5 4 2 11 1 8 5 4 2 10 3 6 12 7 9 5 4 2 11 1 8 12 7 9 10 3 6 11 1 8 5 4 2
10 3 6 11 1 8 10 3 6 11 1 8 5 4 2 12 7 9 11 1 8 12 7 9 10 3 6 5 4 2 10 3 6 5 4 2 10 3 6 11 1 8 10
3 6 11 1 8 10 3 6 11 1 8 5 4 2 12 7 9 10 3 6 5 4 2 10 3 6 5 4 2 10 3 6 12 7 9 5 4 2 12 7 9 5 4 2
12 7 9 5 4 2 11 1 8 12 7 9 5 4 2 11 1 8 12 7 9 5 4 2 10 3 6 5 4 2 10 3 6 5 4 2 12 7 9 10 3 6 5 4 2
11 1 8 10 3 6 12 7 9 10 3 6 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 11 1 8 10 3 6 11 1 8 5 4 2 10 3 6
11 1 8 5 4 2 12 7 9 5 4 2 12 7 9 11 1 8 10 3 6 5 4 2 10 3 6 5 4 2 12 7 9 11 1 8 12 7 9 10 3 6 5 4
2 10 3 6 12 7 9 5 4 2 11 1 8 10 3 6 11 1 8 10 3 6 12 7 9 10 3 6 5 4 2 12 7 9 10 3 6 11 1 8 10 3 6
11 1 8 10 3 6 12 7 9 10 3 6 5 4 2 12 7 9 11 1 8 5 4 2 10 3 6 5 4 2 12 7 9 11 1 8 10 3 6 11 1 8 12
7 9 11 1 8 10 3 6 5 4 2 10 3 6 12 7 9 5 4 2 12 7 9 11 1 8 5 4 2 12 7 9 10 3 6 11 1 8 10 3 6 5 4 2
10 3 6 5 4 2 12 7 9 11 1 8 10 3 6 11 1 8 10 3 6 11 1 8 12 7 9 10 3 6 11 1 8 12 7 9 10 3 6 12 7 9
11 1 8 10 3 6 11 1 8 5 4 2 10 3 6 12 7 9 11 1 8 12 7 9 11 1 8 10 3 6 11 1 8 12 7 9 5 4 2 10 3 6 5
4 2 12 7 9 5 4 2 12 7 9 11 1 8 12 7 9 10 3 6 5 4 2 10 3 6 12 7 9 5 4 2 10 3 6 5 4 2 12 7 9 5 4 2
11 1 8 5 4 2 10 3 6 5 4 2 12 7 9 11 1 8 5 4 2 10 3 6 11 1 8 5 4 2 11 1 8 10 3 6 11 1 8 10 3 6 5 4
2 11 1 8 10 3 6 11 1 8 10 3 6 5 4 2 11 1 8 10 3 6 12 7 9 10 3 6 12 7 9 5 4 2 10 3 6 11 1 8 5 4 2
10 3 6 12 7 9 11 1 8 5 4 2 12 7 9 5 4 2 10 3 6 11 1 8 12 7 9 5 4 2 12 7 9 5 4 2 12 7 9 5 4 2 11 1
8 10 3 6 12 7 9 5 4 2 10 3 6 12 7 9 11 1 8 12 7 9 10 3 6 12 7 9 5 4 2 11 1 8 10 3 6 5 4 2 11 1 8
5 4 2 12 7 9 11 1 8 10 3 6 12 7 9 10 3 6 12 7 9 11 1 8 12 7 9 11 1 8 12 7 9 5 4 2 10 3 6 11 1 8
10 3 6 12 7 9 10 3 6 5 4 2 12 7 9 10 3 6 5 4 2 10 3 6 11 1 8 5 4 2 10 3 6 11 1 8 10 3 6 12 7 9 5
4 2 11 1 8 5 4 2 10 3 6 12 7 9 5 4 2 10 3 6 5 4 2 12 7 9 11 1 8 5 4 2 12 7 9 10 3 6 5 4 2 11 1 8 5
4 2 10 3 6 5 4 2 12 7 9 5 4 2 10 3 6 11 1 8 5 4 2 10 3 6 12 7 9 10 3 6 11 1 8 10 3 6 12 7 9 10 3
6 12 7 9 5 4 2 10 3 6 12 7 9 10 3 6 11 1 8 5 4 2 12 7 9 11 1 8 10 3 6 12 7 9 10 3 6 5 4 2 11 1 8
12 7 9 10 3 6 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 10 3 6 12 7 9 10 3 6 12 7 9 10 3 6 5 4 2 11 1 8 12
7 9 10 3 6 5 4 2 11 1 8 12 7 9 5 4 2 10 3 6 11 1 8 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 11 1 8 12 7 9
10 3 6 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 11 1 8 5 4 2 12 7 9 10 3 6 5 4 2 12 7 9 5 4 2 10 3 6 11 1
8 10 3 6 11 1 8 5 4 2 10 3 6 11 1 8 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 10 3 6 11 1 8 5 4 2
12 7 9 11 1 8 12 7 9 5 4 2 12 7 9 5 4 2 10 3 6 5 4 2 11 1 8 10 3 6 5 4 2 12 7 9 10 3 6 12 7 9 5 4
2 11 1 8 5 4 2 12 7 9 11 1 8 5 4 2 11 1 8 5 4 2 12 7 9 11 1 8 10 3 6 11 1 8 10 3 6 5 4 2 10 3 6
11 1 8 12 7 9 5 4 2 11 1 8 5 4 2 11 1 8 12 7 9 11 1 8 12 7 9 11 1 8 5 4 2 11 1 8 12 7 9 10 3 6 5
4 2 12 7 9 10 3 6 11 1 8 10 3 6 5 4 2 11 1 8 12 7 9 11 1 8 10 3 6 11 1 8 5 4 2 11 1 8 10 3 6 12 7
9 5 4 2 11 1 8 12 7 9 5 4 2 10 3 6 12 7 9 10 3 6 11 1 8 10 3 6 12 7 9 10 3 6 12 7 9 10 3 6 11 1 8
12 7 9 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 5 4 2 11 1 8 10 3 6 12 7 9 10 3 6 11 1 8 5 4 2 10 3 6 11
1 8 12 7 9 5 4 2 12 7 9 5 4 2 10 3 6 11 1 8 5 4 2 11 1 8 12 7 9 11 1 8 5 4 2 10 3 6 12 7 9 11 1 8
10 3 6 11 1 8 10 3 6 12 7 9 11 1 8 10 3 6 5 4 2 12 7 9 11 1 8 5 4 2 11 1 8 12 7 9 5 4 2 10 3 6 12
7 9 10 3 6 11 1 8 12 7 9 10 3 6 11 1 8 5 4 2 11 1 8 12 7 9 10 3 6 5 4 2 10 3 6 11 1 8 12 7 9 10 3
6 5 4 2 10 3 6 5 4 2 12 7 9 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 5 4 2 10 3 6 11 1 8 10 3 6 11
1 8 12 7 9 5 4 2 11 1 8 10 3 6 5 4 2 10 3 6 11 1 8 10 3 6 5 4 2 12 7 9 5 4 2 12 7 9 5 4 2 12 7 9
10 3 6 5 4 2 12 7 9 5 4 2 12 7 9 5 4 2 11 1 8 12 7 9 11 1 8 5 4 2 12 7 9 11 1 8 10 3 6 11 1 8 5 4
2 11 1 8 10 3 6 12 7 9 5 4 2 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 5 4 2 10 3 6 12 7 9 11 1 8 5 4 2
11 1 8 12 7 9 11 1 8 12 7 9 11 1 8 5 4 2 12 7 9 5 4 2 12 7 9 11 1 8 12 7 9 10 3 6 11 1 8 12 7 9
11 1 8 10 3 6 5 4 2 11 1 8 12 7 9 5 4 2 11 1 8 12 7 9 5 4 2 11 1 8 12 7 9 11 1 8 10 3 6 11 1 8 5
4 2 12 7 9 10 3 6 11 1 8 5 4 2 11 1 8 12 7 9 5 4 2 11 1 8 5 4 2 12 7 9 11 1 8 5 4 2 10 3 6 12 7 9
5 4 2 12 7 9 11 1 8 12 7 9 11 1 8 5 4 2 12 7 9 5 4 2 11 1 8 5 4 2 11 1 8 10 3 6 12 7 9 11 1 8 5 4
2 11 1 8 5 4 2 12 7 9 10 3 6 5 4 2 12 7 9 10 3 6 12 7 9 11 1 8 10 3 6 12 7 9 5 4 2 12 7 9 11 1 8

12 7 9 11 1 8 10 3 6 12 7 9 10 3 6 12 7 9 10 3 6 12 7 9 5 4 2 11 1 8 12 7 9 11 1 8 12 7 9 5 4 2 11 1 8 12 7 9 11 1 8 10 3 6 5 4 2 10 3 6 11 1 8 10 3 6 11 1 8 5 4 2 12 7 9 5 4 2 12 7 9 11 1 8 12 7 9 11 1 8 5 4 2 12 7 9 10 3 6 12 7 9 10 3 6 5 4 2 11 1 8 5 4 2 12 7 9 11 1 8 5 4 2 11 1 8 12 7 9 11 1 8 12 7 9 10 3 6 5 4 2 11 1 8 10 3 6 11 1 8 12 7 9 10 3 6 5 4 2 11 1 8 5 4 2 10 3 6 11 1 8 5 4 2 12 7 9 5 4 2 12 7 9 11 1 8 12 7 9 10 3 6 5 4 2 12 7 9 11 1 8 5 4 2 12 7 9 11 1 8 5 4 2 12 7 9 10 3 6 12 7 9 11 1 8 5 4 2 10 3 6 12 7 9 5 4 2 12 7 9 10 3 6 5 4 2 12 7 9 11 1 8 10 3 6 12 7 9 11 1 8 12 7 9 11 1 8 12 7 9 5 4 2 11 1 8 10 3 6 11 1 8 5 4 2 12 7 9 5 4 2 10 3 6 5 4 2 10 3 6 5 4 2 11 1 8 5 4 2 10 3 6 5 4 2 10 3 6 10 3 6

A.2.2. Structured Stream 2 in words

melugi rafinu pukemi rafinu tonapo rafinu pukemi tonapo rafinu tonapo pukemi tonapo rafinu melugi pukemi melugi pukemi melugi tonapo pukemi rafinu pukemi tonapo rafinu melugi rafinu pukemi rafinu melugi pukemi tonapo rafinu pukemi rafinu pukemi tonapo melugi tonapo melugi tonapo pukemi rafinu pukemi rafinu tonapo pukemi tonapo rafinu pukemi melugi pukemi rafinu pukemi rafinu tonapo melugi rafinu pukemi melugi rafinu tonapo pukemi melugi tonapo pukemi tonapo melugi pukemi rafinu pukemi rafinu tonapo pukemi melugi rafinu melugi rafinu melugi pukemi tonapo melugi rafinu tonapo pukemi rafinu melugi pukemi rafinu pukemi rafinu melugi tonapo rafinu tonapo pukemi melugi pukemi melugi pukemi rafinu pukemi rafinu pukemi rafinu melugi tonapo pukemi melugi pukemi melugi pukemi tonapo melugi tonapo melugi tonapo melugi rafinu tonapo melugi rafinu tonapo melugi pukemi melugi pukemi melugi tonapo pukemi melugi rafinu pukemi tonapo pukemi rafinu pukemi tonapo melugi tonapo rafinu pukemi rafinu melugi pukemi rafinu melugi tonapo melugi tonapo rafinu pukemi melugi pukemi melugi tonapo rafinu tonapo pukemi melugi pukemi tonapo melugi rafinu pukemi rafinu pukemi tonapo pukemi melugi tonapo pukemi rafinu pukemi rafinu pukemi tonapo pukemi melugi tonapo rafinu melugi pukemi melugi tonapo rafinu pukemi rafinu tonapo rafinu pukemi melugi pukemi tonapo melugi tonapo rafinu melugi tonapo pukemi rafinu pukemi melugi pukemi melugi tonapo rafinu pukemi rafinu pukemi rafinu tonapo pukemi rafinu tonapo pukemi tonapo rafinu pukemi rafinu melugi pukemi tonapo rafinu tonapo rafinu pukemi rafinu tonapo melugi pukemi melugi tonapo melugi tonapo rafinu tonapo pukemi melugi pukemi tonapo melugi pukemi melugi tonapo melugi rafinu melugi pukemi melugi tonapo rafinu melugi pukemi rafinu melugi rafinu pukemi rafinu pukemi melugi rafinu pukemi rafinu pukemi melugi rafinu pukemi tonapo pukemi tonapo melugi pukemi rafinu melugi pukemi tonapo rafinu melugi tonapo melugi pukemi rafinu tonapo melugi tonapo melugi tonapo melugi rafinu pukemi tonapo melugi pukemi tonapo rafinu tonapo pukemi tonapo melugi rafinu pukemi melugi rafinu melugi tonapo rafinu pukemi tonapo pukemi tonapo rafinu tonapo rafinu tonapo melugi pukemi rafinu pukemi tonapo pukemi melugi tonapo pukemi melugi pukemi rafinu melugi pukemi rafinu pukemi tonapo melugi rafinu melugi pukemi tonapo melugi pukemi melugi tonapo rafinu melugi tonapo pukemi melugi rafinu melugi pukemi melugi tonapo melugi pukemi rafinu melugi pukemi tonapo pukemi rafinu pukemi tonapo pukemi tonapo melugi pukemi tonapo pukemi rafinu melugi tonapo rafinu pukemi tonapo pukemi melugi rafinu tonapo pukemi melugi rafinu pukemi tonapo melugi pukemi tonapo pukemi tonapo pukemi melugi rafinu tonapo pukemi melugi rafinu tonapo melugi pukemi rafinu melugi rafinu pukemi tonapo melugi rafinu tonapo pukemi melugi rafinu pukemi tonapo melugi rafinu melugi tonapo pukemi melugi tonapo melugi pukemi rafinu pukemi rafinu melugi pukemi rafinu melugi rafinu pukemi tonapo melugi tonapo pukemi rafinu melugi tonapo rafinu tonapo melugi tonapo melugi pukemi melugi rafinu pukemi melugi tonapo pukemi tonapo melugi rafinu melugi tonapo rafinu melugi rafinu melugi tonapo rafinu pukemi rafinu pukemi melugi pukemi rafinu tonapo melugi rafinu melugi rafinu tonapo rafinu tonapo rafinu melugi rafinu tonapo pukemi melugi tonapo pukemi rafinu pukemi melugi rafinu tonapo rafinu pukemi rafinu melugi rafinu pukemi tonapo melugi rafinu tonapo melugi pukemi tonapo pukemi rafinu pukemi tonapo pukemi tonapo pukemi rafinu tonapo rafinu pukemi tonapo melugi tonapo melugi rafinu pukemi tonapo pukemi rafinu melugi pukemi rafinu tonapo melugi tonapo melugi pukemi rafinu melugi rafinu tonapo rafinu melugi pukemi tonapo rafinu pukemi rafinu pukemi tonapo rafinu pukemi melugi tonapo rafinu melugi rafinu tonapo melugi pukemi tonapo pukemi rafinu tonapo pukemi rafinu melugi rafinu tonapo pukemi melugi pukemi rafinu tonapo pukemi melugi pukemi melugi tonapo melugi rafinu pukemi tonapo melugi tonapo melugi pukemi rafinu pukemi rafinu tonapo melugi rafinu pukemi melugi pukemi rafinu pukemi melugi tonapo melugi tonapo melugi tonapo pukemi melugi tonapo melugi tonapo melugi rafinu tonapo rafinu melugi tonapo rafinu pukemi rafinu melugi rafinu pukemi tonapo melugi rafinu pukemi tonapo melugi tonapo melugi pukemi tonapo rafinu melugi rafinu tonapo rafinu tonapo rafinu melugi tonapo melugi tonapo rafinu tonapo pukemi rafinu tonapo rafinu pukemi melugi rafinu tonapo melugi rafinu tonapo melugi rafinu tonapo rafinu pukemi rafinu melugi tonapo pukemi rafinu melugi rafinu tonapo melugi rafinu melugi tonapo rafinu melugi pukemi tonapo melugi tonapo rafinu tonapo rafinu melugi tonapo melugi rafinu melugi rafinu pukemi tonapo rafinu melugi rafinu melugi tonapo pukemi melugi tonapo pukemi tonapo rafinu pukemi tonapo melugi tonapo rafinu tonapo rafinu pukemi tonapo pukemi tonapo pukemi tonapo melugi rafinu tonapo rafinu tonapo melugi rafinu tonapo rafinu pukemi melugi pukemi rafinu pukemi rafinu melugi tonapo melugi tonapo rafinu tonapo rafinu melugi tonapo pukemi tonapo pukemi melugi rafinu melugi tonapo rafinu melugi rafinu tonapo rafinu tonapo pukemi melugi rafinu pukemi rafinu tonapo pukemi melugi rafinu melugi pukemi rafinu melugi tonapo melugi tonapo rafinu tonapo pukemi melugi tonapo rafinu melugi tonapo rafinu melugi tonapo pukemi tonapo rafinu melugi pukemi tonapo melugi tonapo pukemi melugi tonapo rafinu pukemi tonapo rafinu tonapo rafinu tonapo melugi rafinu pukemi rafinu melugi tonapo melugi pukemi melugi pukemi melugi rafinu melugi pukemi melugi pukemi pukemi

A.2.3. Consonant phoneme order Structured Stream 2

 # P = labial
 # K = dorsal
 # T = coronal

PTK TPT PKP TPT TTP TPT PKP TTP TPT TTP PKP TTP TPT PTK PKP PTK PKP PTK TTP PKP TPT PKP TTP TPT PTK TPT PKP TPT PTK PKP TTP TPT PKP TPT PKP TTP PTK TTP PTK TTP PKP TPT PKP TPT TTP PKP TTP TPT PKP PTK PKP TPT PKP TPT TTP PTK TPT PKP PTK TPT TTP PKP PTK TTP PKP TTP PTK PKP TPT PKP TPT TTP PKP PTK TPT PTK TPT PTK PKP TTP PTK TPT TTP PKP TPT PTK PKP TPT PKP TPT PTK TTP TPT TTP PKP PTK PKP PTK PKP TPT PKP TPT PKP TPT PTK TTP PKP PTK PKP PTK PKP TTP PTK TTP PTK TTP PTK TPT TTP PTK TPT TTP PTK PKP PTK PKP PTK TTP PKP PTK TPT PKP TTP PKP TPT PKP TTP PTK TTP TPT PKP TPT PTK PKP TPT PTK TTP PTK TTP TPT PKP PTK PKP PTK TTP TPT TTP PKP PTK PKP TTP PTK TPT PKP TPT PKP TTP PKP PTK TTP PKP TPT PKP TPT PKP TTP PKP PTK TTP TPT PTK PKP PTK TTP TPT PKP TPT TTP TPT PKP PTK PKP TTP PTK TTP TPT PTK TTP PKP TPT PKP PTK PKP PTK TTP TPT PKP TPT PKP TPT TTP PKP TPT TTP PKP TTP TPT PKP TPT PTK PKP TTP TPT TTP TPT PKP TPT TTP PTK PKP PTK TTP PTK TTP TPT TTP PKP PTK PKP TTP PTK PKP PTK TTP PTK TPT PTK PKP PTK TTP TPT PTK PKP TPT PTK TPT PKP TPT PKP PTK TPT PKP TPT PKP PTK TPT PKP TTP PKP TTP PTK PKP TPT PTK PKP TTP TPT PTK TTP PTK PKP TPT TTP PTK TTP PTK TTP PTK TPT PKP TTP PTK PKP TTP TPT TTP PKP TTP PTK TPT PKP PTK TPT PTK TTP TPT PKP TTP PKP TTP TPT TTP TPT TTP PTK PKP TPT PKP TTP PKP PTK TTP PKP PTK PKP TPT PTK PKP TPT PKP TTP PTK TPT PTK PKP TTP PTK PKP PTK TTP TPT PTK TTP PKP PTK TPT PTK PKP PTK TTP PTK PKP TPT PTK PKP TTP PKP TPT PKP TTP PKP TTP PTK PKP TTP PKP TPT PTK TTP TPT PKP TTP PKP PTK TPT TTP PKP PTK TPT PKP TTP PTK PKP TTP PKP TTP PKP PTK TPT TTP PKP PTK TPT TTP PTK PKP

TPT PTK TPT PKP TTP PTK TPT TTP PKP PTK TPT PKP TTP PTK TPT PTK TTP PKP PTK TTP PTK PKP TPT PKP TPT PTK PKP TPT PTK TPT PKP TTP PTK TTP PKP TPT PTK TTP TPT TTP PTK TTP PTK PKP PTK TPT PKP PTK TTP PKP TTP PTK TPT PTK TTP TPT PTK TPT PTK TTP TPT PKP TPT PKP PTK PKP TPT TTP PTK TPT PTK TPT TTP TPT TTP TPT PTK TPT TTP PKP PTK TTP PKP TPT PKP PTK TPT TTP TPT PKP TPT PTK TPT PKP TTP PTK TPT TTP PTK PKP TTP PKP TPT PKP TTP PKP TTP PKP TPT TTP TPT PKP TTP PTK TTP PTK TPT PKP TTP PKP TPT PTK PKP TPT TTP PTK TTP PTK PKP TPT PTK TPT TTP TPT PTK PKP TTP TPT PKP TPT PKP TTP TPT PKP PTK TTP TPT PTK TPT TTP PTK PKP TTP PKP TPT TTP PKP TPT PTK TPT TTP PKP PTK PKP TPT TTP PKP PTK PKP PTK TTP PTK TPT PKP TTP PTK TTP PTK PKP TPT PKP TPT TTP PTK TPT PKP PTK PKP TPT PKP PTK TTP PTK TTP PTK TTP PKP PTK TTP PTK TTP PTK TPT TTP TPT PTK TTP TPT PKP TPT PTK TPT PKP TTP PTK TPT PKP TTP PTK TTP PTK PKP TTP TPT PTK TPT TTP TPT TTP TPT PTK TTP PTK TTP TPT TTP PKP TPT TTP TPT PKP PTK TPT TTP PTK TPT TTP PTK TPT TTP TPT PKP TPT PTK TTP PKP TPT PTK TPT TTP PTK TPT PTK TTP TPT PTK PKP TTP PTK TTP TPT TTP TPT PTK TTP PTK TPT PTK TPT PKP TTP TPT PTK TPT PTK TTP PKP PTK TTP PKP TTP TPT PKP TTP PTK TTP TPT TTP TPT PKP TTP PKP TTP PKP TTP PTK TPT TTP TPT TTP PTK TPT TTP TPT PKP PTK PKP TPT PKP TPT PTK TTP PTK TTP TPT TTP TPT PTK TTP PKP TTP PKP PTK TPT PTK TTP TPT PTK TPT TTP TPT TTP PKP PTK TPT PKP TPT TTP PKP PTK TPT PTK PKP TPT PTK TTP PTK TTP TPT TTP PKP PTK TTP TPT PTK TTP TPT PTK TTP PKP TTP TPT PTK PKP TTP PTK TTP PKP PTK TTP TPT PKP TTP TPT TTP TPT TTP PTK TPT PKP TPT PTK TTP PTK PKP PTK PKP PTK TPT PTK PKP PTK PKP PKP

A.3. Random Stream 1

A.3.1. Random Stream 1 as provided by B&P2017

6 5 1 2 7 10 2 7 3 11 10 9 2 4 5 1 12 11 12 2 11 7 12 4 8 7 6 8 10 11 3 11 2 5 2 8 7 10 12 9 11
4 10 5 10 9 8 4 3 9 3 5 2 8 6 10 12 8 9 5 7 8 5 12 5 9 6 11 10 9 10 6 5 11 10 8 6 7 1 7 1 6 11 10
7 5 4 8 5 8 7 6 9 1 7 5 9 12 6 2 8 7 12 5 3 10 6 5 3 11 6 2 1 4 1 6 11 2 9 5 8 5 4 3 4 12 3 6 3 5 4
8 9 10 8 7 1 6 12 10 12 3 9 5 8 1 3 10 7 1 11 8 9 8 7 5 2 5 2 11 6 9 1 10 9 7 1 4 10 9 8 10 3 10
11 1 5 7 6 9 11 2 10 5 2 5 10 3 10 4 2 5 8 12 7 4 8 9 5 8 1 2 10 1 11 1 8 3 9 6 1 6 1 9 7 2 5 4 3
8 2 8 12 9 4 3 5 9 11 12 10 6 9 5 11 9 4 1 7 6 12 1 7 11 10 9 8 12 11 7 2 7 1 6 2 4 7 1 3 4 3 10
3 7 4 2 9 10 5 12 1 9 5 12 8 5 7 9 3 10 9 2 10 11 2 1 12 6 1 6 3 7 9 6 8 5 10 6 10 7 9 8 6 7 12
10 7 4 2 11 7 4 5 6 2 7 5 3 12 10 7 2 5 12 1 3 10 4 7 2 9 5 9 12 9 11 1 6 10 12 1 7 5 12 2 1 4 1
8 12 11 5 6 2 7 10 1 12 6 7 11 4 12 1 12 8 11 8 6 4 9 6 3 5 10 11 5 11 6 7 4 1 3 7 6 7 6 3 4 11 2
3 7 11 5 2 6 10 4 5 10 8 2 1 3 5 3 9 11 7 2 1 8 6 4 6 11 7 4 1 3 7 1 4 7 4 5 11 3 2 9 1 12 10 1 6
11 3 9 6 4 7 1 7 5 2 4 2 12 3 12 5 9 1 12 6 8 5 6 2 10 1 10 9 6 5 6 2 3 10 7 11 1 4 1 10 12 10 6
4 7 12 7 11 8 11 5 4 10 1 12 4 1 3 8 12 6 4 2 7 4 2 6 2 8 2 4 10 7 8 3 11 10 3 7 5 4 11 7 3 6 10
11 2 5 8 1 8 6 2 6 2 5 3 7 10 9 6 4 11 3 4 3 7 8 6 3 8 6 10 6 5 7 11 9 2 1 6 10 6 3 7 1 6 11 8 12
11 8 3 12 7 6 10 12 9 8 11 9 12 6 8 6 5 4 2 10 8 4 8 5 7 6 12 3 6 2 4 3 9 5 12 8 4 10 11 4 3 12 7
8 12 1 11 3 12 4 12 9 4 7 2 11 12 4 3 1 7 1 9 1 2 12 5 10 6 8 12 3 7 12 1 6 4 11 3 11 1 6 8 11 7
5 2 8 2 11 2 7 6 10 3 4 6 4 7 1 2 9 4 3 7 4 1 11 10 7 9 4 12 1 5 2 6 9 1 2 8 11 12 2 11 6 5 12 1 9
12 6 7 9 10 9 6 11 8 4 3 2 3 6 10 7 1 12 3 12 2 6 1 4 10 9 10 11 9 7 4 2 5 2 10 2 9 3 7 5 8 5 9
10 1 10 4 5 6 2 11 12 8 1 2 4 6 2 5 4 1 4 11 10 1 8 9 8 2 3 8 9 4 11 12 9 8 12 4 9 3 9 7 12 1 7
11 7 3 5 12 6 11 12 3 4 11 4 1 9 4 5 11 4 3 6 4 7 8 5 1 2 5 6 1 10 5 3 12 4 5 11 4 6 1 4 8 7 8 9 2
8 6 10 3 11 7 6 12 6 4 12 4 8 7 11 5 11 12 5 6 1 4 3 9 5 4 6 9 8 5 6 12 11 2 10 3 9 7 12 10 11 6
9 5 3 8 9 3 2 12 9 3 10 2 8 12 5 8 1 12 7 12 5 11 5 6 8 10 9 3 7 11 2 1 2 7 1 11 3 8 11 6 7 8 4
12 5 6 10 11 7 5 4 11 8 10 5 7 10 3 9 7 1 7 6 3 6 7 6 2 1 12 4 8 1 3 9 12 2 8 9 5 10 5 10 8 1 5 2
11 5 1 3 7 3 2 8 9 8 10 5 9 1 11 1 2 9 8 1 8 1 3 5 9 2 8 5 11 9 3 1 6 9 3 1 6 3 4 5 1 6 11 3 5 12 8
7 5 1 5 4 3 9 6 2 3 7 5 7 9 6 12 1 4 1 7 12 9 12 1 2 5 11 12 4 1 8 5 4 2 7 6 5 1 8 10 5 7 4 11 3 2
5 12 4 7 4 6 7 8 1 5 7 11 9 12 11 10 2 3 8 1 8 9 11 10 7 9 11 6 9 8 4 2 6 4 1 9 4 9 1 7 6 4 1 10 8
4 8 3 5 7 4 11 10 1 9 11 2 8 12 4 3 8 5 6 1 5 8 10 9 1 3 7 10 7 3 9 4 2 12 8 6 10 7 3 8 7 2 5 3 12
4 1 6 7 4 5 1 10 2 7 2 10 2 8 11 1 10 11 1 4 7 12 7 3 8 9 2 7 4 9 3 9 3 10 1 8 10 6 12 10 6 5 3 6
7 10 5 11 8 12 9 10 7 3 7 10 4 12 2 4 6 7 2 5 1 8 3 5 12 8 2 7 1 8 1 12 1 7 2 11 7 6 12 6 4 12 4
12 5 2 6 3 7 9 1 10 3 9 3 7 10 2 11 6 2 11 4 12 6 9 1 4 5 9 6 2 3 1 11 3 11 7 12 7 12 2 10 2 11 2
11 10 8 2 10 1 4 3 6 2 3 10 6 8 4 10 9 7 8 10 8 4 3 8 10 11 7 1 10 7 8 3 12 6 10 5 11 8 7 8 11 9
7 10 2 11 5 2 4 9 1 8 4 6 1 6 1 8 12 1 8 3 10 3 10 5 4 9 6 11 12 7 3 10 12 9 10 12 11 2 1 12 6 1
8 4 11 1 12 5 1 6 11 4 9 3 7 8 6 4 7 5 9 4 11 2 9 4 3 10 8 12 7 8 12 10 9 3 1 11 4 3 9 6 3 1 6 12
9 11 9 5 6 1 5 7 8 11 6 9 2 3 10 3 4 7 11 12 9 4 2 5 11 4 5 8 12 7 5 2 4 3 5 1 10 6 9 6 4 10 7 3 2
5 12 10 8 3 7 5 10 9 5 11 9 11 6 12 2 8 9 4 9 11 8 12 9 11 3 4 12 1 2 3 9 6 11 3 12 4 9 5 12 9 7
1 4 7 3 6 10 2 1 12 5 10 11 4 1 10 12 3 12 9 7 1 2 4 3 5 10 9 2 11 2 4 7 2 11 1 6 10 11 9 4 6 4
10 12 3 5 11 3 1 9 7 5 3 4 11 3 7 11 12 4 9 10 11 7 3 9 10 1 10 7 11 4 10 8 4 5 7 3 1 11 2 11 6
9 7 12 1 9 1 10 11 6 2 6 8 6 2 8 7 8 12 5 3 10 5 3 12 5 6 9 6 4 3 5 10 2 1 7 8 3 1 11 1 11 3 11 9
11 8 1 10 1 8 7 12 2 1 10 9 11 8 5 1 9 8 12 10 12 2 9 7 3 8 2 9 5 9 7 6 10 4 5 4 3 4 5 6 8 2 5 10
4 11 6 4 8 3 10 5 2 7 2 6 4 6 5 3 12 5 12 7 12 6 10 9 1 7 5 3 5 9 6 8 3 1 8 7 5 9 8 4 11 9 2 4 12
2 11 12 11 2 10 8 12 6 8 1 12 7 1 8 9 8 9 10 1 7 4 3 8 1 8 9 2 9 6 9 2 8 9 8 6 12 10 9 10 7 4 5 1
10 2 4 6 12 8 7 12 3 10 8 12 2 10 8 1 11 10 11 3 2 12 3 7 10 1 2 11 8 1 4 5 8 9 1 7 9 2 9 7 11 2
8 4 12 7 4 12 10 8 11 4 12 4 7 9 1 2 8 1 7 2 10 12 7 12 9 11 12 10 9 2 1 3 10 12 2 5 10 5 11 12
1 4 1 6 9 2 8 6 9 5 10 4 7 9 2 4 11 2 3 7 2 6 11 6 9 8 1 3 9 7 10 4 11 4 2 1 2 6 2 5 1 5 12 10 6 4
9 3 2 3 6 2 12 11 1 12 4 3 2 10 3 10 4 11 8 2 4 5 6 10 6 1 12 9 3 6 11 3 5 11 10 2 3 11 12 9 7 2
3 9 1 3 5 12 4 5 4 8 5 4 2 11 9 11 7 8 1 3 7 1 11 5 9 12 3 4 7 12 6 11 12 3 2 8 12 11 12 4 6 11 8
7 2 5 10 11 8 9 7 12 2 9 11 8 10 8 12 1 2 9 4 7 11 1 12 10 11 2 4 9 7 3 12 3 2 8 5 7 10 3 2 10 3
12 9 11 10 11 12 2 9 3 2 9 7 4 1 12 11 2 4 6 2 9 2 7 4 1 10 8 6 9 5 1 12 10 8 12 11 12 8 7 2 12
10 2 5 10 6 4 10 8 10 8 6 3 6 8 3 12 8 7 4 1 7 2 5 2 11 12 11 6 2 10 2 12 4 8 12 8 6 12 10 7 1 4

10 3 5 2 5 9 6 8 4 2 11 7 10 1 2 4 12 6 11 10 5 8 5 12 5 8 9 3 12 5 12 2 1 12 6 5 11 9 11 9 8 10 9 2 5 1 7 12 8 10 5 10 9 4 1 5 11 6 10 1 6 12 5 2 5 4 8 10 1 7 3 6 3 9 8 4 6 9 11 5 7 8 11 5 7 5 4 5 7 11 3 4 3 10 1 2 1 8 1 9 12 11 3 6 10 5 4 11 9 6 1 12 1 3 8 4 12 6 8 3 2 8 4 10 5 3 12 4 5 11 12 1 12 6 12 10 2 3 10 3 9 1 11 5 11 3 7 6 5 1 2 7 3 2 8 10 6 7 9 4 3 9 4 2 10 11 1 9 11 12 10 8 6 4 10 3 4 7 1 9 10 3 8 5 11 8 5 11 9 6 10 3 1 3 10 9 5 4 10 11 4 8 9 7 3 8 2 4 6 12 4 3 5 11 2 10 5 8 11 1 3 7 5 6 9 2 12 1 9 11 5 11 1 9 6 2 4 6 8 5 9 6 3 6 5 7 10 3 6 4 1 5 3 5 3 11 2 12 3 8 1 12 6 2 11 5 12 9 7 4 12 1 6 7 9 2 5 12 5 8 2 11 6 12 6 10 6 2 3 8 6 5 8 3 1 3 11 10 3 12 3 11 9 12 2 11 10 11 8 2 12 11 8 10 11 6 4

A.3.2. Random Stream 1 in syllables

ku go bi bu la ro bu la da ti ro pi bu do go bi tu ti tu bu ti la tu do pa la ku pa ro ti da ti bu go bu pa la ro tu pi ti do ro go ro pi pa do da pi da go bu pa ku ro tu pa pi go la pa go tu go pi ku ti ro pi ro ku go ti ro pa ku la bi la bi ku ti ro la go do pa go pa la ku pi bi la go pi tu ku bu pa la tu go da ro ku go da ti ku bu bi do bi ku ti bu pi go pa go do da do tu da ku da go do pa pi ro pa la bi ku tu ro tu da pi go pa bi da ro la bi ti pa pi pa la go bu go bu ti ku pi bi ro pi la bi do ro pi pa ro da ro ti bi go la ku pi ti bu ro go bu go ro da ro do bu go pa tu la do pa pi go pa bi bu ro bi ti bi pa da pi ku bi ku bi pi la bu go do da pa bu pa tu pi do da go pi ti tu ro ku pi go ti pi do bi la ku tu bi la ti ro pi pa tu ti la bu la bi ku bu do la bi da do da ro da la do bu pi ro go tu bi pi go tu pa go la pi da ro pi bu ro ti bu bi tu ku bi ku da la pi ku pa go ro ku ro la pi pa ku la tu ro la do bu ti la do go ku bu la go da tu ro la bu go tu bi da ro do la bu pi go pi tu pi ti bi ku ro tu bi la go tu bu bi do bi pa tu ti go ku bu la ro bi tu ku la ti do tu bi tu pa ti pa ku do pi ku da go ro ti go ti ku la do bi da la ku la ku da do ti bu da la ti go bu ku ro do go ro pa bu bi da go da pi ti la bu bi pa ku do ku ti la do bi da la bi do la do go ti da bu pi bi tu ro bi ku ti da pi ku do la bi la go bu do bu tu da tu go pi bi tu ku pa go ku bu ro bi ro pi ku go ku bu da ro la ti bi do bi ro tu ro ku do la tu la ti pa ti go do ro bi tu do bi da pa tu ku do bu la do bu ku bu pa bu do ro la pa da ti ro da la go do ti la da ku ro ti bu go pa bi pa ku bu ku bu go da la ro pi ku do ti da do da la pa ku da pa ku ro ku go la ti pi bu bi ku ro ku da la bi ku ti pa tu ti pa da tu la ku ro tu pi pa ti pi tu ku pa ku go do bu ro pa do pa go la ku tu da ku bu do da pi go tu pa do ro ti do da tu la pa tu bi ti da tu do tu pi do la bu ti tu do da bi la bi pi bi bu tu go ro ku pa tu da la tu bi ku do ti da ti bi ku pa ti la go bu pa bu ti bu la ku ro da do ku do la bi bu pi do da la do bi ti ro la pi do tu bi go bu ku pi bi bu pa ti tu bu ti ku go tu bi pi tu ku la pi ro pi ku ti pa do da bu da ku ro la bi tu da tu bu ku bi do ro pi ro ti pi la do bu go bu ro bu pi da la go pa go pi ro bi ro do go ku bu ti tu pa bi bu do ku bu go do bi do ti ro bi pa pi pa bu da pa pi do ti tu pi pa tu do pi da pi la tu bi la ti la da go tu ku ti tu da do ti do bi pi do go ti do da ku do la pa go bi bu go ku bi ro go da tu do go ti do ku bi do pa la pa pi bu pa ku ro da ti la ku tu ku do tu do pa la ti go ti tu go ku bi do da pi go do ku pi pa go ku tu ti bu ro da pi la tu ro ti ku pi go da pa pi da bu tu pi da ro bu pa tu go pa bi tu la tu go ti go ku pa ro pi da la ti bu bi bu la bi ti da pa ti ku la pa do tu go ku ro ti la go do ti pa ro go la ro da pi la bi la ku da ku la ku bu bi tu do pa bi da pi tu bu pa pi go ro go ro pa bi go bu ti go bi da la da bu pa pi pa ro go pi bi ti bi bu pi pa bi pa bi da go pi bu pa go ti pi da bi ku pi da bi ku da do go bi ku ti da go tu pa la go bi go do da pi ku bu da la go la pi ku tu bi do bi la tu pi tu bi bu go ti tu do bi pa go do bu la ku go bi pa ro go la do ti da bu go tu do la do ku la pa bi go la ti pi tu ti ro bu da pa bi pa pi ti ro la pi ti ku pi pa do bu ku do bi pi do pi bi la ku do bi ro pa do pa da go la do ti ro bi pi ti bu pa tu do da pa go ku bi go pa ro pi bi da la ro la da pi do bu tu pa ku ro la da pa la bu go da tu do bi ku la do go bi ro bu la bu ro bu pa ti bi ro ti bi do la tu la da pa pi bu la do pi da pi da ro bi pa ro ku tu ro ku go da ku la ro go ti pa tu pi ro la da la ro do tu bu do ku la bu go bi pa da go tu pa bu la bi pa bi tu bi la bu ti la ku tu ku do tu do tu go bu ku da la pi bi ro da pi da la ro bu ti ku bu ti do tu ku pi bi do go pi ku bu da bi ti da ti la tu la tu bu ro bu ti bu ti ro pa bu ro bi do da ku bu da ro ku pa do ro pi la pa ro pa do da pa ro ti la bi ro la pa da tu ku ro go ti pa la pa ti pi la ro bu ti go bu do pi bi pa do ku bi ku bi pa tu bi pa da ro da ro go do pi ku ti tu la da ro tu pi ro tu ti bu bi tu ku bi pa do ti bi tu go bi ku ti do pi da la pa ku do la go pi do ti bu pi do da ro pa tu la pa tu ro pi da bi ti do da pi ku da bi ku tu pi ti pi go ku bi go la pa ti ku pi bu da ro da do la ti tu pi do bu go ti do go pa tu la go bu do da go bi ro ku pi ku do ro la da bu go tu ro pa da la go ro pi go ti pi ti ku tu bu pa pi do pi ti pa tu pi ti da do tu bi bu da pi ku ti da tu do pi go tu pi la bi do la da ku ro bu bi tu go ro ti do bi ro tu da tu pi la bi bu do da go ro pi bu ti bu do la bu ti bi ku ro ti pi do ku do ro tu da go ti da bi pi la go da do ti da la ti tu do pi ro ti la da pi ro bi ro la ti do ro pa do go la da bi ti bu ti ku pi la tu bi pi bi ro ti ku bu ku pa ku bu pa la pa tu go da ro go da tu go ku pi ku do da go ro bu bi la pa da bi ti bi ti da ti pi ti pa bi ro bi pa la tu bu bi ro pi ti pa go bi pi pa tu ro tu bu pi la da pa bu pi go pi la ku ro do go do da do go ku pa bu go ro do ti ku do pa da ro go bu la bu ku do ku go da tu go tu la tu ku ro pi bi la go da go pi ku pa da bi pa la go pi pa do ti pi bu do tu bu ti tu ti bu ro pa tu ku pa bi tu la bi pa pi pa pi ro bi la do da pa bi pa pi bu pi ku pi bu pa pi pa ku tu ro pi ro la do go bi ro bu do ku tu pa la tu da ro pa tu bu ro pa bi ti ro ti da bu tu da la ro bi bu ti pa bi do go pa pi bi la pi bu pi la ti bu pa do tu la do tu ro pa ti do tu do la pi bi bu pa bi la bu ro tu la tu pi ti tu ro pi bu bi da ro tu bu go ro go ti tu bi do bi ku pi bu pa ku pi go ro do la pi bu do ti bu da la bu ku ti ku pi pa bi da pi la ro do ti do bu bi bu ku bu go bi go tu ro ku do pi da bu da ku bu tu ti bi tu do da bu ro da ro do ti pa bu do go ku ro ku bi tu pi da ku ti da go ti ro bu da ti tu pi la bu da pi bi da go tu do go do pa go do bu ti pi ti la pa bi da la bi ti go pi tu da do la tu ku ti tu da bu pa tu ti tu do ku ti pa la bu go ro ti pa pi la tu bu pi ti pa ro pa tu bi bu pi do la ti bi tu ro ti bu do pi la da tu da bu pa go la ro da bu ro da tu pi ti ro ti tu bu pi da bu pi la do bi tu ti bu do ku bu pi bu la do bi ro pa ku pi go bi tu ro pa tu ti tu pa la bu tu ro bu go ro ku do ro pa ro pa ku da ku pa da tu pa la do bi la bu go bu ti tu ti ku bu ro bu tu do pa tu pa ku tu ro la bi do ro da go bu go pi ku pa do bu ti la ro bi bu do tu ku ti ro go pa go tu go pa pi da tu go tu bu bi tu ku go ti pi ti pi pa ro pi bu go bi la tu pa ro go ro pi do bi go ti ku ro bi ku tu go bu go do pa ro bi la da ku da pi pa do ku pi ti go la pa ti go la go do go la ti da do da ro bi bu bi pa bi pi tu ti da ku ro go do ti pi ku bi tu bi da pa do tu ku pa da bu pa do ro go da tu do go ti tu bi tu ku tu ro bu da ro da pi bi ti go ti da la ku go bi bu la da bu pa ro ku la pi do da pi do bu ro ti bi pi ti tu ro pa ku do ro da do la bi pi ro da pa go ti pa go ti pi ku ro da bi da ro pi go do ro ti do pa pi la da pa bu do ku tu do da go ti bu ro go pa ti bi da la go ku pi bu tu bi pi ti go ti bi pi ku bu do ku pa go pi ku da ku go la ro da ku do bi go da go da ti bu tu da pa bi tu ku bu ti go tu pi la do tu bi ku la pi bu go tu go pa bu ti ku tu ku ro ku bu da pa ku go pa da bi da ti ro da tu da ti pi tu bu ti ro ti pa bu tu ti pa ro ti ku do

A.3.3. Consonant phoneme order Random Stream 1

 # P = labial
 # K = dorsal
 # T = coronal

K K P P T T P T T T T P P T K P T T T P T T T T P T K P T T T T P K P P T T T P T T T K
T P P T T P T K P P K T T P P K T P K T K P K T T P T K K T T P K T P T P K T T T K T
P K P T K P P T K P T K P P T T K T T K K T T K P P T P K T P P K P K T T T T T K T K
T P P T P T P K T T T T P K P P T T T P T P P P T K P K P T K P P T P T P T T P P T T T T
P K T K P T P T K P K T T T T P K P T T T P P K P P P T P T P P T P K P K P P T P K T T
P P P T P T T K P T T T K P K T P T P T K T P T T T P P T T T P T P K P T T P T T T T T
T T P P T K T P P K T P K T P T T P P T T P P T K P K T T P K P K T K T T P P K T T T T
T P T T T K K P T K T T T T P K T P T T T T P P K P T P T P K T T P T K T P P T P P T T
K K P T T P T K T T T T P T P T P K T P K T K T T K T K T T P T T K T K T T T P T T T
K P K T T K T P P P T K T P T T P P P K T K T T T P T T P T T T K T T P P P T T P K T T
P K T T P T K P T P T T T K P P T K P K K P T P T P K K K P T T T T P T P T T T K T T T
T T P T K T T P T T P T P T K T P T T P K P P P T T T P T T T T T K T T T T K T T P K P
P P K P K P K T T T P K T T T T T T P K T P K T K K T T P P P K T K T T P K T P T T P
T T T K T T P P T P T K P K K T P T P T P K T K T T K P T T P K T P T T T T T T T T P T P
T T T T T P T T P T T T T P T P P P P T K T K P T T T T P K T T T T P K P T T K P P P T

P T K T T T K T T P P P T T T T P T T T P T T P K P K P P P P T T P T K K T P P T K T P
T P K T P T T P T K T T P T T T P K P T T P T T P T T P K P T P P T T K P K P T P T T K
K P T T P P P T K P K T P T T T P P P P P T P P T T T P P T T P T P T T P T T T T K T K T
T T T T T P P T K T T T K T T P K P P K K P T K T T T K T T K P T P T P P P P K T T T T
K T K T T T P T T K T T K K P T T P K T K P P K K T T P T T P T T T T K P K T P P T P
T P T T P P T K P P T T T K T K K P T P T T T P P P T P T T P T K T P T T K K T T T K T
T P T K T T T P T P T K T K T K P P T T P P T P T P P P K T K T P P K P T K P T T T P P
P P T K P P T P P P P P P P T K P P P K T P T P K P T P K T T K P K T T K T P T K P K T
T P K P T T K T P K T P T P T T P T P P K T T T P P K T P T K K P P T K T T T T P K T T
T T K T P P K T T P T T T P T P P P P T T T P T K P P T P K T P P T P P T K T P T P T P T
K T T T T P P T P P T T T P K K P K P T P P T T T T T T P T P T P K T T T P T P K T T T P
K T T K P T P T P T P P T P T T P T T T T T P P P T T P T P T T P P T K T T K K T K T T
K T P T P T T T T T T T P T K T P K P P T K T P P T P P P T P T P T T K T K T T T T K P
K T T P P T T P T T T P T K P T T T K P P T K P K P T P T T T T T T T P T P T P T T P P
T P T T K P T T K P T T P T P T P T T P T T T P T T P T T K T K T P T P T P T T P T K P
T P P P T K P K P P T P P T T T T K T P K T T T T T T P T T T P P T K P P T T P T K P K
T T P T T P K T T K P T T P P T T T P T T P T T P T P T T T P K T P K T P T P K K P K T
P T K P P T T T T T T T P T P K T T K P T T K P T T K P T K P K T T T T P K T T P T T K
T P K T P T K T P P P T P T P T P T T T T P P T P K T T T T P K T P T P T T T K T P P T
K T T T P T T T T P T P P T T K T P P T P T T P T P K T T P T K T T T T K T T P P T K T
T T T T T T T P T T T T P T P T T T T T T P T K T T P T P T K P T T P P P T T K P K P K P
P T P T K T T K T T K K P K T T K T P P T P T P T P T T T P T P P T P P T T P P T P T P
K P P P T T T P P T T P P P K P T K T T K T T T K K P P K T T T K T P T T K P T P K T K
K T T K T T T K T P P T K T K P K P T P P T K P P T T P P T T P T T T P T P T K P P T T
P P P P P T P T T T P P P P P P K P P P P P K T T P T T T K P T P T K T P T T T T P T P T
P P T T T T P T T T T P P T P P T K P P P T P P P T T P P T T T T T T P T T T T T P P P P
P T P T T T T P T T T P P P T T T P K T K T T P T P K P P P K P K T T T P P T T P T T P
K T K P P P T P T T T T T P P P K P K P K T T K T P T P T K P T T P T T T P T T T T T P
P T K K T K P T P T K T T K T T P T T T P T P T P P T K T T K T P K T P T P T T P P T T
P T K P T T T T T K T T T P P T T T T K T P T P K T T P P T T P P T P T P T P P P T T T P
T T T P T P T T T T P P K T T T P T T T P T T T T P P T P P T T P T T P T K P P P T T P T
P K P K P T T P T T T P T P T T P K T K T T P T P K T K P T T P T T P T P K P T T T K P
T P T T P T P K T T T P T T T K P K P K P T P T T T P P T T K T T K P K T K P P T T K T
P P T K K T P T P P T P P K P T T P T K T P T P K T K T P K T K P K T P T P T T K T P P
T K P T K T P T K T K T K T T T T T T P P P P P T T T K T K T T P K P T P T P T T K P
T P P T T K T T T K T T P T K T T P T T T P P T K T T T K K P P T T P P T K T P T T P T
P T T P P T T T P K T T T T T P P T T P K T P K T P K T T P T T P K T T T T P P T T P P
T K T T T K T P T K P T P T T K K P P T P P T K T P P K P T K P K P K T K K T T T K T
P K T K T T P T T P P T K P T K T P T T T P K T P P K T K P P T K T K T K P T P K K P
T P T T T T T T T T P T P T T T P P T T P T T K T

A.4. Random Stream 2

A.4.1. Random Stream 2 as provided by B&P2017

12 7 3 6 4 11 9 7 4 10 1 5 4 12 1 8 12 1 5 12 3 5 6 7 9 8 4 7 10 6 3 11 1 11 5 3 5 3 11 1 3 1 7
10 7 3 4 6 5 7 11 9 11 5 10 4 12 6 10 9 5 11 2 11 8 3 4 5 11 9 10 1 2 7 2 12 9 3 6 10 2 11 5 9
12 8 11 2 5 8 2 5 11 7 11 9 8 2 6 7 1 5 10 4 8 1 2 8 6 8 3 10 11 6 10 6 1 9 2 4 1 12 8 12 5 4 5
10 12 4 12 4 3 4 2 10 7 9 6 2 8 10 5 11 5 2 5 3 4 9 2 4 12 7 3 4 2 5 11 12 3 7 12 5 9 4 6 1 3 7
11 8 10 5 11 6 1 6 2 12 6 11 3 4 5 12 1 10 3 1 9 5 8 3 9 11 3 1 5 1 12 2 5 3 12 6 5 10 4 12 7 10
12 8 12 1 5 10 3 12 9 6 9 2 3 9 4 2 1 3 8 4 6 3 8 11 5 7 3 6 9 11 1 7 2 5 7 11 7 4 12 4 11 10 9
11 7 12 7 9 2 12 1 2 12 3 8 3 9 11 2 6 10 4 7 4 7 2 3 5 4 10 4 3 11 3 11 2 1 8 9 11 4 3 12 3 1 11
12 3 8 6 11 10 6 1 3 1 6 12 1 11 9 8 4 7 5 9 1 9 10 9 4 8 10 12 4 11 12 10 8 2 8 9 12 10 9 10 4
5 4 11 9 1 9 2 9 11 6 12 10 4 11 7 3 2 12 8 1 5 1 9 8 4 6 10 3 5 11 2 11 2 6 7 11 9 1 5 4 10 12 7
10 11 10 11 5 10 7 6 5 12 1 8 11 1 5 2 4 8 7 10 8 1 6 5 1 2 9 2 12 1 10 6 9 7 1 8 5 6 9 10 1 4 3
9 8 10 6 2 8 12 5 7 9 8 7 9 6 7 3 1 9 4 9 11 8 12 1 4 8 10 6 11 3 2 7 2 1 12 5 2 8 7 10 4 2 3 10 1
2 4 12 1 7 10 11 4 9 3 6 8 9 1 5 7 3 8 7 1 11 7 6 2 7 12 8 10 2 11 4 7 5 8 3 1 9 4 5 9 7 12 3 11 3
2 6 1 2 8 12 4 9 4 3 8 3 1 7 3 6 2 1 11 12 6 4 6 4 12 10 4 1 9 8 4 2 12 3 2 11 2 5 6 7 5 11 7 11 6
10 12 5 1 7 9 1 10 3 11 8 6 5 3 2 11 3 2 1 10 1 10 2 6 9 6 4 12 1 9 11 1 4 7 2 10 1 6 4 1 9 10 9
7 4 3 6 9 8 12 1 10 2 9 7 12 1 5 7 2 11 1 10 12 5 4 8 6 2 11 6 9 7 6 3 7 9 6 7 5 11 9 12 4 7 8 6 2
9 7 2 4 8 2 7 12 10 3 11 5 3 11 4 3 11 5 12 5 8 9 3 4 7 6 3 10 1 2 9 11 1 12 4 7 5 9 4 12 1 11 8
5 4 3 2 6 8 2 11 2 8 10 5 4 3 2 5 6 2 3 5 6 11 5 7 11 4 8 10 9 2 9 11 9 12 3 5 2 1 9 4 8 6 7 5 8
12 4 12 6 7 10 11 4 10 11 6 5 3 10 9 11 7 8 7 5 6 9 12 11 3 5 8 2 5 7 6 2 7 3 9 11 7 11 3 9 1 6 7
11 4 8 2 1 5 10 2 11 12 2 11 12 1 10 8 12 8 5 7 8 4 1 10 7 11 10 1 2 4 3 6 2 5 12 11 1 6 3 9 6 7
12 10 12 1 7 6 7 8 4 12 3 11 6 11 2 12 8 9 2 5 6 3 9 4 9 2 12 5 3 9 5 4 9 6 11 3 6 7 12 9 12 4 3
7 10 9 1 11 5 10 2 12 10 1 5 2 10 4 5 1 7 3 10 2 5 11 4 11 7 10 5 10 12 8 2 11 7 12 10 11 3 4 6
11 9 11 10 11 10 6 11 8 9 6 8 4 1 4 8 3 9 3 10 11 12 2 10 6 1 11 1 3 5 11 4 10 1 2 10 7 9 2 8 1
4 3 9 10 12 8 9 6 9 3 9 1 5 6 4 7 12 10 8 2 10 5 2 11 9 6 2 11 3 4 7 11 9 5 10 2 11 6 9 2 7 1 12
5 7 2 5 4 1 11 9 1 5 8 7 4 10 1 10 3 8 7 8 10 11 3 8 4 9 10 6 10 6 2 4 11 1 7 9 3 8 7 6 8 5 11 10
7 4 12 3 11 7 4 11 4 3 6 7 6 2 7 5 9 6 11 7 4 12 1 11 3 11 2 3 9 8 11 1 6 7 3 1 5 6 12 7 12 1 5 8
9 10 6 10 4 9 6 11 1 3 5 9 7 3 5 12 4 9 1 6 2 7 10 6 8 2 6 8 3 11 7 12 7 10 5 12 10 8 4 5 1 3 10
7 4 9 3 1 3 4 7 12 8 1 8 7 1 2 7 11 6 4 5 8 12 9 7 9 3 4 8 11 1 3 4 6 2 1 2 11 2 1 8 2 1 2 3 12 8
11 10 6 8 6 4 12 4 11 10 5 7 8 5 3 6 9 3 5 6 4 8 5 7 4 11 9 3 9 10 11 10 5 2 10 12 3 7 5 6 4 8 10
1 7 11 12 1 12 10 3 1 11 8 4 7 2 4 12 4 3 8 6 12 11 4 9 1 9 6 9 6 7 12 2 9 5 6 4 7 6 9 8 12 2 8 2
1 9 12 6 4 11 1 2 5 7 5 10 6 1 8 10 12 7 2 12 3 2 1 9 4 1 8 3 4 6 8 5 12 4 8 10 12 3 5 2 6 8 7 2 4
2 9 12 8 6 9 10 11 1 5 10 2 7 10 6 12 7 10 5 2 3 6 11 2 9 2 1 10 11 3 10 2 8 1 12 11 6 1 6 11 10
7 3 11 9 10 12 3 4 7 9 8 2 9 7 2 4 6 3 1 2 8 12 11 10 11 10 9 3 9 5 6 9 6 2 6 1 9 3 2 5 12 9 6 5
12 6 7 12 9 1 3 12 7 5 3 5 9 1 3 6 10 11 7 6 7 3 4 8 1 2 7 12 5 11 4 9 12 3 12 2 7 4 2 12 6 9 3 9
3 1 8 12 2 3 4 3 12 9 4 10 12 2 7 5 6 11 12 10 4 10 5 3 5 8 3 9 12 1 11 1 5 7 8 6 1 7 6 2 4 8 11
7 6 9 11 8 6 7 3 10 3 2 6 7 12 11 3 12 8 10 3 10 3 11 5 10 5 6 5 6 5 8 2 8 9 3 1 3 8 10 5 8 5 6 4
11 12 5 12 11 3 11 10 12 10 11 3 11 2 1 6 4 6 5 8 7 2 6 10 9 2 3 5 1 10 9 2 5 8 1 12 7 11 12 7 1
8 5 8 6 8 12 8 12 10 3 7 2 1 11 10 2 10 11 8 5 3 6 4 9 11 10 2 11 1 7 4 7 12 7 9 6 3 7 10 7 1 4
12 9 12 2 4 3 10 8 5 3 5 9 3 8 5 6 2 6 8 10 8 2 4 10 7 10 6 5 2 12 11 12 5 12 6 8 3 10 2 5 2 4 7
11 4 10 4 6 7 10 7 6 3 7 10 8 1 6 8 3 10 3 5 1 6 4 1 9 10 11 9 5 2 1 5 2 12 4 11 2 7 10 6 1 10 6
3 8 3 1 12 2 6 1 4 9 4 1 10 5 4 1 10 3 10 8 6 5 12 8 6 10 8 4 5 4 12 1 12 2 1 9 5 9 3 5 1 2 12 3 2
8 10 2 1 12 9 4 10 4 3 12 1 5 3 9 2 4 11 4 6 11 12 9 8 11 8 12 5 9 2 11 6 4 8 2 10 5 12 5 2 4 12
2 7 11 12 11 12 9 1 11 7 5 11 10 8 5 10 2 11 9 8 3 5 7 3 9 10 4 9 6 8 11 5 9 11 10 12 8 12 11 8
1 12 11 3 9 5 4 12 10 1 5 9 11 12 1 3 6 3 4 10 8 12 4 2 8 6 3 11 5 4 10 7 1 8 6 8 1 12 9 2 1 4 7
4 7 8 1 9 5 6 3 10 12 6 4 12 8 1 4 2 9 10 12 7 9 4 9 1 11 7 4 9 5 11 1 5 8 9 10 8 4 5 1 7 2 1 4 8
4 10 6 3 7 3 12 2 3 6 12 1 6 1 7 8 12 4 6 7 4 8 6 5 9 3 1 2 7 6 5 6 7 8 11 4 2 9 10 3 10 6 12 3 12
3 6 11 8 2 4 6 3 7 10 2 6 2 10 4 12 9 1 8 9 3 1 7 5 9 8 11 12 10 11 6 12 6 5 8 2 8 4 11 2 4 6 2 8
3 1 12 3 8 6 2 1 12 4 1 5 6 9 12 6 9 7 11 8 9 6 12 2 11 7 12 2 4 10 2 7 8 2 11 5 2 8 11 3 4 12 10
8 9 4 1 8 10 8 10 3 7 12 8 3 10 2 3 1 11 4 8 3 8 10 2 4 2 6 12 10 7 5 11 10 3 8 7 1 7 12 1 4 12 5

8 5 4 2 3 7 8 11 8 9 11 8 9 5 1 9 5 12 7 10 4 5 10 9 8 10 2 5 4 8 1 8 11 3 1 7 9 8 1 2 6 5 10 7 12 5 1 9 1 11 8 6 5 7 10 7 2 10 3 11 2 4 9 8 1 7 1 2 12 9 11 4 9 12 8 12 9 12 9 6 7 3 5 10 3 12 9 11 3 5 3 10 11 6 5 2 9 11 8 11 10 4 11 6 9 6 1 3 1 10 8 9 12 8 10 12 1 8 3 10 5 4 10 12 7 1 8 4 11 12 8 9 4 6 9 7 1 11 3 6 1 11 7 5 9 11 5 6 7 1 7 5 7 9 5 1 5 7 6 7 2 3 8 12 10 3 12 7 5 2 6 10 8 2 6 10 6 12 2 4 12 3 12 3 7 12 9 8 7 6 7 10 12 7 9 4 8 12 7 1 8 7 4 12 4 6 4 5 6 3 11 7 5 1 10 8 9 8 6 11 9 4 2 6 1 2 5 10 8 4 1 3 7 5 9 5 3 4 2 5 12 4 8 2 10 11 3 4 5 9 4 2 9 3 5 8 2 1 6 2 4 5 11 12 6 7 2 1 5 9 5 1 10 3 6 7 9 4 7 11 3 9 1 5 6 10 9 10 7 1 5 1 4 6 7 8 10 1 9 10 3 11 5 7 1 8 10 2 1 2 6 5 2 3 2 4 7 11 7 5 6 9 12 8 1 2 1 4

A.4.2. Random Stream 2 in syllables

to na ke mi lu ra po na lu pu fi me lu to fi nu to fi me to ke me mi na po nu lu na pu mi ke ra fi ra me ke me ke ra fi ke fi na pu na ke lu mi me na ra po ra me pu lu to mi pu po me ra gi ra nu ke lu me ra po pu fi gi na gi to po ke mi pu gi ra me po to nu ra gi me nu gi me ra na ra po nu gi mi na fi me pu lu nu fi gi nu mi nu ke pu ra mi pu mi fi po gi lu fi to nu to me lu me pu to lu to lu ke lu gi pu na po mi gi nu pu me ra me gi me ke lu po gi lu to na ke lu gi me ra to ke na to me po lu mi fi ke na ra nu pu me ra mi fi mi gi to mi ra ke lu me to fi pu ke fi po me nu ke po ra ke fi me fi to gi me ke to mi me pu lu to na pu to nu to fi me pu ke to po mi po gi ke po lu gi fi ke nu lu mi ke nu ra me na ke mi po ra fi na gi me na ra na lu to lu ra pu po ra na to na po gi to fi gi to ke nu ke po ra gi mi pu lu na lu na gi ke me lu pu lu ke ra ke ra gi fi nu po ra lu ke to ke fi ra to ke nu mi ra pu mi fi ke fi mi to fi ra po nu lu na me po fi po pu po lu nu pu to lu ra to pu nu gi nu po to pu po pu lu me lu ra po fi po gi po ra mi to pu lu ra na ke gi to nu fi me fi po nu lu mi pu ke me ra gi ra gi mi na ra po fi me lu pu to na pu ra pu ra me pu na mi me to fi nu ra fi me gi lu nu na pu nu fi mi me fi gi po gi to fi pu mi po na fi nu me mi po pu fi lu ke po nu pu mi gi nu to me na po nu na po mi na ke fi po lu po ra nu to fi lu nu pu mi ra ke gi na gi fi to me gi nu na pu lu gi ke pu fi gi lu to fi na pu ra lu po ke mi nu po fi me na ke nu na fi ra na mi gi na to nu pu gi ra lu na me nu ke fi po lu me po na to ke ra ke gi mi fi gi nu to lu po lu ke nu ke fi na ke mi gi fi ra to mi lu mi lu to pu lu fi po nu lu gi to ke gi ra gi me mi na me ra na ra mi pu to me fi na po fi pu ke ra nu mi me ke gi ra ke gi fi pu fi pu gi mi po mi lu to fi po ra fi lu na gi pu fi mi lu fi po pu po na lu ke mi po nu to fi pu gi po na to fi me na gi ra fi pu to me lu nu mi gi ra mi po na mi ke na po mi na me ra po to lu na nu mi gi po na gi lu nu gi na to pu ke ra me ke ra lu ke ra me to me nu po ke lu na mi ke pu fi gi po ra fi to lu na me po lu to fi ra nu me lu ke gi mi nu gi ra gi nu pu me lu ke gi me mi gi ke me mi ra me na ra lu nu pu po gi po ra po to ke me gi fi po lu nu mi na me nu to lu to mi na pu ra lu pu ra mi me ke pu po ra na nu na me mi po to ra ke me nu gi me na mi gi na ke po ra na ra ke po fi mi na ra lu nu gi fi me pu gi ra to gi ra to fi pu nu to nu me na nu lu fi pu na ra pu fi gi lu ke mi gi me to ra fi mi ke po mi na to pu to fi na mi na nu lu to ke ra mi ra gi to nu po gi me mi ke po lu po gi to me ke po me lu po mi ra ke mi na to po to lu ke na pu po fi ra me pu gi to pu fi me gi pu lu me fi na ke pu gi me ra lu ra na pu me pu to nu gi ra na to pu ra ke lu mi ra po ra pu ra pu mi ra nu po mi nu lu fi lu nu ke po ke pu ra to gi pu mi fi ra fi ke me ra lu pu fi gi pu na po gi nu fi lu ke po pu to nu po mi po ke po fi me mi lu na to pu nu gi pu me gi ra po mi gi ra ke lu na ra po me pu gi ra mi po gi na fi to me na gi me lu fi ra po fi me nu na lu pu fi pu ke nu na nu pu ra ke nu lu po pu mi pu mi gi lu ra fi na po ke nu na mi nu me ra pu na lu to ke ra na lu ra lu ke mi na mi gi na me po mi ra na lu to fi ra ke ra gi ke po nu ra fi mi na ke fi me mi to na to fi me nu po pu mi pu lu po mi ra fi ke me po na ke me to lu po fi mi gi na pu mi nu gi mi nu ke ra na to na pu me to pu nu lu me fi ke pu na lu po ke fi ke lu na to nu fi nu na fi gi na ra mi lu me nu to po na po ke lu nu ra fi ke lu mi gi fi gi ra gi fi nu gi fi gi ke to nu ra pu mi nu mi lu to lu ra pu me na nu me ke mi po ke me mi lu nu me na lu ra po ke po pu ra pu me gi pu to ke na me mi lu nu pu fi na ra to fi to pu ke fi ra nu lu na gi lu to lu ke nu mi to ra lu po fi po mi po mi na to gi po me mi lu na mi po nu to gi nu gi fi po to mi lu ra fi gi me na me pu mi fi nu pu to na gi to ke gi fi po lu fi nu ke lu mi nu me to lu nu pu to ke me gi mi nu na gi lu gi po to nu mi po pu ra fi me pu gi na pu mi to na pu me gi ke mi ra gi po gi fi pu ra ke pu gi nu fi to ra mi fi mi ra pu na ke ra po pu to ke lu na po nu gi po na gi lu mi ke fi gi nu to ra pu ra pu po ke po me mi po mi gi mi fi po ke gi me to po mi me to mi na to po fi ke to na me ke me po fi ke mi pu ra na mi na ke lu nu fi gi na to me ra lu po to ke to gi na lu gi to mi po ke po ke fi nu to gi ke lu ke to po lu pu to gi na me mi ra to pu lu pu me ke me nu ke po to fi ra fi me na nu mi fi na mi gi lu nu ra na mi po ra nu mi na ke pu ke gi mi na to ra ke to nu pu ke pu ke ra me pu me mi me mi me nu gi nu po ke fi ke nu pu me nu me mi lu ra to me to ra ke ra pu to pu ra ke ra gi fi mi lu mi me nu na gi mi pu po gi ke me fi pu po gi me nu fi to na ra to na fi nu me nu mi nu to nu to pu ke na gi fi ra pu gi pu ra nu me ke mi lu po ra pu gi ra fi na lu na to na po mi ke na pu na fi lu to po to gi lu ke pu nu me ke me po ke nu me mi gi mi nu pu nu gi lu pu na pu mi me gi to ra to me to mi nu ke pu gi me gi lu na ra lu pu lu mi na pu na mi ke na pu nu fi mi nu ke pu ke me fi mi lu fi po pu ra po me gi fi me gi to lu ra gi na pu mi fi pu mi ke nu ke fi to gi mi fi lu po lu fi pu me lu fi pu ke pu nu mi me to nu mi pu nu lu me lu to fi to gi fi po me po ke me fi gi to ke gi nu pu gi fi to po lu pu lu ke to fi me ke po gi lu ra lu mi ra to po nu ra nu to me po gi ra mi lu nu gi pu me to me gi lu to gi na ra to ra to po fi ra na me ra pu nu me pu gi ra po nu ke me na ke po pu lu po mi nu ra me po ra pu to nu to ra nu fi to ra ke po me lu to pu fi me po ra to fi ke mi ke lu pu nu to lu gi nu mi ke ra me lu pu na fi nu mi nu fi to po gi fi lu na lu na nu fi po me mi ke pu to mi lu to nu fi lu gi po pu to na po lu po fi ra na lu po me ra fi me nu po pu nu lu me fi na gi fi lu nu lu pu mi ke na ke to gi ke mi to fi mi fi na nu to lu mi na lu nu mi me po ke fi gi na mi me mi na nu ra lu gi po pu ke pu mi to ke to ke mi ra nu gi lu mi ke na pu gi mi gi pu lu to po fi nu po ke fi na me po nu ra to pu ra mi to mi me nu gi nu lu ra gi lu mi gi nu ke fi to ke nu mi gi fi to lu fi me mi po to mi po na ra nu po mi to gi ra na to gi lu pu gi na nu gi ra me gi nu ra ke lu to pu nu po lu fi nu pu nu pu ke na to nu ke pu gi ke fi ra lu nu ke nu pu gi lu gi mi to pu na me ra pu ke nu na fi na to fi lu to me nu me lu gi ke na nu ra nu po ra nu po me fi po me to na pu lu me pu po nu pu gi me lu nu fi nu ra ke fi na po nu fi gi mi me pu na to me fi po fi ra nu mi me na pu na gi pu ke ra gi lu po nu fi na fi gi to po ra lu po to nu to po to po mi na ke me pu ke to po ra ke me ke pu ra mi me gi po ra nu ra pu lu ra mi po mi fi ke fi pu nu po to nu pu to fi nu ke pu me lu pu to na fi nu lu ra to nu po lu mi po na fi ra ke mi fi ra na me po ra me mi na fi na me na po me fi me na mi na gi ke nu to pu ke to na me gi mi pu nu gi mi pu mi to gi lu to ke to ke na to po nu na mi na pu to na po lu nu to na fi nu na lu to lu mi lu me mi ke ra na me fi pu nu po nu mi ra po lu gi mi fi gi me pu nu lu fi ke na me po me ke lu gi me to lu nu gi pu ra ke lu me po lu gi po ke me nu gi fi mi gi lu me ra to mi na gi fi me po me fi pu ke mi na po lu na ra ke po fi me mi pu po pu na fi me fi lu mi na nu pu fi po pu ke ra me na fi nu pu gi fi gi mi me gi ke gi lu na ra na me mi po to nu fi gi fi lu

A.4.3. Consonant phoneme order Random Stream 2

 # P = labial
 # K = dorsal
 # T = coronal

T T K P T T P T T P P P T T P T T P P T K P P T P T T T P P K T P T P K P K T P K P T P
T K T P P T T P T P P T T P P P P T K T T K T P T P P P K T K T P K P P K T P P T T T K
P T K P T T T P T K P T P P P T T P K T P T K P T P P P P P K T P T T T P T P P T T T T K
T K P T P P K T P P T P K P K T P K T T T K T K P T T K T T P P T P P K T T T P P T P P
P K T P T K T P T P P K P P P T K P T K P P P T K P K T P P P T T T P T T T P P P K T P
P P K K P T K P K T T P K T T P T K P P T P T K P T T T T T T T P P T T T T P K T P K T
K T K P T K P P T T T T K K P T P T K T K T K P T P T T K T K P T T K T P T P P P K P
P T P T P T T T P P P P P P T T P T T T T P T K T P T P P P T P T T P P P K P T P T P T T
T K K T T P P P P T T P P K P T K T K P T T P P P T P T T P T P T P P T P P T P T T P P K
T T T P T P P P P K P K T P P P P T P T P P P P T K P T P P K T T P T P T T P P T K P P
T P T T T P T T P P T K K T K P T P K T T P T K K P P K T T P T P T T P K P T P P P T K
T T P T T P K T T T P K T T T P T K P P T P P T T K T K K P P K T T T P T K T K P T K P

K P T T P T P T T P T P P T T K T K K T K P P T P T T T P P T P P T P P P K T T P P K K
T K K P P P P K P P P T T P P T P T T K P P P T P P P P T T K P P T T P P K P T T P P T K
T P P T P T T P K T P P T P K T P P T P T P T T T T P K P T K T T K T T P K T P K T T K
T P T P T P K T T P K P P K P T P T T T P P T T P T T P T K K P T K T K T P P T K K P P
K K P P T P T T T T P P K P T P T K P K P P T T P T P T T T T P T P T T P T P P K P P T
T T T P P P T T K P T K P T P K T K P T T T K P P P T T T T K P P P K T T K T T P P T T
T P T T T P P T T P P K T K P K P T T P P K P P T T P T P T P T T T T K T P T K T T P K
P P K P T P K T P K P P T P P T K P T T P T T K T P P P T P P K T P P P K P T P P T K P
K P T T T T P P P T T K T T T P T K T P T P T P T P P T T P P T T P T T K P K P T T K P
P P T P K P T T P P K P T P K T P T K P P T T P P P K P P P P T T T P T K P P K T P P K
T K T T T P P P K T P P K T P T P T K P T P T P P P T T T P P P K T T T P T K T T P P P P
P K T T P T P K T T P T P T P T T T K T T T T T K P T P K T P P P T T T T P T K T K K P
T T P P T K P P P T T T P P T P P P P T P P T P K P P T K P T T P P P K T P P T K P T K T
T T T P P T P T T P P K P T T P K P K T T T T P T T P K T T P T P T T P T P K T T T P K
T P K P K T K P T K P K K T T T P P T P T T T T P P T T P K P P K P P T T P T T T P K P
P T P P K P T K T P P T T P P T T T P T P K P T T T T K T T T K T P T T T P P P P P P T T
K P P P T T P P T T K T K P P T P T T P K P T P P P P T P T T K T K K P P T P T K T P T
P T T T P T K P K P T T K T K P T T P P P T P P P K T P P T T P P K K P T K P K P P T K
P K T P T T P P P T P T K T P P T K T T P T K P T K T P K P K T T T P T P P K P P P P P
K P P P K K P T P P P T P T T P P K T T P K P P P K P P T T P T K T T P K T T P T T P T
K T K T T K T P P K P K P T T K K T K T P T P T K T P P T T P T P P K P T K P T P T P P
T T P P T P K T T T T P P T T P T K P K K P T T T K T T P K P K T P P P P P P P T K T P
K P K T P P T P P T T T P T T K T P T P T K T K P P T P P T T K P P P K K P P P P K P T
P T T T T T P T P T P T T T T P K T K P T P K P T T P K P T P T P K T P T T T T T P P K
T P T P T T P T K T K P T P K P P K T P P K P T P T K T P T P P P K T T T P T P T K P K
P K T T T T P T P T P T P K T P T P P T K P K P P P T P P P T P P K P P K T T T K T P P P
P P K T K P T K P P T P T P P P T P P K P T P P T T P P T T P T T P T K P P P P K P P K T
K K T P K P T P T P T K T P P K P K T T T P T T P T T T T P P K T P T T K P P T P K T T
K T T T T T P P T T P T P T P P K T P T K P T K P P T P P T T P P T P T T T T T T P T T K
P P T T P P P P T T P K P K T P T T T K T P K T P T P T P T P T P K P T T T T T T P P P
P K P T P T T T P T K P P T T P T P P T T T P P T P P T P P T T P P T K P T T T P P K T K
T K K P T P P P T T T T P T T T P P P K P K T P P P T T T T K P P K P P T K T K P T T K
T P K T P K P K P T T P P T P K P T P P T T T P T P T P P T K T T T K T P K T K P T K T
P K P T T P P P P T P P T T T P P T K T T T K T P K T T K T P K T T K T T P T P T P T P
T P K T T T K P K K P T T T K T P K T K P T P T P T P K T T P T T P T T P T P T K K T T
T T P T T P P P P P T T P T P P P T P K P T T P T T K P T P T P K P P P T T P P P P T T P
P T P T K P K T K T P T P T P K T P T T P T T T P T P P T K P P K T P T K P K P T P P K
P T T T P T T P P P P K P P T P T T P T P T K P P T P T T P T T T T T P T P P T P T K P P
T T P P T P P T P T P T P P P P T P T K K T T P K T T P K P P T K P P P T K T T K T K T
T P T T P T P T T P T T T T P T T T T T P T P P K T T P P P T P T P T P T K P P K P P T T
P K T P P P K T K P T T T K P T K T P P T K P K P T K P P K T P T T P T K P P P P P P K
P T P T T T K P P P P P P P T P P P T P T T P P P P K T P T P T P K P K P P K K K T T T T
P P P T T P K P T

Appendix B. Iterative model report for the TP and OCP models.

Iterative model report for the TP model. Each line reports the assessment of improved model fit after adding a single predictor.

Nr.	-2 LL	nr. of parameters	p model fit (chisquare distribution)	model comparison	predictor added	action
Model 0	47425,3	3			Empty model with Participant as random intercept	-
Model 1	41873	4	< .0001	better	Word random intercept	keep
Model 2	41399,9	5	< .0001	better	Condition Fixed	keep
Model 3	40448,5	7	< .0001	better	Block Fixed	keep
Model 4	39147	9	< .0001	better	Condition Block Interaction	keep
Model 5	39147	10	0,9638	not better	Stream Order of Presentation Fixed	remove

Iterative model report for the OCP model. Each line reports the assessment of improved model fit after adding a single predictor.

Nr.	-2 LL	nr. of parameters	p model fit (chisquare distribution)	model comparison	predictor added	action
Model 0	47425,3	3			Empty model with Participant as random intercept	-
Model 1	41873	4	< .0001	better	Word random intercept	keep
Model 2	41813,7	5	< .0001	better	Centered OCP Fixed	keep
Model 3	40700,9	7	< .0001	better	Block Fixed	keep
Model 4	39013,7	9	< .0001	better	Centered OCP Block Interaction	keep
Model 5	39013,7	10	0,9497	not better	Stream Order of Presentation Fixed	remove

Appendix C. Inter-Trial Coherence as a function of frequency per block

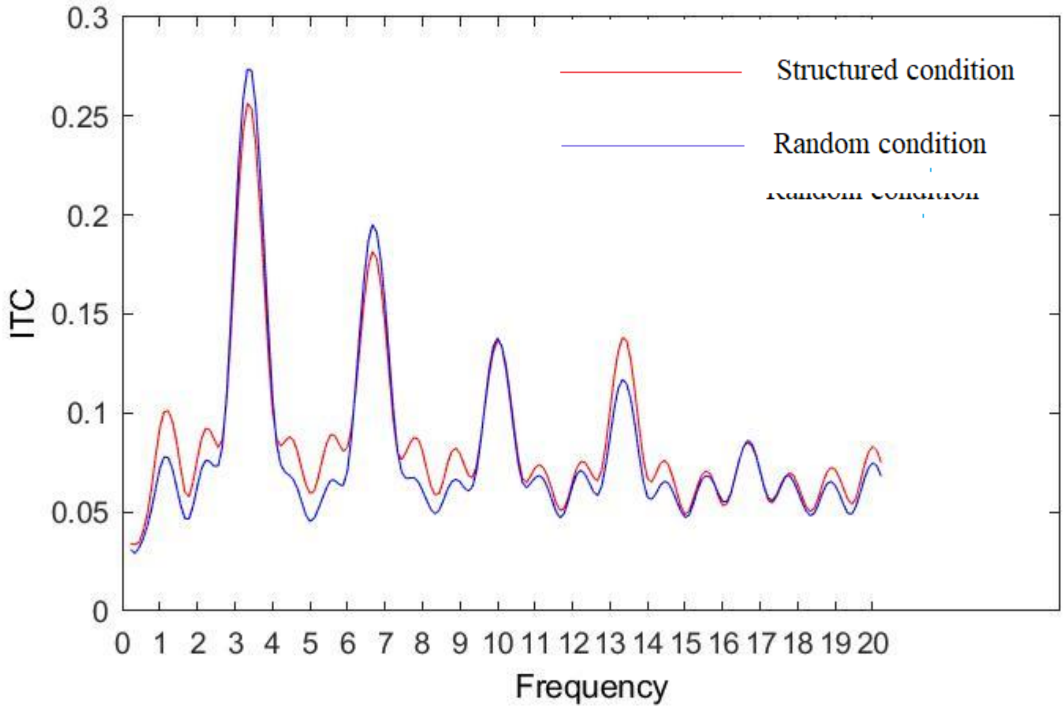


Figure C1. Inter-Trial Coherence (ITC) as a function of frequency during block 1 for all measured frequencies (0.1-20 Hz).

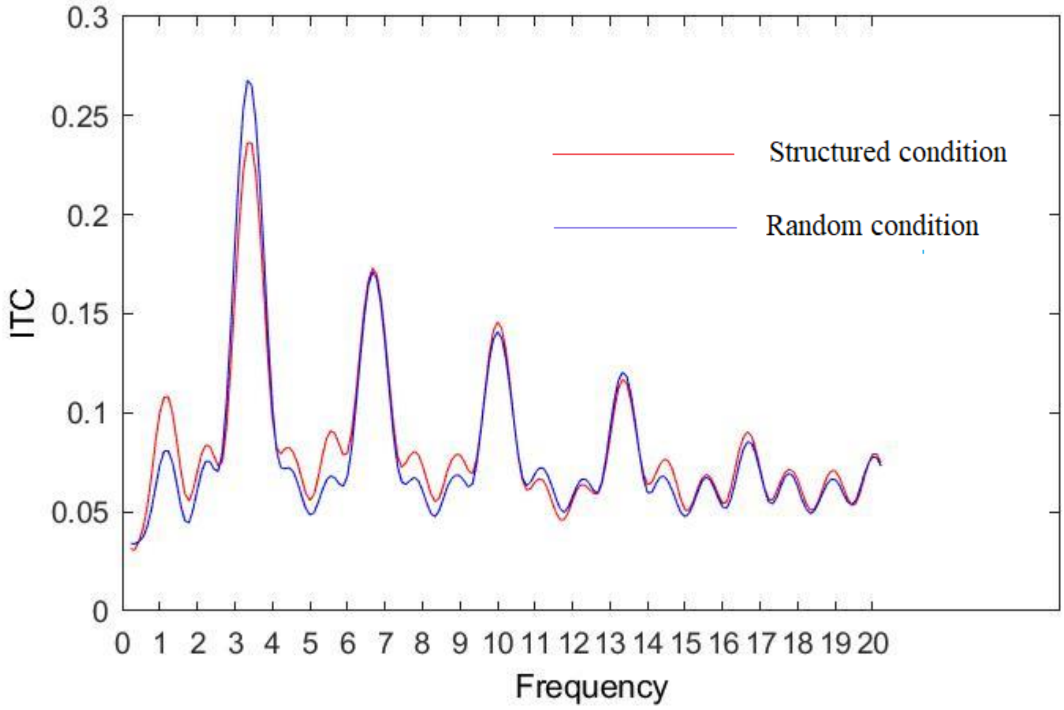


Figure C2. Inter-Trial Coherence (ITC) as a function of frequency during block 2 for all measured frequencies (0.1-20 Hz).

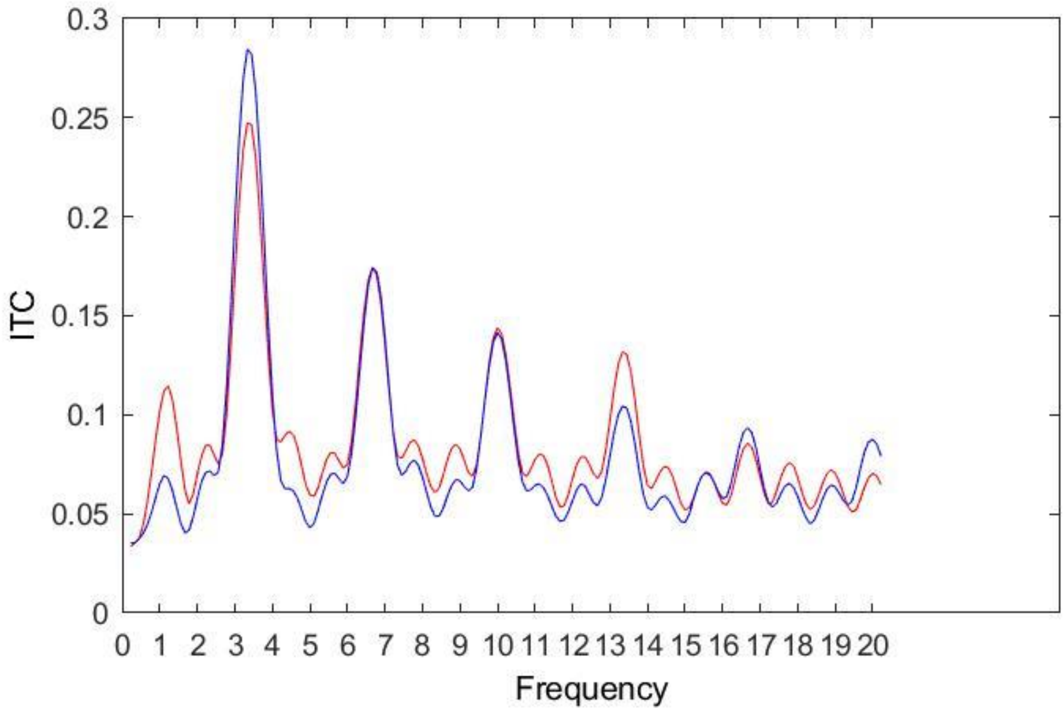


Figure C3. Inter-Trial Coherence (ITC) as a function of frequency during block 3 for all measured frequencies (0.1-20 Hz).

Appendix D. Summary of our TP model and OCP model.

Summary of LMM of B&P2017 data re-analysis

	TP model				**OCP model**			
Predictors	*Estimates*	*Statistic*	*p*	*df*	*Estimates*	*Statistic*	*p*	*df*
(Intercept)	0.40 (0.31, 0.50)	8.24	**<0.001**	71811.00	0.45 (0.35, 0.54)	9.35	**<0.001**	71811.00
Condition [Structured]	0.10 (0.09, 0.11)	15.79	**<0.001**	71811.00				
Block [2]	-0.02 (-0.03, -0.01)	-4.06	**<0.001**	71811.00	0.01 (0.01, 0.02)	5.01	**<0.001**	71811.00
Block [3]	-0.02 (-0.03, -0.01)	-4.70	**<0.001**	71811.00	0.10 (0.09, 0.10)	32.62	**<0.001**	71811.00
Condition [Structured] * Block [2]	0.07 (0.06, 0.08)	11.59	**<0.001**	71811.00				
Condition [Structured] * Block [3]	0.21 (0.20, 0.22)	35.57	**<0.001**	71811.00				
centered_ocp					0.23 (0.20, 0.25)	18.71	**<0.001**	71811.00
centered_ocp * Block [2]					0.17 (0.14, 0.20)	12.48	**<0.001**	71811.00
centered_ocp * Block [3]					0.60 (0.57, 0.62)	40.98	**<0.001**	71811.00
Random Effects								
σ^2	0.10				0.10			
τ_{00}	0.00 $_{Word}$				0.00 $_{Word}$			
	0.11 $_{Participant}$				0.10 $_{Participant}$			
ICC	0.51				0.51			
N	45 $_{Participant}$				45 $_{Participant}$			
	1244 $_{Word}$				1244 $_{Word}$			
Deviance	39146.959				39013.670			
log-Likelihood	-19573.479				-19506.835			

Appendix E. Full table of pairwise comparisons for the TP model

Table E.
*Pairwise Comparisons Condition*Block in the TP model*[a]

(I) reference Random Block 1	Mean difference (I-J)	Std. Error	df	*t*	Sig.[b]
Random 1 – Structured 1	-0.10	0.006	291	-15.79	< .001
Random 2 – Structured 2	-0.17	0.006	292	-26.64	< .001
Random 3 – Structured 3	-0.31	0.006	292	-49.05	< .001
Random 1 – Random 2	0.02	0.004	33937	4.06	< .001
Random 2 – Random 3	0.01	0.004	34055	0.64	0.99
Random 1 – Random 3	0.02	0.004	35644	4.70	< .001
Structured 1 – Structured 2	-0.05	0.004	71741	-12.41	< .001
Structured 2 – Structured 3	-0.14	0.004	71743	-33.63	< .001
Structured 1 – Structured 3	-0.19	0.004	71764	-45.96	< .001

a. Dependent Variable: WLI (B&P2017)
b. Adjustment for multiple comparisons: Tukey