

Syllables in Sync Form a Link: Neural Phase-locking Reflects Word Knowledge during Language Learning

Laura Batterink

Abstract

■ Language is composed of small building blocks, which combine to form larger meaningful structures. To understand language, we must process, track, and concatenate these building blocks into larger linguistic units as speech unfolds over time. An influential idea is that phase-locking of neural oscillations across different levels of linguistic structure provides a mechanism for this process. Building on this framework, the goal of the current study was to determine whether neural phase-locking occurs more robustly to novel linguistic items that are successfully learned and encoded into memory, compared to items that are not learned. Participants listened to a continuous speech stream composed of repeating nonsense

words while their EEG was recorded and then performed a recognition test on the component words. Neural phase-locking to individual words during the learning period strongly predicted the strength of subsequent word knowledge, suggesting that neural phase-locking indexes the subjective perception of specific linguistic items during real-time language learning. These findings support neural oscillatory models of language, demonstrating that words that are successfully perceived as functional units are tracked by oscillatory activity at the matching word rate. In contrast, words that are not learned are processed merely as a sequence of unrelated syllables and thus not tracked by corresponding word-rate oscillations. ■

INTRODUCTION

A hallmark of human language is that it is composed of building blocks, which combine hierarchically to form an infinite number of meaningful expressions. For example, phonemes are combined into syllables, which in turn form words, then phrases, and, finally, full sentences. During speech comprehension, the brain must simultaneously track and concatenate these different linguistic structures across time. Recent evidence suggests that this task may be accomplished through phase-locking of endogenous neural oscillations to linguistic units unfolding at different timescales (Gross et al., 2013; Giraud & Poeppel, 2012; Peelle & Davis, 2012). Neural phase-locking is established by demonstrating a consistent phase lag between recorded neural responses and some sort of external stimulus, as measured over time, over trials, or over participants (Peelle & Davis, 2012). According to several prominent models (Giraud & Poeppel, 2012; Peelle & Davis, 2012), ongoing neural oscillations phase-lock to linguistic segments occurring at different rates, and coupling between phase-locked signals at slow and faster frequencies supports the integration of smaller linguistic elements into larger units. The alignment of neural oscillations with a periodic or quasiperiodic stimulus stream has also been described as “neural tracking”

or “neural entrainment” (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Peelle & Davis, 2012) and is a general phenomenon that also occurs to nonlinguistic stimuli.

Importantly, neural tracking of slow auditory fluctuations (< 10 Hz) does not merely reflect low-level acoustic features of a stimulus but is sensitive to listeners’ abstract knowledge and subjective perceptions of a stimulus, as guided, for example, by imagined rhythms (Nozaradan, Peretz, & Keller, 2016; Nozaradan, Peretz, Missal, & Mouraux, 2011) or syntactic rules (e.g., Ding et al., 2016, 2017). For instance, Ding et al. (2016, 2017) demonstrated that sequences of monosyllabic words, hierarchically organized into phrases and sentences, elicit spectral magnetoencephalography (MEG)/EEG peaks at frequencies corresponding to syllable, word, and phrase presentation rates. Critically, phrasal and word peaks were observed only when sentences were presented in language known to participants, and not an unknown foreign language, indicating that neural tracking of higher-level units depends on language-specific knowledge.

Neural tracking of linguistic structures also emerges during language learning, reflecting the moment-by-moment acquisition of new linguistic representations. Getz, Ding, Newport, and Poeppel (2018) recorded learners’ MEG while they listened to a miniature artificial language that contained embedded phrases made up of words presented at an isochronous rate. Within 3.5 min

Western University, London, ON, Canada

of exposure, learners showed a robust spectral peak in MEG power at the phrase structure rate, reflecting phrase-level tracking. We found similar results in two recent statistical learning studies (Batterink & Paller, 2017, 2019), in which participants were exposed to an isochronous artificial speech stream composed of repeating trisyllabic nonsense words, concatenated together without pauses (e.g., *tupirogolabu...*). In both studies, neural phase-locking to the hidden component words increased over the exposure period and predicted performance on an implicit RT-based measure of statistical word knowledge. Finally, Buiatti, Peña, and Dehaene-Lambertz (2009) also found neural tracking of trisyllabic repeating “AXC” pseudowords with nonadjacent dependencies, although only when words were cued by subliminal 25-msec pauses; this trisyllabic spectral EEG response also correlated with the number of correctly reported words. Taken together, these results indicate that neural tracking of embedded, novel linguistic structures emerges during the learning process and predicts subsequent linguistic knowledge at the behavioral level.

Notably, all of the above studies follow a “frequency-tagging” approach: The experimental stimuli are presented at a steady, isochronous rate, which drives the neural population that codes for the stimulus to oscillate at the same rate. Thus, clear peaks in the frequency spectrum of the recorded neural signal can be detected at the stimulus presentation frequency, providing a “frequency tag” to identify the associated brain response. Although this approach offers an excellent signal-to-noise ratio relative to classical ERP analysis, it requires averaging the neural signal across continuous stimulation blocks, precluding the isolation of neural responses to individual items (e.g., *tupiro* vs. *golabu*; from here on, “item” refers to the combined collection of individual instances of the same word or linguistic unit, at the single participant level). Thus, these past studies are unable to address whether the neural tracking of these isochronous linguistic structures reflects the specific learning of individual items in the artificial language or whether this rhythmic neural response primarily reflects more general individual differences in statistical learning ability. Disentangling these two possibilities would provide novel and important insight into the functional significance of neural phase-locking responses during language learning. Below, I elaborate further on each of these two possibilities in turn.

According to the first possibility, neural phase-locking during language learning may directly reflect the discovery, perception, and encoding of individual linguistic items (e.g., a word in speech). An individual item that is successfully learned should be perceived and categorized as a relevant linguistic segment rather than as an unrelated sequence of syllables and should be represented by an underlying neural population that codes for this item as a meaningful chunk. Thus, better-learned items should elicit stronger neural phase-locking at the

corresponding presentation rate. In contrast, an individual item that has not been learned would be encoded and represented only at the syllabic level, eliciting phase-locking at the syllable rate but not at the word rate. Under this scenario, the previous findings that better learners show higher neural entrainment over the learning period (Batterink & Paller, 2017, 2019; Getz et al., 2018; Buiatti et al., 2009) would be driven by these learners’ successful discovery of more total items in the language and/or their stronger representations for each item.

A second, alternative possibility is that neural tracking of rhythmic linguistic structures is not sensitive to item-level differences in learning but is primarily driven by differences at the individual level. To illustrate, the neural entrainment response observed in previous studies (Batterink & Paller, 2017, 2019; Getz et al., 2018; Buiatti et al., 2009) may index an individual’s general sensitivity to rhythmic temporal patterns, which could represent a stable individual trait that in turn predicts statistical learning ability. In line with this idea, it has been shown that stronger endogenous neural entrainment at the beat frequency to auditory rhythms is associated with superior temporal prediction abilities (Nozaradan et al., 2016). Another recent study found that individuals vary distinctly in their sensitivity to external patterns, as assessed by whether they spontaneously synchronize their own speech to an isochronous speech rhythm (Assaneo et al., 2019). This individual predisposition predicts statistical learning performance and also correlates with neuroanatomical and neurophysiological outcomes, suggesting that it is a stable individual trait. Given that previous studies showing a link between neural entrainment and statistical language learning all presented isochronous stimuli (Batterink & Paller, 2017, 2019; Getz et al., 2018; Buiatti et al., 2009), better learners may be those individuals who are generally more sensitive to temporal rhythms or who demonstrate a high degree of spontaneous synchronization to external stimuli. These traits could both produce a higher neural entrainment response and lead to better statistical learning performance on subsequent tests. Under this hypothesis, the contribution of individual items to variance in the neural tracking response would be negligible, and the relationship between neural entrainment and word learning would be driven primarily by stable individual differences (such as temporal prediction ability) that operate similarly across all items.

The goal of this study was to address the question of whether neural phase-locking is sensitive to the discovery of individual items during language learning, thereby advancing our understanding of the functional significance of neural phase-locking in the context of language. In particular, I tested whether neural phase-locking during learning reflects acquired knowledge of specific words in an artificial language, as opposed to more generally indexing interindividual differences in processing that would operate similarly across all items. Following a

classic statistical learning design, participants listened to a continuous stream composed of repeating nonsense words and then completed a recognition test. I hypothesized that neural phase-locking to a given word during learning should be higher when that word is successfully perceived as a functional unit, such that phase-locking and subsequent recognition performance should correlate at the individual item level. Alternatively, if neural phase-locking during learning and subsequent word recognition correlate at the individual participant level (replicating prior findings), but not at the individual item level, this would provide evidence for the alternative hypothesis: the idea that neural phase-locking during statistical learning is primarily driven by interindividual differences that operate similarly across all items.

METHODS

Participants

Twenty-one participants (10 women) contributed data to this study. Participants were recruited at Northwestern University and were paid \$10/hr. They were all fluent English speakers, between 19 and 23 years old (mean = 20.5 years), and had no history of neurological problems. The study was undertaken with the understanding and written consent of each participant.

Data from other tasks completed by this sample of participants have been reported in a previous publication (implicit training group; Batterink, Reber, & Paller, 2015). Briefly, the goal of this previous study was to test whether the ability to predict incoming stimuli, a key function of statistical learning, can be enhanced through explicit training. This previous study did not analyze EEG data recorded during the statistical learning exposure period. This previous study also included an additional group of participants assigned to an explicit training condition, whose data are not included here.

Stimuli

Stimuli in the exposure phase were modeled after previous auditory statistical learning studies (e.g., Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Saffran, Newport, & Aslin, 1996). This language consists of 11 syllables combined to create six trisyllabic nonsense words (*babupu*, *bupada*, *dutaba*, *patubi*, *pidabu*, and *tutibu*). Some members of the syllable inventory occur in more words than others, which produce varying transitional probabilities between the syllables within the words, as in natural language. Each nonsense word was repeated 300 times in pseudorandom order, with the restriction that the same word never occurred consecutively. Because the speech stream contained no pauses or other acoustic indications of word onsets, the only cues to word boundaries were transitional probabilities, which

were higher within words than across word boundaries (cf. Saffran et al., 1996, 1997).

A speech synthesizer (Mac text-to-speech application, female voice “Victoria”) was used to generate a continuous speech stream composed of the six trisyllabic nonsense words. To achieve more natural-sounding speech, speech synthesis technology makes use of automated techniques to produce acoustic variations in the speech output. Thus, as in speech produced by a human talker, individual tokens of a given word type in the speech stream were acoustically variable (e.g., each instance of “babupu” was not uttered in an identical manner). Descriptive statistics summarizing token durations for each of the six words are shown in Table 1. As can be seen from the table, there was considerable acoustic variability across the tokens within each word type. Across the synthesized stream, the average syllable-to-syllable latency was 235 msec ($SD = 39.8$ msec).

The speech stream was edited to include 31 pitch changes. Each pitch change represented either a 20-Hz increase or decrease from the baseline frequency (~160 Hz). Pitch changes occurred randomly, rather than systematically on certain syllables, and thus could not provide a cue for segmentation. Syllables that spanned pitch changes were excluded from EEG analysis. The detection of infrequent pitch changes was used as a cover task during the learning period, to ensure adequate attention to the auditory stimuli.

EEG event codes were sent at the onset of each syllable. The timing of syllable onsets in the continuous speech stream was determined by three trained raters using both auditory information and visual inspection of sound spectrographs, with the mean rating used. Any discrepancy > 20 msec among one or more raters was resolved by a fourth independent rater. The first 30 syllables of the stream were not coded and thus not included in the analysis to avoid auditory onset effects.

Table 1. Descriptive Statistics for the Durations of Individual Tokens within Each of the Six Words in the Artificial Language

Word	Mean Duration (msec)	Standard Deviation (msec)	Min Duration (msec)	Max Duration (msec)	Word Presentation Rate for ITPC Calculation (Hz)
<i>babupu</i>	705	54	588	887	1.4
<i>bupada</i>	749	49	645	836	1.3
<i>dutaba</i>	658	48	576	746	1.5
<i>patubi</i>	694	36	587	818	1.4
<i>pidabu</i>	711	52	595	822	1.4
<i>tutibu</i>	708	24	644	756	1.4

Token durations represent the total time between the onset of the first word to the onset of the subsequent word in the continuous speech stream.

For the recognition test phase, six nonword foils were created (batabu, bipabu, butipa, dupitu, pubada, and tubuda). The nonwords consisted of syllables from the language's syllable inventory that never directly followed each other in the speech stream, even across word boundaries. The frequency of individual syllables across words and nonword foils was matched.

Procedure

At the beginning of the experiment, participants were fitted with an elastic EEG cap embedded with electrodes. After EEG setup, participants were informed that they would listen to a speech stream of nonsense words. They were instructed that the speech stream contained occasional pitch changes and that they should detect these pitch changes using the keypad, with one button indicating a low pitch change and another indicating a high pitch change. To increase interest in the task, participants earned a small amount of additional money (12 cents) for each successfully detected pitch change. Because of technical issues, behavioral data for the pitch-detection task from two participants could not be analyzed. Overall, the remaining participants performed well on the pitch-detection task, detecting 94.6% ($SD = 4.2\%$) of the 31 pitch changes. The total exposure stream of approximately 21 min was divided into three equal blocks, and participants were given a brief break between each block.

After finishing the listening phase of the experiment, participants were informed that the nonsense language that they had just listened to was composed of individual words. They were then given a free recall task in which they were asked to recall the six words by writing them down on a piece of paper. Overall performance was very poor on this task (mode of 0 words correct across participants) and was not analyzed further.

Participants then completed a forced-choice recognition judgment task. Each trial included a word and a nonword foil. Participants gave two responses for each trial, (1) indicating which of the two sound strings sounded more like a word from the language and (2) reporting on their awareness of memory retrieval, with "remember" indicating confidence based on retrieving specific information from the learning episode, "familiar" indicating a vague feeling of familiarity with no specific retrieval, and "guess" indicating no confidence in the selection. Each of the six words and six nonword foils were paired exhaustively for a total of 36 trials. In half of the trials, the word was presented first, whereas in the other half, the nonword foil was played first; presentation order for each individual trial (whether presented first/second) was counterbalanced across participants. Participants then completed a final speeded target detection task, designed to assess implicit memory of the syllable patterns. Behavioral and EEG data for the remember/familiar/guess component of the recognition test and for the target

detection task have been analyzed previously (Batterink et al., 2015) and are not included in the current article.

Behavioral Data Analysis (2AFC Recognition Task)

For each participant (1–21) and word type (1–6), I computed a "recognition score," which represents the total number of correct trials out of six for each word. These 126 values were then analyzed as the dependent variable in the main linear mixed-effects model (described below), to examine the relationship between recognition and neural phase-locking at the item level.

In addition, a one-sample t test was used to test whether recognition performance was above chance, with 50% correct representing chance-level performance. Finally, a repeated-measures ANOVA with Word type (1–6) as a within-participant factor was used to test whether recognition performance differed across the different component words of the language.

EEG Recording and Analysis

Recording and Preprocessing

EEG during the exposure phase was recorded with a sampling rate of 512 Hz from 64 Ag/AgCl-tipped electrodes attached to an electrode cap using the 10–20 system. Recordings were made with the Active-Two system (Biosemi), which does not require impedance measurements, an online reference, or gain adjustments. Additional electrodes were placed on the left and right mastoids, at the outer canthi of both eyes, and below both eyes. Scalp signals were recorded relative to the Common Mode Sense active electrode and then rereferenced off-line to the algebraic average of the left and right mastoids.

All EEG analyses were carried out using EEGLAB (Delorme & Makeig, 2004). First, EEG data were band-pass filtered from 0.1 to 30 Hz. Sections of data in which no auditory cues were present (i.e., during breaks or pauses in the auditory stimulation) were removed from the continuous data set. Next, the data were submitted to an automatic artifact correction procedure, based on the Artifact Subspace Reconstruction algorithm developed by Mullen et al. (2015), which is designed for the removal of occasional large-amplitude noise/artifacts. Critical parameters for the implementation of this algorithm were selected conservatively based on empirical testing and previously established guidelines (Chang et al., 2018; high-pass transition = 0.25–0.75 Hz, minimum channel correlation = 0.8, line noise = "off," burst criterion = 5, window criterion = 0.25). This step resulted in the removal of the noisiest sections of data (mean = 1.4%, $SD = 1.6\%$) and interpolation of an average of 4.47 of 64 scalp electrodes ($SD = 4.46$). Next, for each participant and for each word type (1–6), epochs time-locked from -2.5 to 2.5 sec relative to word onset were extracted from the continuous data set and baseline corrected using

mean amplitude across the whole epoch, producing 126 separate data sets.

Neural Phase-locking Analysis

For each of the 126 item-level data sets, neural phase-locking was quantified by measuring intertrial phase coherence (ITPC) across all epochs. ITPC is a measure of event-related phase locking. ITPC values range from 0, indicating purely non-phase-locked activity, to 1, indicating strictly phase-locked activity. A significant ITPC indicates that the EEG activity in single trials is phase-locked at a given time and frequency, rather than phase-random with respect to the time-locking experimental event. ITPC was computed using a continuous Morlet wavelet transformation from 0.3 to 6.0 Hz via the *newtimef* function of EEGLAB. Wavelet transformations were computed in 0.1-Hz steps with one cycle at the lowest frequency (0.3 Hz) and increasing by a scaling factor of 0.5, reaching 10 cycles at the highest frequency (6.0 Hz). A scaling factor of 0.5 indicates that the width of the wavelet used for the highest frequency is half (0.5) the width of the wavelet used at the lowest frequency (Dickter & Kieffaber, 2014), allowing better frequency resolution at higher frequencies than wavelet approaches using constant cycle lengths (Delorme & Makeig, 2004). Two hundred output times were computed for each frequency; time points spanned an interval from -643 to 643 msec, separated by an average of 6.5 msec. For each word type, a specific frequency of interest was selected, corresponding to the mean token duration (range across word types = 1.3–1.5 Hz; see Table 1 for specific frequency values by word type). Note that the decision to select the most appropriate frequency for each word type, rather than using a constant frequency across word types, had no major impact on the results, as the ITPC values at the different frequency bins of interest (1.3, 1.4, and 1.5 Hz) were highly correlated (mean $r = .983$). After artifact correction, an average of 268 trials contributed to each item-level data set (range = 147–301, $SD = 26.5$).

For each item, ITPC values were averaged from word onset to 576 msec, which corresponds to the minimum token duration across all words (see Table 1), to capture neural phase-locking across the full duration of each word. All 64 scalp electrodes were used in this calculation because of the widespread nature of the effect (see Results). These 126 average item-level ITPC values (henceforth referred to as “item-level ITPC”) were used as predictors in the main mixed-effects model.

Statistical Testing of Item-level ITPC

To test whether neural oscillations at the word presentation rate specifically track word identities, observed item-level ITPC values were compared against a null distribution of ITPC values. The null ITPC distribution—representing the null hypothesis that item-level ITPC is not higher to

words than to pseudorandomly selected syllable triplets—was estimated by creating “surrogate” data sets for each participant and each word type. To control for word duration (both mean and standard deviation of durations of words in the actual data sets), surrogate data sets were created through an item-level matching procedure. For each “true” word in an actual item-level data set, a randomly selected triplet was selected for assignment to the surrogate data set, based on the following criteria: (1) The triplet was not simply another repetition of the true word, and (2) The triplet had not already been selected previously for assignment to the surrogate data set. The triplet with the closest duration to the true word was then selected from the pool of all candidate triplets that met these two criteria. In cases where more than one candidate triplet was an equally close match, the surrogate word was selected randomly from the closest candidates. Thus, this procedure ensured that, within a given item-level data set, onsets for each surrogate word occurred pseudorandomly across actual syllable positions and word identities. For each item-level surrogate data set, ITPC at the corresponding word presentation rate was then computed as in the original analysis. This entire procedure was performed 100 times, producing a surrogate ITPC distribution of 100 group-averaged values for each word type. The critical ITPC value was defined as the value in this surrogate distribution corresponding to the 95th percentile ($p < .05$). If the observed item-level ITPC values within a given condition (i.e., word type and electrode) exceeded this critical value, ITPC was considered significant, providing evidence of word identity-specific phase-locking that exceeds phase-locking to randomly selected syllable triplets in the stream.

Statistical Testing of Item-level ITPC by Word Type

A linear mixed-effects model was used to test whether item-level ITPC differs as a function of word type. The model included word type (1–6) as a fixed effect, participant intercept as a random effect, and item-level ITPC as the dependent variable, using maximum likelihood estimation. In addition, to explore potential acoustic factors driving any potential differences in ITPC between word types, a separate linear mixed-effects model tested whether mean word duration and variability in word duration across utterances predicted ITPC. This model included the mean token duration and standard deviation of token durations for each word as fixed effects (see Table 1 for values), participant intercept as a random effect, and item-level ITPC as the dependent variable, using maximum likelihood estimation.

Statistical Testing of Relationship between Item-level ITPC and Recognition

My main hypothesis was that neural phase-locking to each word in the speech stream should predict knowledge of

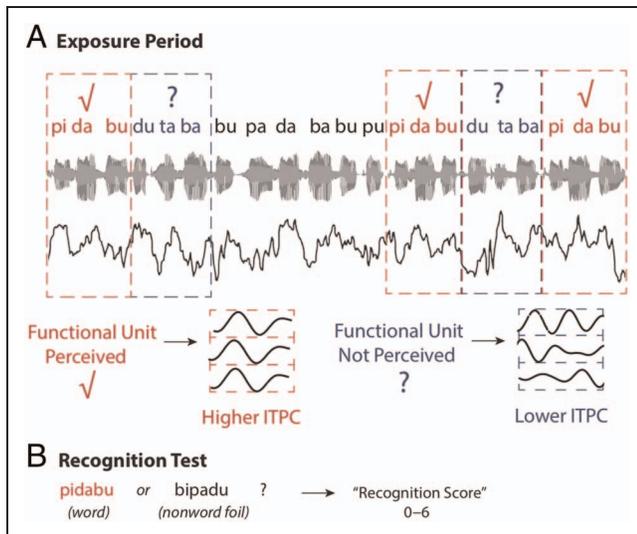


Figure 1. Summary of paradigm and analysis approach. (A) During the exposure period, a continuous speech stream made up of six repeating nonsense words was presented while listeners' EEG was recorded. If statistical learning of a given component word occurs, acoustically variable tokens should be perceived as functionally equivalent units and thus elicit more similar brain responses across word presentations. This similarity across word presentations is quantified by ITPC. (B) Knowledge of each of the six component words was assessed in a subsequent 2AFC recognition test. Across 36 trials, each word is pitted against six nonword foils, such that the maximum possible recognition score is 6.

that word, as assessed during the forced-choice recognition task. The underlying logic here is that, to the extent that a word is successfully "learned," acoustically variable tokens should be segmented from the continuous speech stream and perceived as a cohesive, functional unit. These units should in turn be tracked by neural oscillations at the word presentation rate, with better learned words eliciting stronger neural phase-locking across word tokens (Figure 1). Thus, ITPC should be higher to words that are better recognized compared to words that are more poorly recognized.

A linear mixed-effects model was used to test whether recognition differs as a function of item-level ITPC. To account for the potential effect of acoustic factors on word learning, an initial model was run with maximum likelihood estimation including token variability (standard deviation of duration across tokens; Table 1), mean token duration (Table 1), and item-level ITPC as fixed effects and recognition score as the dependent variable. The Wald Z statistic was used to estimate variance at the participant level and to test whether a random intercept for participant should be included in the model, with $p > .05$ indicating that a random effect is needed (Seltman, 2012). Because the Wald Z test was not significant (Wald $Z = 0.97$, $p = .33$), participant was not included as a random intercept in the model.

As discussed in the Introduction, one theoretical possibility is that an individual's average neural phase-locking

to all words in the language—indexing a general sensitivity to temporal patterns—could entirely account for the previously demonstrated relationship between phase-locking (or neural entrainment) and statistical learning outcomes (Batterink & Paller, 2017, 2019; Buiatti et al., 2009). I therefore tested whether item-level ITPC accounts for item-level word recognition over and above an individual's average ITPC values. Each participant's average word-rate ITPC (henceforth referred to as "average ITPC") was computed by averaging the six item-level ITPC values included in the original analysis. To directly compare the predictive value of item-level ITPC to average ITPC in item-level recognition performance, I ran a linear mixed-effects model with recognition score as the dependent variable and both item-level ITPC and average ITPC as fixed effects. In addition, I tested a linear mixed-effects model with average ITPC as a fixed effect and compared this model against the winning item-level ITPC model (as described above) using Bayesian Information Criterion (BIC). Finally, Spearman's correlation was used to test the relationship between average ITPC and overall recognition performance across individuals, as a conceptual replication of previous findings (Batterink & Paller, 2017, 2019; Buiatti et al., 2009).

Time Course Analysis of ITPC

To examine the time course of ITPC over the course of exposure to the artificial language, a fine-grained time course analysis was carried out. Given previous evidence that statistical learning in the context of artificial speech segmentation paradigms occurs within 2 min in infants (Saffran et al., 1996) and that learning occurs most rapidly during early stages of exposure and follows a logarithmic curve (Choi, Batterink, Black, Paller, & Werker, 2020; Siegelman, Bogaerts, Elazar, Arciuli, & Frost, 2018), I expected the most reliable changes in ITPC to occur relatively early on during exposure. Thus, the time course analysis was restricted to the first block of exposure (corresponding to roughly 7 min or ~800 total word presentations). One participant was excluded from this analysis, as part of the first exposure block for this participant was not recorded because of experimenter error.

For each word type, single-trial wavelet decompositions were computed and stored as complex coefficients using EEGLAB's *ifdata* output variable, using the same parameters as in the original analysis. To improve the signal-to-noise ratio of single-trial estimates, every 10 consecutive trials were then grouped together using a moving window approach (Trials 1–10, 2–11, 3–12, etc.). Following our previous approach (Choi et al., 2020), ITPC was then computed for each group of 10 consecutive trials. The key prediction here is that the phase of neural oscillations at the word rate should become more consistent if word learning occurs over the course of exposure, such that ITPC at the word rate should increase over time (see Batterink & Paller, 2017,

2019). A linear mixed-effects model was used to test whether item-level ITPC significantly increased over the course of the first block. The model included the word presentation number of the first trial in a given group of 10 trials (i.e., number of presented items within each word type; 1–80), word type (1–6), and the interaction between word type and number of word presentations as fixed factors; participant intercept as a random effect; and ITPC values for each group of 10 consecutive trials as the dependent variable, using maximum likelihood estimation.

Temporal Dynamics of ITPC

Finally, to explore the temporal dynamics of neural phase-locking across the unfolding of a word over time, a separate analysis was run in which item-level ITPC was computed only within the corresponding word presentation rate bins of interest (i.e., 1.3–1.5 Hz; see Table 1), as well as at the average syllable presentation rate (4.3 Hz). Discarding lower frequencies from the analysis allowed for computing ITPC across a longer time window (i.e., –1570 to 1570 msec, compared to –643 to 643 msec in the original analysis). All other parameters were matched to the original analysis parameters. The time course of ITPC at both the word and syllable frequencies was plotted across time to visualize the temporal trajectory of entrainment. A running paired t test across all post-word-onset time intervals was used to test at which time points ITPC values exceeded the prestimulus value (i.e., the value occurring immediately before word onset).

All p values are from two-tailed tests with an alpha of .05. Greenhouse–Geisser corrections are reported for factors with more than two levels.

RESULTS

Behavioral Results (2AFC Recognition Test)

Recognition performance was significantly above chance across participants (mean = 59.3%, 11.7%), $t(20) = 3.62$, $p = .002$, providing evidence of successful word segmentation because of statistical learning. Recognition performance varied significantly across the six component words of the language words (word type: $F(5, 100) = 2.76$, $p = .049$; Figure 2B), indicating that some words were learned better and recognized at higher levels than others.

EEG Results (Exposure Period)

Item-level ITPC Shows Peaks at Word and Syllable Frequencies

The average of all item-level ITPC values is plotted as a function of frequency in Figure 2A. As shown in the figure, there is a clear ITPC peak in the frequency range

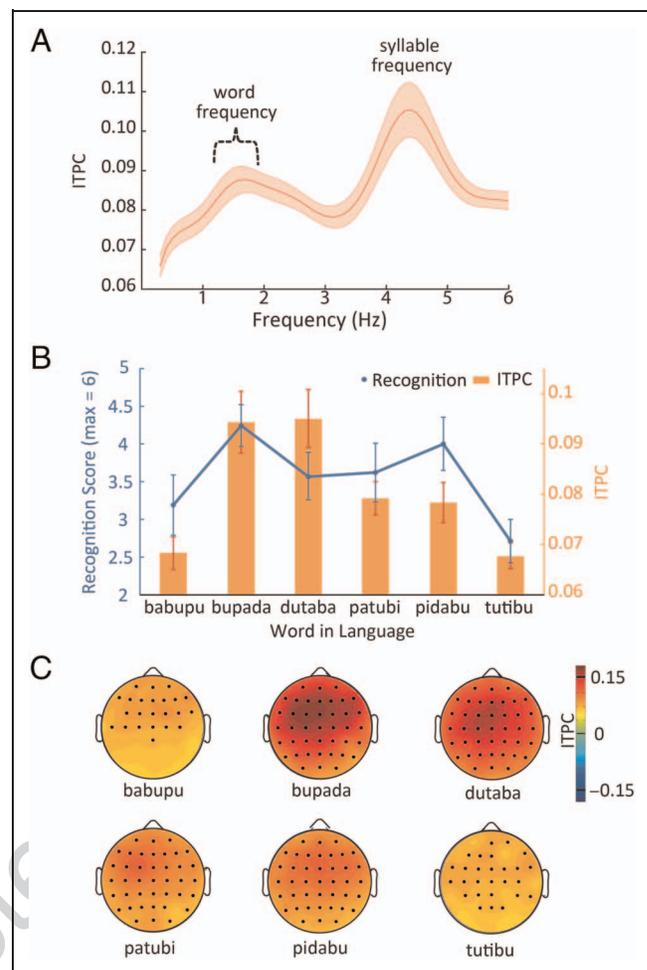


Figure 2. Quantification of neural phase-locking to component words in the speech stream at the word presentation rate. (A) ITPC as a function of frequency, averaged across the six component words of the speech stream. Two peaks are observed, corresponding to the word rate (range = 1.3–1.5 Hz across words) and the syllable rate (mean = ~4.3 Hz). Shaded regions represent *SEM*. (B) ITPC and recognition score as a function of word type, across all participants. Across participants, words that showed higher ITPC values were also better recognized. (C) Distribution of word-rate ITPC across the scalp for each of the six words in the language. Electrode locations with significant word-rate ITPC as contrasted against the shuffled surrogate distribution are denoted with a black circular marker.

corresponding to the average word rate of the speech stream, providing evidence of word-level neural phase-locking. A second peak between 4 and 5 Hz can also be seen, corresponding to the average syllabic rate of the speech stream.

Item-level ITPC Shows Significant Tracking of Word Identities

Across all electrodes, item-level ITPC within each word type was highly significant when tested against the null distribution of ITPC values, which reflects phase-locking to randomly selected syllable triplets that were equated

for duration (all p s < .01, with observed ITPC values exceeding the 99th percentile of the surrogate distribution). This result indicates that neural oscillations at the word rate phase-lock to words in the speech stream over and above phase-locking to random syllable triplets, providing evidence that individual-item ITPC tracks word identities.

Item-level ITPC Varies as a Function of Word Type

Item-level ITPC varied significantly as a function of word type, with some words in the language eliciting significantly higher ITPC than other words (effect of word type on item-level ITPC: $F(5, 30) = 10.6, p < .001$). As shown in Figure 2B, word types that elicited greater ITPC also appeared to show higher recognition performance. Effects of word type and item-level ITPC on recognition performance were statistically tested through linear mixed-effects modeling, with results described below (under Item-level ITPC Predicts Item-level Word Recognition section). In addition, item-level ITPC varied significantly as a function of variability in token duration (effect of standard deviation of token duration: $F(1, 40) = 8.59, p = .006$) as well as average token duration, $F(1, 48) = 5.36, p = .025$. Higher ITPC values were associated with higher duration variability and shorter mean durations across tokens.

The distribution of ITPC across the scalp for each word type is plotted in Figure 2C. As shown in the figure, ITPC values at a large majority of individual electrodes reached statistical significance when tested independently against the null distribution ($p < .05$). Overall, the distribution of ITPC across the scalp was relatively widespread, with a frontocentral maximum, consistent with an auditory response.

Item-level ITPC Predicts Item-level Word Recognition

Critically, supporting my main prediction, the full three-predictor model indicated that item-level ITPC significantly predicted subsequent item-level word recognition, $F(1, 83) = 8.60, p = .004$. Figure 3A shows item-level ITPC as a function of recognition score (1–6). The number of items that comprise each recognition score “bin” are as follows: score 1 = 13; score 2 = 22; score 3 = 21; score 4 = 28; score 5 = 24; and score 6 = 16.

Word variability in token duration also significantly predicted recognition performance (standard deviation of token duration: $F(1, 57) = 5.23, p = .026$). Mean token duration was not a significant predictor in the model, $F(1, 42) = 2.27, p = .14$. Together, these results indicate that item-level ITPC and word variability both independently and positively predicted word learning (ITPC parameter estimate = 16.9, 95% CI [5.44, 28.4], $SE = 5.76$; word variability parameter estimate = 0.027, 95% CI [0.0034, 0.051], $SE = 0.012$).

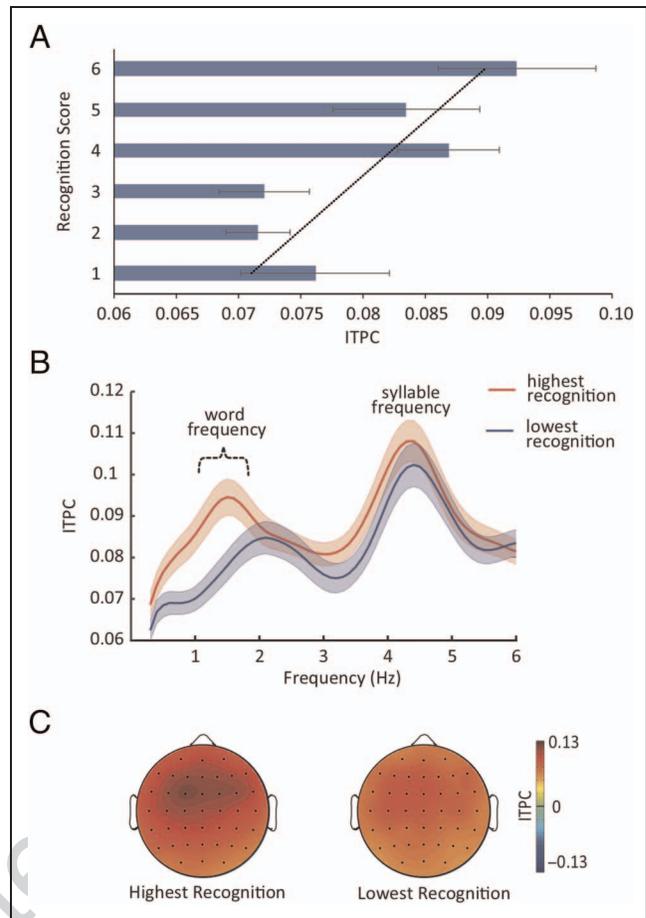


Figure 3. Relationship between ITPC and recognition score pooling across words and participants. (A) ITPC predicts recognition score at the item level. (Here, an “item” is defined as the combined collection of individual instances of the same word at the single participant level; thus, each participant contributes six items in total to the analysis.) Items are binned by recognition score (1–6); note that a recognition score of 0 is not plotted because this bin contains only two items in total. The dotted line represents the best linear fit between ITPC and recognition score. Error bars represent SEM. (B) ITPC as a function of frequency for high and low recognition items, divided by median split for data visualization purposes. High recognition items ($n = 68$) have a recognition score of 4–6, and low recognition items ($n = 58$) have a recognition score of 0–3. Shaded regions represent SEM. (C) Distribution of item-level ITPC across the scalp for high and low recognition items.

The Wald Z test was not significant (Wald $Z = 0.97, p = .33$), indicating that unmeasured variance at the individual participant level did not significantly contribute to the model. When item-level ITPC was included as a single predictor in the model, the random effect of participant again did not significantly predict word recognition (Wald $Z = 0.52, p = .60$).

For visualization purposes, ITPC for the items with the highest recognition score (4–6; 68 items) and the lowest recognition score (0–4; 58 items), divided by median split, are plotted as a function of frequency and scalp distribution in Figure 3B and 3C.

Item-level ITPC Predicts Item-level Word Recognition Better than Average ITPC

When both item-level ITPC and average ITPC (at the participant level) were included as predictors in the model, only item-level ITPC significantly predicted item-level word recognition (item-level ITPC: $F(1, 91) = 5.45, p = .022$; overall ITPC = $F(1, 102) = 0.37, p = .55$). This result indicates that item-level ITPC accounts for item-level word recognition over and above an individual's average ITPC value. A model including average ITPC as a single predictor did not significantly predict item-level word recognition, $F(1, 115) = 8.78, p = .004$; however, this model performed more poorly than the comparison model in which item-level ITPC was used as a predictor (BIC for model with overall ITPC = 504.3; BIC for model with item-level ITPC = 500.7).

Across individuals, average ITPC predicted overall recognition performance (Spearman's $r = .44, p = .048$), which conceptually replicates previous reports of correlations between neural entrainment and statistical learning at the individual level (Batterink & Paller, 2017, 2019; Buiatti et al., 2009). In summary, overall neural phase-locking at the individual level predicts statistical learning performance as measured by word recognition but is not as good a predictor as item-level neural entrainment.

Item-level ITPC Increases over the First Block of Exposure

As shown in Figure 4, across all word types, item-level ITPC showed a significant increase over the first block of exposure (effect of number of word repetitions: $F(1, 8854) =$

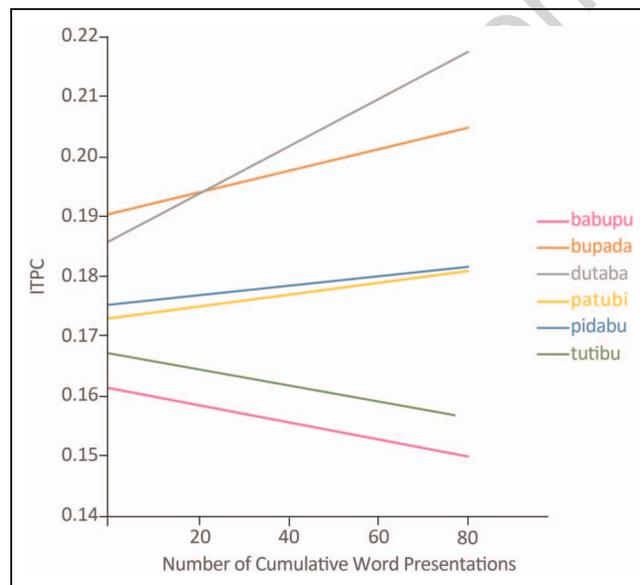


Figure 4. Modeled progression of word-rate ITPC by word as a function of exposure. Values are based on parameter estimates of fixed effects in the linear mixed-effects model within the first block (~7 min) of exposure. ITPC showed a significant increase as a function of exposure across all words, reflecting online learning. The change in ITPC over exposure varied significantly by word.

$7.32, p = .007$). The item-level ITPC slope over time varied significantly as a function of word type (Word Type \times Number of Word Repetitions: $F(5, 8852) = 8.08, p < .001$). Follow-up analyses indicated that the two words associated with the highest recognition performance (i.e., *bupada* and *dutaba*) both independently showed significant increases in ITPC over exposure (both $ps < .010$), whereas the two words with the lowest recognition performance (*babupu* and *tutibu*) showed negative or marginally negative ITPC decreases over the first block of exposure ($p = .023-.053$).

Temporal Dynamics of Item-level ITPC Show Early Onset and Relatively Late Peak

As shown in Figure 5A, item-level ITPC at the word presentation rate peaked at approximately 420 msec after word

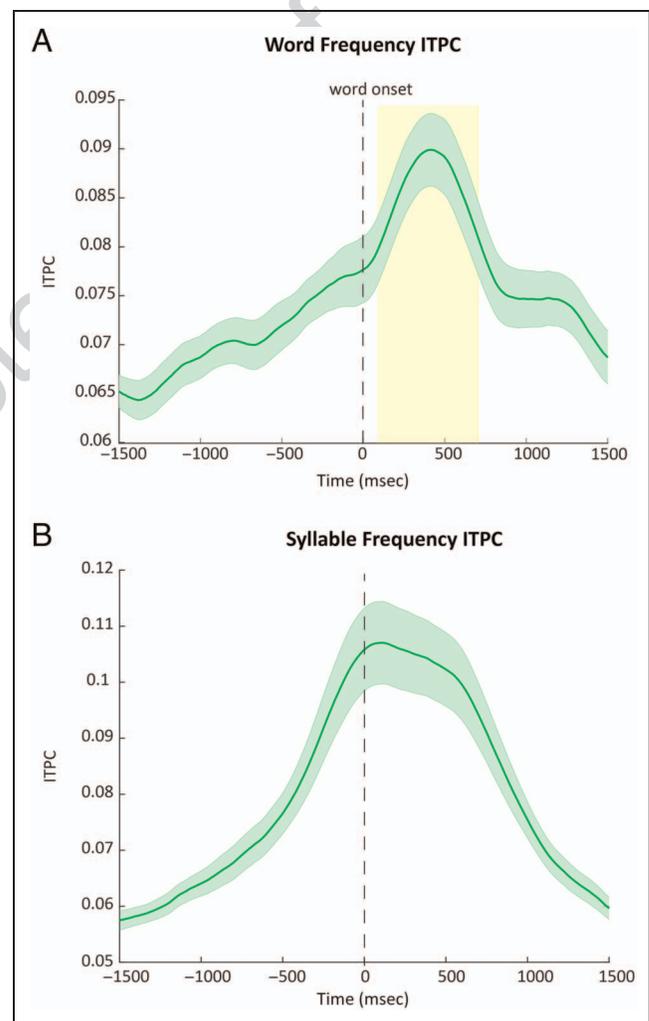


Figure 5. Temporal dynamics of item-level ITPC over the course of word presentation, relative to word onset. (A) Temporal evolution of item-level ITPC at the word rate. ITPC peaks approximately 418 msec after word onset. Shaded green regions around the line represent SEM. The shaded yellow region (spanning 86–701 msec) shows time intervals when ITPC significantly exceeds prestimulus baseline ITPC. (B) Temporal evolution of item-level ITPC at the syllable rate, demonstrating a much earlier peak (~110 msec).

onset and significantly exceeded the prestimulus value from 86 to 701 msec ($p < .05$). ITPC during the prestimulus interval also showed a steady increase over time, which may be attributed to reduced temporal jitter relative to word onset over the prestimulus interval. In contrast, ITPC at the syllable presentation rate showed a much earlier peak, at approximately 110 msec (Figure 5B); however, these peak values were not significantly different from the prestimulus value. Again, ITPC during the prestimulus interval increased strongly over time, reflecting reduced temporal variability relative to syllable onset.

DISCUSSION

The current findings demonstrate a robust association between neural phase-locking and subsequent linguistic knowledge at the individual word level, providing novel evidence that phase-locking to “hidden” linguistic units in continuous speech delineates perceived linguistic boundaries on a word-by-word basis. Using a classical statistical learning task, participants were passively exposed to a continuous speech stream made up of repeating nonsense words. After the exposure period, learners’ memory of the nonsense words was assessed using an explicit 2AFC recognition test. Words that elicited stronger neural phase-locking during exposure, as quantified by ITPC, were recognized at higher rates on the subsequent memory test. Neural phase-locking at the word rate also significantly increased over the first block of the exposure period. These results indicate that neural phase-locking over repeated word presentations reflects the discovery, encoding, and perception of individual linguistic items acquired as a result of statistical learning.

These findings support the hypothesis that continuous speech is segmented into meaningful functional units through nested, hierarchically organized neural oscillations (Gross et al., 2013; Giraud & Poeppel, 2012; Peelle & Davis, 2012). According to these models, speech is parsed into meaningful units by neural oscillations operating across a range of specific frequencies that match the rhythms of relevant linguistic components (e.g., phonemes, syllables, words, and phrases). Consistent with this idea, I found that neural phase-locking is higher to words that are successfully recognized, compared to those that are not. Presumably, words with higher recognition performance were perceived as functional units and tracked by oscillatory activity at the matching word rate. In contrast, words with poor recognition performance were processed merely as a sequence of unrelated syllables rather than as a word unit and thus were not tracked by corresponding word-rate oscillations.

Neural Entrainment and Word Knowledge May Interact Bidirectionally

The current results are correlational in nature and cannot directly disentangle the causality between neural phase-

locking and word learning. However, on the basis of previous findings, I propose that there may be bidirectional interactions between phase-locking and linguistic knowledge: (1) Neural phase-locking to underlying patterns may influence the formation of high-level linguistic representations, and (2) word representations may exert a top-down influence on phase-locking of ongoing oscillations. The first idea—that modulation of neural phase may influence word learning—is supported by several recent transcranial alternating current stimulation studies (Riecke, Formisano, Sorger, Başkent, & Gaudrain, 2018; Wilsch, Neuling, Obleser, & Herrmann, 2018; Zoefel, Archer-Boyd, & Davis, 2018). By directly manipulating neural oscillations, these studies demonstrated that the phase lag between brain and speech rhythms influenced the neural responses to intelligible speech in superior temporal gyrus (Zoefel et al., 2018) as well as speech comprehension (Riecke et al., 2018; Wilsch et al., 2018). These results indicate that the phase alignment between neural oscillations and an ongoing speech signal plays a causal role in high-level speech processing and, by extension, could also (in principle) influence speech segmentation and statistical word learning.

A major mechanism underlying neural phase-locking to speech is phase-resetting of low-frequency oscillations in the auditory cortex to “acoustic landmarks” in the speech envelope, such as speech onsets or sharp acoustic transients (Doelling, Arnal, Ghitza, & Poeppel, 2014; Gross et al., 2013). Certain words may thus be more learnable because they contain acoustic features that elicit stronger or more consistent phase-resetting across word presentations. Consistent with this idea, in this study, I found that some words elicited higher ITPC than others and that word-level differences in ITPC accounted for word-level differences in recognition (Figure 2B). Furthermore, these word-level ITPC differences emerged very early on and were relatively stable across exposure; words that showed high ITPC values during early learning continued to elicit relatively higher ITPC values throughout the first block of exposure (Figure 4). These findings suggest that “baseline” acoustic features influence phase-locking and that degree of phase-locking predicts whether a given word is more learnable. Over multiple word exposures, phase-locked oscillations at the word frequency could mediate the binding of syllables into larger temporal chunks, thereby supporting word learning. This idea is consistent with the proposal that neural entrainment functions to align phases of neural excitability to repeated temporal patterns, providing a mechanism for identifying specific patterns in upcoming sensory input (Schroeder & Lakatos, 2009).

A second possibility, which is not mutually exclusive, is that high-level word knowledge has a top-down influence on neural phase-locking. This idea is supported by the present finding that ITPC significantly increased over exposure, reflecting the gradual acquisition of word knowledge that in turn may facilitate predictive processing. This

significant increase in phase-locking over time cannot be accounted for by bottom-up factors alone, given that the stimulus stream did not differ systematically over exposure, and replicates previous findings showing that neural phase-locking to words (or phrases) in an artificial language increases gradually over the course of exposure (Choi et al., 2020; Batterink & Paller, 2017, 2019; Getz et al., 2018). Together, these results converge with mounting evidence that neural phase-locking is critically modulated by top-down processes such as selective attention and expectations (e.g., Rimmele, Zion Golumbic, Schröger, & Poeppel, 2015; Horton, D’Zmura, & Srinivasan, 2013; Zion Golumbic et al., 2013; Ding & Simon, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008). Mechanistically, a recent MEG study demonstrated that top-down signals from frontal brain areas causally influence the phase of speech-coupled oscillations in auditory cortex, enhancing speech-brain coupling (Park, Ince, Schyng, Thut, & Gross, 2015). The idea that neural phase-locking is influenced by top-down processing is also compatible with theoretical proposals that neural phase-resetting provides an instrument for sensory selection by enabling phases of higher neural excitability to align with important stimulus events (Thut, Miniussi, & Gross, 2012; Schroeder & Lakatos, 2009). In the context of speech segmentation, high-level word knowledge enables predictions to be made about upcoming syllables. In turn, these top-down predictions may function to optimally align ongoing neural oscillations with the most important or informative moments of the speech signal, acting to increase sensitivity to relevant acoustic cues and thereby facilitating speech processing (Pelle & Davis, 2012).

The current results also suggest that bottom-up acoustic factors may interact with statistical learning and top-down knowledge, with words that are the most initially “trackable” also benefiting the most from exposure and showing continual gains in learning (Figure 5). The time course analysis demonstrated that words with the highest ITPC estimates at the beginning of the exposure period (i.e., *bupada* and *dutaba*) showed significant increases in ITPC over exposure. In contrast, words with low initial-phase locking (i.e., *babupu* and *tutibu*) did not show increases in neural phase-locking over this period, suggesting that words that are not initially trackable (as measured by baseline ITPC estimates) do not benefit from exposure. In summary, both bottom-up and top-down factors appear to contribute to the observed relationship between item-level ITPC and subsequent word learning, and furthermore, these different mechanisms are likely to interact with one another.

Word Variability across Utterances Influences Both ITPC and Word Learning

At the behavioral level, I found a significant impact of word type on recognition performance, indicating that some words were more easily learned than others. This

finding aligns well with previous findings that language-specific knowledge influences linguistic statistical learning, with words that more closely follow the phonotactic regularities of a participant’s native language being learned better (Siegelman, Bogaerts, Kronenfeld, & Frost, 2018; Finn & Hudson Kam, 2015). A more novel, unexpected finding was that words with more variable durations in the stream were learned better compared to words that had less variability. Some caution is warranted in interpreting this result, given that there were only six words in the language and that a full exploration of acoustic differences between words is beyond the scope of this article. Nonetheless, this finding is consistent with prior evidence showing that variability facilitates speech learning and generalization to novel instances (e.g., Bradlow & Bent, 2008; Singh, 2008; Clopper & Pisoni, 2004; Greenspan, Nusbaum, & Pisoni, 1988). For example, the perception of new sentences produced with synthetic speech improves when participants are exposed to a larger set of training stimuli compared to a restricted set (Greenspan et al., 1988). Within the context of statistical learning, Gómez (2002) demonstrated that infants’ and adults’ learning of nonadjacent dependencies (e.g., *pel-X-jic*) depends on sufficient variability, occurring only when the middle, nonpredictive element of the dependency (i.e., X) is drawn from a sufficiently large pool. Taken together, these results indicate that exposure to a greater variety of exemplars allows learners to better ignore irrelevant features and identify the most predictable, informative, or invariant structures in a stimulus stream. In the current study, words with greater variability across utterances may promote the acquisition of more abstract word representations, as opposed to more specific, stimulus-based, acoustic representations (Vouloumanos, Brosseau-Liard, Balaban, & Hager, 2012). This in turn may facilitate generalization and better recognition performance when the same word presented in a new context (i.e., in isolation during the 2AFC recognition task, rather than embedded in a continuous speech stream as during the exposure phase).

Word variability in duration across utterances also influenced ITPC, with greater word variability predicting higher ITPC values. This finding does not follow from a straightforward bottom-up mechanistic account of neural phase-locking, which would predict that ITPC should be higher to items that have a more stable (i.e., less variable) duration across presentations. Rather, this finding suggests that greater word variability facilitates word learning, which in turn leads to stronger phase-locking to the embedded words, providing additional support for top-down influences on neural phase-locking.

Neural Mechanisms Underlying Statistical Learning

On a more specific note, the current findings also provide new insights into the neural mechanisms that underlie statistical learning in the context of word

segmentation, extending previous work in this area. As described in the Introduction, prior studies have shown that neural tracking of repeating nonsense words predicts statistical learning performance on subsequent behavioral tests at the individual level; participants who show stronger neural entrainment responses to the underlying linguistic structures perform better on subsequent learning tests (Batterink & Paller, 2017, 2019; Buiatti et al., 2009). The current results conceptually replicate these results, demonstrating that average ITPC during learning predicts subsequent overall recognition performance across individuals. At the same time, the current findings go beyond a demonstration of interindividual correlations, showing that item-level neural entrainment predicted item-level recognition more strongly than individual-level average entrainment. Furthermore, the effect of individual participant did not significantly account for variability in item-level word recognition when item-level ITPC was accounted for.

Taken together, these results indicate that neural phase-locking in the context of language learning primarily reflects the discovery and perception of individual items in the language inventory, rather than indexing more general interindividual differences that would operate similarly across all items, such as the tendency to “spontaneously synchronize” one’s behavior to external stimuli (Assaneo et al., 2019). In other words, it appears that the previously documented relationship between neural entrainment and statistical learning performance (Batterink & Paller, 2017, 2019; Buiatti et al., 2009) primarily reflects the specific content of linguistic knowledge and can be accounted for by better learners’ higher rates of word learning.

ITPC results also hint that learners may have engaged in a suboptimal parsing strategy for words that were not successfully learned (i.e., “lowest recognition items” with a score of 50% accuracy or below on the 2AFC test—a recognition score). As shown in Figure 3B, low recognition items show a peak at approximately 2.1 Hz, which corresponds to the average bigram rate in the speech stream. This finding suggests that poorly learned items may be parsed as bigrams on some proportion of trials. For example, for a triplet such as “babupu,” participants may segment the bigram “babu” on some occurrences, “bupu” on other occurrences, and neither possible bigram on still other occurrences. Across all trials, this would produce a weak signature of bigram tracking. Because overall ITPC values corresponding to the bigram presentation rate are similar across highest recognition and lowest recognition items (Figure 3B), it appears that some (relatively weak) degree of erroneous bigram parsing also occurs for better learned words. This finding highlights that ITPC at the word rate specifically is a signature of statistical word learning and that phase-locking at other low frequencies more generally (< 10 Hz) does not distinguish between better learned and poorly learned items.

Temporal Trajectory of Item-level Neural Entrainment

The temporal dynamics of ITPC provide additional insights into the neural mechanisms that support statistical learning of novel words. As a given word unfolds, ITPC showed a steep increase beginning immediately after word onset (see Figure 5). The rapid nature of this effect converges with previous demonstrations that word onsets modulate ongoing neural responses very quickly. For example, a recent MEG study modeled neural responses to continuous narrative speech and found a highly significant effect of word onsets with a peak latency of 103 msec (Brodbeck, Hong, & Simon, 2018). This finding was interpreted as evidence that word boundaries are detected essentially as they occur, rather than after incorporating cues occurring subsequent to word onset. Similarly, an ERP study found an early sensory-related N100 effect to onsets of nonsense words in continuous speech (Sanders, Newport, & Neville, 2002). This effect was observed only in learners who showed the strongest behavioral evidence of word knowledge, suggesting that high-level linguistic knowledge is a prerequisite for this early response. In the context of neural entrainment frameworks (e.g., Schroeder & Lakatos, 2009), word onsets may represent privileged sensory events, as syllables occurring at the beginning of a word are relatively information-rich and highly predictive of subsequent syllables. Successful word learning may therefore be accompanied by the rapid alignment of neural oscillations to these informative word onsets.

Although showing a rapid increase soon after word onset, the neural entrainment response did not peak until ~ 420 msec, which coincides to roughly 200 msec after the onset of the second syllable. This peak was followed by a decrease in entrainment, which statistically reached baseline levels by 701 msec, very close to the mean word duration of 704 msec (see Table 1). A similar neural tracking trajectory was reported by Ding et al. (2016; see their Figure 4), who found that neural activity reached its peak during the second word of artificial grammar phrases and then progressively decreased with each additional word in the phrase. Taken together, these findings indicate that ITPC tracks the entire time course of a higher-level unit, rather than being a transient response occurring only at unit boundaries (cf. Ding et al., 2016). It is also interesting to note that the observed peak in ITPC is similar to the typical latency of the N400 effect (Kutas & Federmeier, 2011). This suggests that neural tracking of a given word may decline once the word is processed to the point of recognition.

In contrast to ITPC at the word presentation rate, ITPC at the syllable rate peaked very soon after word onset (~ 110 msec; Figure 5B). Overall, the trajectory of neural entrainment at the syllable rate resembles a symmetrical, steep parabolic curve centered just after word onset, consistent with a sensory-evoked response that is not strongly modulated by high-level knowledge.

Conclusions

In summary, the main finding of the study is that neural phase-locking accompanies an individual's subjective perception of an individual word in continuous speech, as acquired in real time during statistical learning. These results indicate that the association between neural phase-locking and statistical learning is not limited to perfectly isochronous syllable sequences (e.g., Batterink & Paller, 2017, 2019; Getz et al., 2018; Ding et al., 2016; Buiatti et al., 2009) but is generalizable to continuous speech containing nonidentical word tokens. The demonstration that neural phase-locking is sensitive to recognition strength of individual words opens up the possibility of tracking the contents of learning in real time. For example, by monitoring the EEG of language learners exposed to a continuous stream of foreign language input, it may be possible to predict which words have been successfully learned and which words require additional training. This neural phase-locking approach may also be applied to investigate other aspects of language that involve the concatenation of smaller linguistic elements into larger units, such as the learning and processing of grammatical rules, as well as perceptual aspects of language acquisition, such as phonetic category learning. Thus, new applications of this approach may significantly advance our understanding of other neural mechanisms underlying language acquisition and processing.

Acknowledgments

The data in this article were collected in the laboratory of Dr. Ken A. Paller, and I am grateful for his support. This work was supported by the National Institutes of Health (Grants T32 NS 047987 and F32 HD 078223).

Reprint requests should be sent to Laura Batterink, Department of Psychology, Brain and Mind Institute, Western University, London, ON N6A 2B3, Canada, or via e-mail: lbatter@uwo.ca.

REFERENCES

- Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*, 627–632.
- Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45.
- Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71.
- Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & Memory*, *22*, 544–556.
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707–729.
- Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*, 3976–3983.
- Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, *44*, 509–519.
- Chang, S. W., McGuire, J. F., Walkup, J. T., Woods, D. W., Scahill, L., Wilhelm, S., et al. (2018). Neurocognitive correlates of treatment response in children with Tourette's disorder. *Psychiatry Research*, *261*, 464–472.
- Choi, D., Batterink, L., Black, A. K., Paller, K., & Werker, J. F. (2020). Prelingual infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. <https://doi.org/10.31234/osf.io/fuqd2>.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*, 207–238.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.
- Dickter, C. L., & Kieffaber, P. D. (2014). *EEG methods for the psychological sciences*. Thousand Oaks, CA: Sage.
- Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience*, *11*, 481.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.
- Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences, U.S.A.*, *109*, 11854–11859.
- Doelling, K. B., Arnal, L. H., Ghizva, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768.
- Finn, A. S., & Hudson Kam, C. L. (2015). Why segmentation matters: Experience-driven segmentation errors impair “morpheme” learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1560–1569.
- Getz, H., Ding, N., Newport, E. L., & Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition*, *181*, 135–140.
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.
- Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 421–433.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*, e1001752.
- Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*, 3082–3093.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.
- Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*, 2553–2567.

- Nozaradan, S., Peretz, I., & Keller, P. E. (2016). Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports*, *6*, 20612.
- Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*, 10234–10240.
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*, 1649–1653.
- Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.
- Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*, 161–169.
- Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, *68*, 144–154.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*, 606–621.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.
- Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*, 700–703.
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*, 9–18.
- Seltman, H. J. (2012). *Experimental design and analysis*. Carnegie Mellon University.
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213.
- Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining “learning” in statistical learning: What does an online measure reveal about the assimilation of visual regularities? *Cognitive Science*, *42*, 692–727.
- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*, 833–870.
- Thut, G., Miniussi, C., & Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Current Biology*, *22*, R658–R663.
- Vouloumanos, A., Brosseau-Liard, P. E., Balaban, E., & Hager, A. D. (2012). Are the products of statistical learning abstract or stimulus-specific? *Frontiers in Psychology*, *3*, 70.
- Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *Neuroimage*, *172*, 766–774.
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, *77*, 980–991.
- Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*, 401–408.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please spell out 2AFC in the section headings to which it appeared.
2. Please provide the expanded form of 2AFC here in its first occurrence in the body/text.
3. Please provide volume and page numbers for this reference (Choi et al., 2020).
4. Please provide publisher location for this reference (Seltman, 2012).

END OF ALL QUERIES

Uncorrected Proof