UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Statistical Word Segmentation in Unfamiliar Speech

Permalink

https://escholarship.org/uc/item/0sz0q4p8

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Lu, Helen Shiyang
Werker, Janet
Black, Alexis K.

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Statistical Word Segmentation in Unfamiliar Speech

Helen Shiyang Lu (helens.lu@ubc.ca)
School of Audiology and Speech Sciences, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

Janet F. Werker (jwerker@psych.ubc.ca)
Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Alexis K. Black (alexis.black@audiospeech.ubc.ca)
School of Audiology and Speech Sciences, University of British Columbia, Vancouver, BC V6T 1Z3, Canada

Abstract

Statistical learning, the ability to detect patterns in sensory input, allows listeners to segment words from continuous speech by tracking transitional probabilities. While this mechanism is robust in familiar contexts, its adaptability to unfamiliar speech with distinct phonological properties remains less understood. This study investigates whether English-speaking adults can use TPs to segment an artificial language modeled on Cantonese. Participants identified words where syllables consistently occurred together (statistical words) and syllables that partially co-occurred (part-words) compared to those that never did (non-words). However, they struggled to distinguish statistical words from part-words when frequency was controlled. Pupillometry results showed participants dilated more to part-words and non-words at test, compared to frequency-controlled statistical words. Pupillary responses during familiarization also predicted test performance, demonstrating the potential of pupillometry to track learning in real time. These findings highlight the flexibility of statistical learning in adapting to novel linguistic contexts while revealing its limitations.

Keywords: statistical learning; word segmentation; artificial language; pupillometry

Statistical learning (SL), the ability to detect regularities in sensory input, is an important mechanism for language learning. This mechanism allows listeners to segment words from continuous speech by tracking transitional probabilities (TPs) between syllables, where higher probabilities indicate within-word sequences and lower probabilities signal word boundaries. Both infants (e.g., Saffran, Aslin, & Newport, 1996) and adults (e.g., Saffran, Newport, & Aslin, 1996) have demonstrated the ability to use TPs for word segmentation, but the robustness of this ability in unfamiliar linguistic contexts remains less explored.

Prior linguistic experience can both facilitate and constrain SL. Adults bring a wealth of prior knowledge and cognitive resources to learning tasks, such as enhanced working memory and attention control, which may support SL in complex or novel conditions (e.g., Misyak & Christiansen, 2012; Toro, Sinnett, & Soto-Faraco, 2005). However, entrenched expectations based on native language can make adapting to unfamiliar input challenging. T. Wang and Saffran (2014) found that English monolingual adults struggled to segment an artificial tonal language, likely due to limited exposure to tonal cues, while bilingual participants and those with tonal language experience performed significantly better. This aligns with findings from Siegelman, Bogaerts, Elazar, Arciuli, and Frost (2018), who highlighted the role of prior knowledge in shaping SL outcomes and emphasized that task demands can either align with or challenge existing knowledge.

Infant studies provide additional insights into the strengths and limitations of SL. For example, Graf Estes, Gluck, and Bastos (2015) showed that English-learning 14-month-olds could segment an artificial language modeled on Mandarin syllables using TPs alone, despite the phonological differences between Mandarin and English. However, success was influenced by factors such as task difficulty and sex. When statistical differences between words (TP = 1) and foils were less pronounced (TP = 0.5 vs. 0), only male infants demonstrated evidence of segmentation. A meta-analysis by Black and Bergmann (2017) further revealed that infants performed better with synthetic stimuli than natural speech, suggesting that greater acoustic complexity and variability pose additional challenges for SL. While these findings are specific to infants, they raise important questions about how adults, with their greater linguistic experience, might handle the added complexity of foreign speech.

Failures to demonstrate successful segmentation in more complex learning contexts, such as non-native speech, may reflect genuine learning difficulties. However, they could also indicate slower or more fragile learning that behavioral tasks fail to fully capture. Explicit tasks, which require participants to consciously reflect on segmented units, have been argued to underestimate SL (Christiansen, 2019). Nevertheless, explicit measures may still capture important aspects of learning that are not reflected in purely implicit tasks. For example, Batterink, Reber, Neville, and Paller (2015) tested adults' SL by combining an explicit recognition task, a reaction-time-based implicit behavioral task, and ERP measures. They found that while the reaction-time-based task and ERP measures were correlated and tested implicit knowledge generated by SL, these measures did not correlate with the explicit recognition task. In particular, the participants demonstrated some evidence of learning in the explicit task, leading them to argue that SL generates both explicit and implicit knowledge, with different tasks tapping into distinct types of knowledge.

More recently, physiological measures such as pupillometry have been adopted to study statistical learning. Changes in pupil size can reflect cognitive processing load and surprisal, providing an implicit measure of learners' responses to structured input. For instance, Marimon, Höhle, and Langus (2022) demonstrated that both German infants and adults exhibited greater pupil dilation when encountering unexpected patterns in speech streams. Among adults, pupillary responses corresponded with their behavioral performance (i.e.,

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

pupils dilated more to sequences they rated as less familiar). Moreover, the degree of alignment between pupil fluctuations and word timing in training predicted subsequent test performance, suggesting that pupillometry can capture meaningful individual differences in the learning process. These findings indicate that pupillometry serves as a valuable measure of statistical learning, capable of indexing both ongoing processing and later outcomes.

The current study investigates whether English-speaking adults can use TPs alone to segment words from a continuous speech stream modeled on Cantonese, a language with distinct phonological properties from English. Processing unfamiliar sounds is often more effortful, which may make segmentation more challenging. To assess learning, we used both an explicit recognition task and pupillometry. We predicted that successful learners would recognize high-probability syllable sequences (i.e., statistical words) and show larger pupil dilation to low-probability sequences (i.e., part-words and non-words). Additionally, we expected pupil responses during familiarization to align with statistical word boundaries, indicating sensitivity to the underlying structure.

If adults succeed in segmenting words from a Cantonese-like speech stream, it would suggest that SL is robust and capable of adapting to typologically unfamiliar linguistic input. Alternatively, if adults struggle, it would emphasize the limits of SL when prior knowledge does not align with the input, raising questions about the extent to which SL functions as a core mechanism in second language acquisition. This would prompt broader theoretical considerations about the role of SL in language learning and the conditions under which it is most effective. These findings will advance our understanding of SL mechanisms and the cognitive processes underlying learning in unfamiliar contexts. By integrating explicit behavioral and implicit physiological measures, this research also demonstrates the value of complementary methodologies in capturing the multifaceted nature of statistical learning.

Methods

Participants

Data were collected from 40 eligible participants recruited through the university's subject pool system ($M_{age} = 21.83$ years old, $SD_{age} = 6.15$, $range: 17 - 45$). All participants were native English speakers who had no prior exposure to tonal languages and reported no history of speech or hearing disorders. Participants provided informed consent before participation, and the study was approved by the university's ethics review board.

Stimuli

Eight Cantonese-like syllables were created specifically for this study: *caa2, ge6, je2, ngo3, wu5, zi4, zo1,* and *zyu5*. These syllables combined legal sounds and tones in unattested ways, ensuring that they did not carry semantic meaning. To confirm the syllables' adherence to phonotactic constraints, two native Cantonese speakers unfamiliar with the study's hypotheses independently verified that the combinations were plausible within Cantonese phonology.

The syllables were recorded individually in a sound-attenuated booth by a 22-year-old male native Cantonese speaker. To standardize duration across stimuli, the recordings were modified using the *lengthen* function in Praat (Boersma, 2001), resulting in syllables of 300 ms each. An additional 10 ms of silence was added to the beginning of each syllable to enhance clarity, bringing the total duration to 310 ms. This corresponds to a syllable rate of 3.2 Hz, consistent with naturalistic speech rates (Ding et al., 2017).

Two language conditions were created for counterbalancing purposes. The eight syllables were randomly combined to form four disyllabic words for each condition. The order of these words in the familiarization stream was pseudo-randomized, ensuring that TPs between syllables spanning word boundaries were lower, ranging from 0.50 to 0.21. No immediate repetitions of the same word were allowed.

After familiarization, participants were tested on their ability to distinguish statistical words, part-words, and non-words. Statistical words were the four disyllabic words that were used to construct the familiarization stream (TP = 1). Part-words were composed of syllable pairs spanning word boundaries during familiarization, with transitional probabilities ranging from 0.50 to 0.21. Non-words consisted of syllable pairs that never occurred together in the familiarization stream (TP = 0).

To control for potential frequency effects (Aslin, Saffran, & Newport, 1998), two of the statistical words were presented 90 times each during familiarization, while the other two were presented 45 times each. In this way, the two part-words formed from the frequent statistical words were matched in frequency (44 and 45 occurrences) to the infrequent statistical words. Table 1 provides an overview of the statistical words, part-words, and non-words, along with their frequencies during the familiarization phase.

Procedure

Participants were told that they would hear an artificial language and be tested on it later. They were instructed to keep their gaze on the screen during the experiment as much as possible. Participants were seated in a sound attenuated booth with ambient lighting, about 60 to 70 cm away from a 16.5 x 10.5-inch screen, and their pupil responses were recorded with a SR Eyelink 1000 Plus eye-tracker. The audio was played through a speaker positioned behind the screen. Before starting, a 5-point calibration was performed to ensure accurate eye-tracking, with a successful calibration defined as having less than 1 degree of error on average.

After calibration, participants listened to a continuous stream of syllables while watching a video of an aquarium with fish moving slowly on the screen. The video was included to help maintain participants' attention. The audio stream included 2480-ms linear fade-in and fade-out ramps at the edges to minimize additional segmentation cues. In addition, the video had relatively consistent luminance (in

Table 1: Syllable combinations for statistical words, part-words, and non-words.

Word type	Language	
	Language 1	Language 2
Statistical (TP = 1)	zo1-ca2 (*90*)	je2-ge6 (*90*)
	zi4-wu5 (*90*)	zyu5-ngo3 (*90*)
	ngo3-je2 (*45*)*	ca2-zi4 (*45*)*
	ge6-zyu5 (*45*)*	wu5-zo1 (*45*)*
Part-words (TP = 0.21-0.5)	ca2-zi4 (*44*)*	ngo3-je2 (*44*)*
	wu5-zo1 (*45*)*	ge6-zyu5 (*45*)*
	wu5-ge6 (*19*)	ge6-wu5 (*19*)
	ca2-ngo3 (*20*)	ngo3-ca2 (*20*)
Non-words (TP = 0)	ge6-ngo3 (*0*)	ge6-ngo3 (*0*)
	wu5-ca2 (*0*)	wu5-ca2 (*0*)
	zi4-zo1 (*0*)	zi4-zo1 (*0*)
	zyu5-je2 (*0*)	zyu5-je2 (*0*)

Note. The frequencies during familiarization are in italicized numbers. Words with asterisks are frequency-controlled and counter-balanced across the two conditions.

gray-scale intensity units, which represent relative brightness where 0 indicates black and 255 indicates white in an 8-bit image: $M = 72.26$, *range* = 71.27–73.97). This consistency minimizes any potential confounding effects on pupil dilation caused by abrupt changes in brightness, ensuring that observed changes in pupil size were more likely attributable to cognitive responses rather than luminance variability.

Immediately after the familiarization phase, participants moved on to the test phase, which included 36 trials organized into three blocks. Each trial began with a small, silent animation (200 x 200 pixels) of a bouncing baby, which moved slightly (10 pixels) in the center of the screen to capture participants' attention. Once participants focused on the animation for at least 1000 ms, a disyllabic test word was played (620 ms), followed by 2500 ms of silence. Pupil sizes were recorded throughout the trial, starting from the 1000 ms fixation period before the word onset (used as a baseline) and continuing for 3120 ms after the word onset (including the word duration and the silence period). The animation remained on the screen throughout the 3120 ms post-word period to prevent sudden changes in luminance. After the silence, participants were prompted to decide if the word they heard was part of the familiarization stream by pressing a "yes" or "no" button on a button box. Each block included all 12 test words (four statistical words, four part-words, and four non-words), and their order was randomized so that no more than two words of the same type appeared in a row. At the start of each block, a drift correction was performed using a looming animation with sounds.

Pre-processing of Pupillary Data

The pre-processing approach was modeled after Marimon et al. (2022)'s paper and modified based on the suggestions from Mathôt and Vilotijević (2023). In the test phase, samples with looks outside of the animation region were removed. Then, trials with more than 25% missing eye-tracker samples during the baseline or after word onset were excluded. Additionally, trials where baseline pupil sizes were more than 2 standard deviations from the participant's average baseline pupil size were excluded. Missing data in the remaining trials were linearly interpolated and extrapolated using the trial-level median pupil size. After this, data from each test trial were normalized using z-scores, linearly detrended to remove slow drifts, baseline corrected using the mean pupil size from the 1000 ms pre-stimulus period, and smoothed using a moving filter. To reduce computational load, pupil size was averaged across every 5 samples, resulting in a sampling rate of 100 Hz. For subsequent analyses of pupillary data, we included only trials with statistical words and part-words that were matched in frequency, as well as non-words, resulting in a maximum of 24 trials per participant.

Data from the familiarization phase were excluded if participants provided less than 75% valid eye-tracker samples. Pupillary changes at the word frequency during familiarization were analyzed from the onset of the first statistical word after the audio ramp-up to the end of the last word before the audio ramp-down, resulting in a duration of 162.44 seconds (i.e., 262 words). Missing data were first linearly interpolated. Then, the pupillary data were linearly detrended, baseline corrected, and bandpass-filtered between 0.2 and 25 Hz. The resulting data were then transformed from the time domain to the frequency domain using a Fast Fourier Transform (FFT) implemented with the *eegfft* function from the *eegkit* package in R (Helwig, 2018), producing the frequency spectra of the pupillary response for each participant.

To assess the temporal alignment of the pupillary response to the word frequency, we focused on the phase shift of the 1.61 Hz component of the frequency spectrum obtained from the FFT. This frequency corresponds to the word rate, calculated as $1/0.620s = 1.61Hz$. The phase represents the temporal shift of the frequency response relative to stimulus onset, measured in radians. Since the analysis window began with a statistical word, a phase of 0 radian (i.e., no temporal shift) would indicate entrainment to statistical words.

Results

Behavioral Responses

The behavioral responses of participants to the test words are presented in Figure 1. To analyze these responses, we used a generalized linear mixed-effects model with a binomial link function, implemented through the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2015). The final model included participants' behavioral responses to individual test words (coded as 1 for "yes" and 0 for "no") as the dependent variable, *word type* (non-word, part-word, statistical word) and *block* (1–3) as the fixed effects, and by-subject random intercepts. *Word type* was coded with statistical words as the reference level.

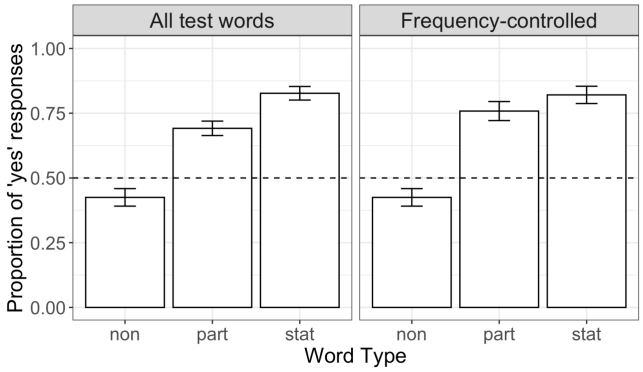


Figure 1: Proportion of "yes" responses to test words with standard error bars.

The model resulted in a significant *Intercept* ($\beta = 2.13, SE = 0.22, z = 9.74, p < .001$), a significant effect of *word type* (statistical words vs. non-words: $\beta = -1.98, SE = 0.16, z = -12.54, p < .001$; statistical words vs. part-words: $\beta = -0.79, SE = 0.16, z = -4.98, p < .001$), and a significant effect of *block* ($\beta = -0.23, SE = 0.07, z = -3.11, p = .002$). Overall, in the test phase, participants were more likely to indicate that they had heard the words than to indicate that they had not. They were also more likely to endorse statistical words than part-words or non-words. However, participants became less likely to respond "yes" to test words as they completed more blocks.

We then performed a follow-up analysis using a subset of the data, focusing only on statistical words and part-words that were matched for frequency. Again, the selected model included participants' binary responses (1 = "yes," 0 = "no") as the dependent variable, *word type* (non-word, part-word, statistical word) and *block* (1–3) as fixed effects, and by-subject random intercepts. *Word type* was still coded with statistical words as the reference level.

The results showed a significant *Intercept* ($\beta = 2.04, SE = 0.28, z = 7.36, p < .001$) and a significant effect of *block* ($\beta = -0.20, SE = 0.09, z = -2.17, p = .030$). Participants still were more likely to select "yes" for statistical words compared to non-words ($\beta = -1.97, SE = 0.20, z = -9.79, p < .001$). However, the effect of *word type* was no longer significant for statistical words and part-words ($\beta = -0.40, SE = 0.23, z = -1.73, p = .084$). These results suggest that when frequency was controlled, participants were less consistent, or less capable in identifying words from part-words, highlighting the potential influence of frequency, as opposed to transitional probabilities, in statistical word segmentation of unfamiliar speech (see below for further discussion).

Pupil Dilation in the Test Phase

After pre-processing, we retained data from 858 valid test trials across 40 participants, with the number of trials per participant ranging from 9 to 24. To examine differences in pupil dilation across word types, we conducted a cluster-based permutation test using linear mixed effects models. Pupil size was the dependent variable, with *word type* as a categorical fixed factor (statistical words, part-words, non-words; statistical words as reference level). Random intercepts were included for both participants and trials. The cluster-based permutation test, implemented using the *clusterperm.lmer* function from the *permutes* package, used a likelihood ratio test for cluster-level statistics (Voeten, 2023). Clusters were identified from consecutive time bins where the fixed factor estimates reached a significance level of $p < .05$. When comparing non-words to statistical words, two significant time windows were identified: one from word onset to 310 ms after onset (*cluster mass LRT* = 228.94, $p < .001$) and another between 2270 ms and 2860 ms after word onset (*cluster mass LRT* = 504.51, $p < .001$), during which non-words elicited a greater pupillary response. Similarly, part-words elicited larger pupil dilation than statistical words in two distinct time windows: from word onset to 290 ms after onset (*cluster mass LRT* = 262.46, $p < .001$) and between 2060 ms and 2690 ms after word onset (*cluster mass LRT* = 515.39, $p < .001$). Additionally, a brief time window towards the end of the trials (3010–3090 ms after word onset) showed the opposite pattern, with statistical words eliciting larger pupil dilation than part-words (*cluster mass LRT* = 58.13, $p < .001$). These relevant time windows are highlighted by the colored regions in Figure 2.

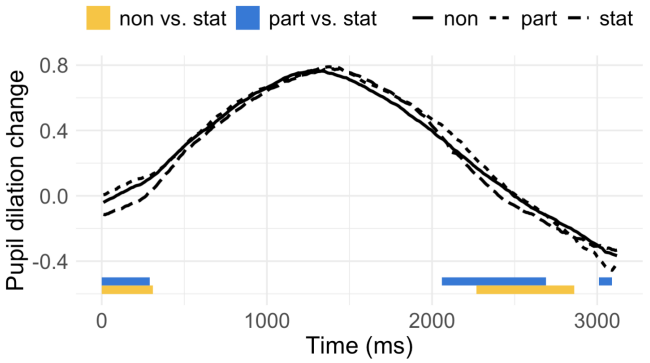


Figure 2: Mean pupil dilation across time for statistical words, part-words, and non-words. Colored regions indicate time frames with significant differences across word types.

Pupillary Entrainment in the Familiarization Phase

To analyze participants' performance during the familiarization phase, we examined their pupillary responses. Specifically, we analyzed the temporal alignment (phase shifts) of pupillary changes at the word frequency (1.61 Hz) using FFT. After the pre-processing steps, all participants ($n = 40$) provided at least 75% valid samples in the familiarization phase. The frequency distribution of their phase shifts is shown in Figure 3. The phases of the 1.61 Hz pupillary response were skewed toward the onset of statistical words, with a median of 1.20 radians ($mean = 1.24, range = 0.11 - 2.86$ radians).

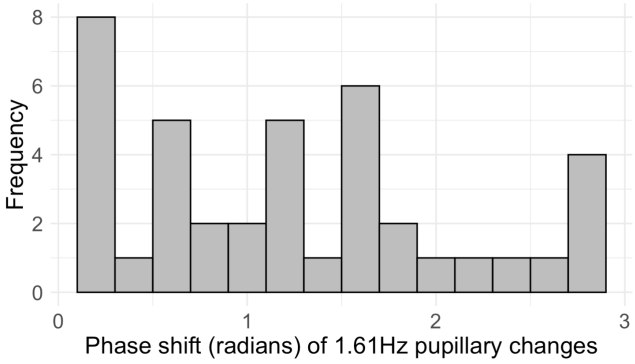


Figure 3: Distribution of participants across phase shift values of pupillary changes at 1.61 Hz, showing skew toward statistical word onsets (0 radian).

To investigate whether pupillary entrainment to statistical words during familiarization predicted participants' performance during the test phase, we assessed the relationship between the phases of these pupillary changes during familiarization and the pupillary responses for frequency-matched statistical words and part-words at test. A cluster-based permutation test was conducted using linear mixed-effects models, focusing on the time window where the interaction between condition and phase was significant ($p < .05$).

The model included pupil dilation as the dependent variable, with *word type* (statistical vs. part-words) as a categorical fixed factor and the *phase shift* of pupillary changes at the word frequency as a continuous fixed factor (ranging from 0 radians, corresponding to statistical word onset, to 3.14 radians, corresponding to part-word onset). Additionally, the interaction between *word type* and *phase shift* was included. Participant and trial were modeled as random factors with random intercepts. The interaction between *word type* and *phase shift* was significant in four time windows: 690-1080 ms after word onset (*cluster mass LRT* = 338.44, $p < .001$), 1770-2130 ms after word onset (*cluster mass LRT* = 348.91, $p < .001$), 2720-2760 ms after word onset (*cluster mass LRT* = 24.48, $p = .017$), and 2800-2910 ms after word onset (*cluster mass LRT* = 100.83, $p < .001$).

To explore further, a post-hoc linear mixed-effects analysis was conducted using the same model, and the average pupil size within the significant clusters was used as the dependent variable. The results revealed a significant *intercept* ($\beta = 0.35, SE = 0.07, t = 5.27, p < .001$), a significant effect of *word type* ($\beta = -0.05, SE = 0.01, t = -6.93, p < .001$), and a significant interaction between *word type* and *phase shift* ($\beta = 0.04, SE = 0.005, t = 7.38, p < .001$). No significant main effect of *phase shift* was found ($\beta = 0.06, SE = 0.04, t = 1.38, p = .176$). It is important to note that this post-hoc analysis was performed on data selected using cluster-based permutation tests, which may have inflated effect sizes.

As shown in Figure 4, participants who had pupillary responses during familiarization closely aligned with statistical word onsets (i.e., lower phase shift values) dilated less to statistical words and more to part-words at test. In contrast, participants with weaker alignment during familiarization (i.e., higher phase shift values) showed less differentiation in dilation between statistical words and part-words.

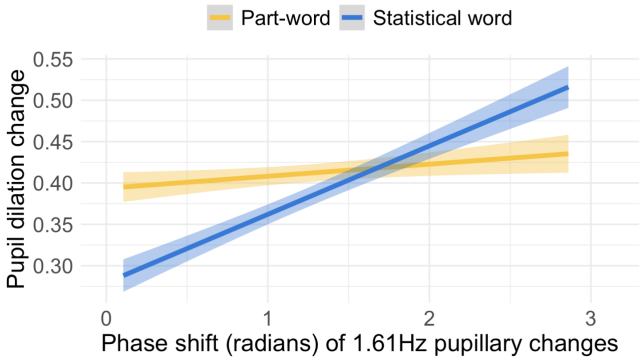


Figure 4: Relationship between pupillary phase shift values at 1.61 Hz during familiarization and average pupil dilation change from baseline in response to statistical words and part-words at test. Shaded regions represent standard errors.

Discussion and Conclusions

This study investigated whether adults could use transitional probabilities (TPs) to segment a continuous speech stream modeled on a foreign language.

Behavioral responses revealed reliable distinctions between statistical words and non-words but less consistent differentiation between statistical words and part-words, particularly when frequency was controlled. This contrasts with findings from familiar speech contexts. For example, F. H. Wang, Luo, and Wang (2024) demonstrated that native Mandarin speakers successfully segmented a speech stream structured according to Mandarin phonology, rating statistical words as more familiar than part-words even when frequency was matched and exposure was minimal. These differences suggest that while TPs can support segmentation within familiar phonological systems, unfamiliar input imposes additional demands that hinder the formation of robust word representations based on statistical cues alone.

One possibility is that entrenched expectations from native language phonology disrupted learning when participants encountered unfamiliar phonological patterns, contributing to the comparable behavioral responses to frequency-controlled statistical words and part-words observed in our study. This interpretation aligns with prior research showing that statistical learning mechanisms are less effective when processing unfamiliar linguistic input (Graf Estes et al., 2015; Siegelman et al., 2018; T. Wang & Saffran, 2014).

Another possibility, though not mutually exclusive with the first, comes from computational modeling work suggesting that statistical learning mechanisms are initially more sensitive to frequencies and only later shift toward prioritizing

transitional probabilities (Mirman, Graf Estes, & Magnuson, 2010). In our study, adults struggled to distinguish statistical words from part-words when frequency cues were controlled, suggesting that this shift may have been delayed by the phonological unfamiliarity of the input.

Pupillary data at test provided more consistent evidence of segmentation, with greater dilation to non-words and part-words than statistical words. These responses suggest that participants implicitly differentiated word types, even though this was not consistently reflected in the explicit recognition task. This discrepancy may reflect multiple factors. One is that individual differences in accessing and reporting knowledge can obscure learning in the recognition task (Christiansen, 2019). In addition, the frequency-controlled statistical words were encountered less often, and limited exposure, combined with the processing challenges of unfamiliar phonology, may have been insufficient to form robust representations. These factors together may have reduced sensitivity in the recognition task. Future studies could consider more implicit measures, such as the rating task used by F. H. Wang et al. (2024), which may better detect subtle or emerging learning in such contexts.

Notably, differences in pupil dilation were observed approximately 2 seconds after word onset, a time frame consistent with prior findings (Marimon et al., 2022). However, we also observed early effects immediately after word onset. These early effects are difficult to interpret, as the earliest known pupillary response, the pupil light reflex, has a latency of about 200 ms, while cognitive effects on pupil size tend to arise more slowly (Mathôt & Vilotijević, 2023).

One possible explanation for these early pupillary effects is carry-over from participants' behavioral responses, as the baseline period began immediately after their decisions. Since pupil dilation can reflect decision uncertainty even after a choice is made, the early effects may reflect lingering processing rather than immediate responses to the stimulus (Urai, Braun, & Donner, 2017). However, this interpretation remains speculative, as our current data cannot directly test it. In future studies, we plan to introduce a delay between participants' decisions and the baseline period to reduce carry-over effects and better isolate pupillary responses.

In addition to pupil dilation differences observed at test, pupillary entrainment during familiarization provided further evidence of sensitivity to statistical structure. Participants generally showed phase alignment toward the TP-defined rhythm of the speech stream, and those whose pupillary responses were more closely aligned with statistical word onsets exhibited greater differentiation between statistical words and part-words at test. These results suggest that implicit alignment with the statistical structure during familiarization predicts later sensitivity to word boundaries.

While the observed phase alignment indicates that participants tracked TPs during familiarization, the considerable variability across individuals suggests that entrainment was not precisely synchronized. Several factors likely contributed to this variability. Compared to Marimon et al. (2022), where prosodic cues and native-language familiarity supported segmentation, our study relied solely on transitional probabilities within an unfamiliar phonological context. Prosodic cues are strong indicators of word boundaries (Thiessen & Saffran, 2003, 2007; Marimon et al., 2022; Marimon, Langus, & Höhle, 2024), and their absence likely increased task demands. Unfamiliar phonetic and rhythmic patterns may have further heightened individual differences in segmentation.

Additionally, differences in equipment may have contributed to the observed variability. While Marimon et al. (2022) used an eye-tracker that recorded pupil diameter in millimeters, our equipment recorded measurements in arbitrary units. Variations in how the devices compensated for factors such as head movement and viewing distance may have introduced additional noise into the pupillary signal.

Nevertheless, despite these sources of variability, the phase alignment of pupillary responses during familiarization reliably predicted performance at test: participants who exhibited stronger alignment with statistical word onsets during familiarization showed greater differentiation between statistical words and part-words at test. This finding suggests that pupillary entrainment captured meaningful individual differences in sensitivity to the statistical structure of the input, providing a valuable real-time index of implicit learning.

Overall, the findings support the hypothesis that statistical learning mechanisms in adults are flexible enough to operate in novel linguistic contexts, such as segmenting unfamiliar speech using only TPs. However, they also highlight challenges in adapting to unfamiliar phonological systems. Behavioral responses showed limited ability to distinguish statistical words from part-words when frequency was controlled. In contrast, pupillary data provided consistent evidence of sensitivity to statistical structure, indicating that some aspects of statistical learning remain robust even when explicit behavioral performance is fragile.

These findings shed light on second language acquisition, suggesting that while adults retain the ability to track novel regularities, first language knowledge can constrain the application of these mechanisms to unfamiliar input. This aligns with prior work showing that native phonotactics can impair the use of new statistical cues for segmentation (Finn & Hudson Kam, 2008). In both cases, the challenge is not merely detecting regularities, but overcoming prior expectations that shape how learners interpret unfamiliar speech. Although statistical learning likely plays a foundational role in second language acquisition, successful adaptation may require recalibrating entrenched linguistic biases and integrating multiple sources of information over time. Thus, while statistical learning remains a flexible mechanism for adapting to novel input, it operates within cognitive and linguistic systems shaped by prior experience, posing both opportunities and constraints for learning new languages in adulthood.

Acknowledgments

We acknowledge that this research was carried out on the traditional, ancestral, and unceded territory of the xʷməθkʷəy̓əm (Musqueam) Nation. This work was supported by a UBC Killam Postdoctoral Research Fellowship (to HSL) and funding from SSHRC (to AKB; 435-2024-0782) and NSERC (to JFW; 2020-05202). We thank Fay Zhuozhuo Han for assistance with analysis scripts, and Grace Bu and Zoe Chung for their help with stimulus preparation and data collection.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78.

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In *Proceedings of the 39th annual conference of the cognitive science society* (pp. 124–129). London, UK.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499.

Graf Estes, K., Gluck, S. C.-W., & Bastos, C. (2015). Flexibility in statistical word segmentation: Finding words in foreign speech. *Language Learning and Development*, *11*(3), 252–269.

Helwig, N. E. (2018). eegkit: Toolkit for electroencephalography data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=eegkit` (R package version 1.0-4)

Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old german-learning infants. *Cognition*, *224*, 105054.

Marimon, M., Langus, A., & Höhle, B. (2024). Prosody outweighs statistics in 6-month-old german-learning infants' speech segmentation. *Infancy*, *29*(5), 750–770.

Mathôt, S., & Vilotijević, A. (2023). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077.

Mirman, D., Graf Estes, K., & Magnuson, J. S. (2010). Computational modeling of statistical learning: Effects of transitional probability versus frequency and links to word learning. *Infancy*, *15*(5), 471–486.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302–331.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

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.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73–100.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637.

Voeten, C. C. (2023). permutes: Permutation tests for time series data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=permutes` (R package version 2.8)

Wang, F. H., Luo, M., & Wang, S. (2024). Statistical word segmentation succeeds given the minimal amount of exposure. *Psychonomic Bulletin & Review*, *31*(3), 1172–1180.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953.