Published in final edited form as:

Cortex. 2017 May ; 90: 31–45. doi:10.1016/j.cortex.2017.02.004.

Online neural monitoring of statistical learning

Laura J. Batterink[*] and **Ken A. Paller**

Northwestern University, Department of Psychology, 2029 Sheridan Road, Evanston IL 60208

Abstract

The extraction of patterns in the environment plays a critical role in many types of human learning, from motor skills to language acquisition. This process is known as statistical learning. Here we propose that statistical learning has two dissociable components: (1) perceptual binding of individual stimulus units into integrated composites and (2) storing those integrated representations for later use. Statistical learning is typically assessed using post-learning tasks, such that the two components are conflated. Our goal was to characterize the online perceptual component of statistical learning. Participants were exposed to a structured stream of repeating trisyllabic nonsense words and a random syllable stream. Online learning was indexed by an EEG-based measure that quantified neural entrainment at the frequency of the repeating words relative to that of individual syllables. Statistical learning was subsequently assessed using conventional measures in an explicit rating task and a reaction-time task. In the structured stream, neural entrainment to trisyllabic words was higher than in the random stream, increased as a function of exposure to track the progression of learning, and predicted performance on the RT task. These results demonstrate that monitoring this critical component of learning via rhythmic EEG entrainment reveals a gradual acquisition of knowledge whereby novel stimulus sequences are transformed into familiar composites. This online perceptual transformation is a critical component of learning.

Keywords

implicit learning; intertrial coherence; neural entrainment; steady-state response; word segmentation

1. Introduction

Spoken words in an unknown foreign language can seem exceedingly rapid and incomprehensible compared to normal speech in one's native language, despite the fact that syllable rate across languages is relatively similar (Pellegrino, 2011). Natural speech consists

[*]To whom correspondence should be addressed: Laura Batterink, Northwestern University, Department of Psychology, 2029 Sheridan Road, Evanston, IL 60208, lbatterink@northwestern.edu.
Corresponding author information: Laura J. Batterink, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, USA, lbatterink@gmail.com.

of a continuous stream of sound with no reliable pauses between words (Lehiste, 1960), and a major challenge for language learners is to discover word boundaries, a process known as speech segmentation. One reason that beginners learning a new language may perceive unfamiliar speech as unfolding quickly is because they are not yet capable of segmenting the speech stream, leading them to perceive a greater number of basic perceptual units (i.e., multiple individual syllables rather than multisyllabic words). Thus, one of the very first stages of word learning is a perceptual process, requiring a shift in the perception of smaller syllable-units (Bertoncini and Mehler, 1981; Mehler et al., 1981) to that of larger word-units. Only after this perceptual shift has been accomplished can the extracted word forms, comprising key building blocks of language, be stored in memory and undergo further processing (e.g., Graf Estes et al., 2007).

Statistical learning, the process of becoming sensitive to statistical structure in the environment, is thought to be a critical learning mechanism underlying speech segmentation (Saffran, 2003). Learners can discover word boundaries by computing the transitional probabilities between neighboring syllables, which are higher within words compared to across word boundaries (Saffran et al., 1996a, 1996b). In the first demonstration of statistical learning, infants were exposed to a continuous auditory stream of repeating trisyllabic nonsense words, in which transitional probabilities served as the only cue to word boundaries (Saffran, 1996a). A subsequent test revealed that infants' visual fixation times differed between words from the stream and non-word foils made up of recombined syllables, demonstrating that infants were sensitive to the statistical properties of the input. Since this seminal study, a large literature has demonstrated that statistical learning operates across ages and sensory modalities (e.g., Bulf, Johnson, & Valenza, 2011; Saffran et al., 1997; Saffran et al., 1999; Fiser & Aslin, 2001; Turk-Browne, Junge, and Scholl, 2005; Conway & Christiansen, 2005), and that it contributes to a wide range of cognitive functions in addition to speech segmentation (e.g., Fiser & Aslin, 2001; Hunt & Aslin, 2001; Goujon & Fagot, 2013; Saffran et al., 1999; Creel et al., 2004).

At a mechanistic level, statistical learning can be conceptualized as comprising at least two dissociable stages or components. As illustrated by the earlier example of foreign speech perception, the initial component is perceptual in nature, involving a transition from the perception and encoding of raw individual stimulus units (e.g., syllables) to that of larger integrated items (e.g., words). This perceptual process, which we will refer to as the *word identification component*, may be considered the central challenge of statistical learning. After words are encoded as units, the extracted representations can be stored as such in long-term memory. This *memory storage component* can in one sense be considered peripheral to the statistical learning process per se, but nonetheless critically influences performance on subsequent tests of statistical learning. Effective memory storage of integrated representations is also a prerequisite for further processing that occurs after initial segmentation, such as acquiring phonological patterns across words (e.g., Saffran & Thiessen, 2003) and mapping words to objects (Graf Estes et al., 2007; Mirman et al., 2008).

Although it is well-established that there are substantial individual differences in general long-term memory abilities (e.g., Bors & MacLeod, 1996), little is known about the extent to which the word identification component of statistical learning shows individual variability.

These differences may be considerable, given that performance on post-exposure statistical learning tasks varies substantially, with at least one third of a sample often failing to perform the task above chance levels (Siegelman & Frost, 2015; Frost et al., 2015). This variability in statistical learning performance could in principle be attributable to the word identification component of statistical learning, subsequent long-term memory storage, or both. One theoretical possibility is that the word identification component of statistical learning is relatively invariant across individuals with normal sensory processing, as some studies have suggested for implicit learning (e.g., Reber, Walkenfeld & Hernstadt, 1991; Reber, 1993). In this case, individual differences in performance on statistical learning tasks would be driven primarily by variability in long-term memory storage. Alternatively, there may also be substantial individual variability in the online perceptual and encoding processes contributing to word identification. If so, variability in the word identification component should at least partially account for observed individual differences in statistical learning task performance (Siegelman & Frost, 2015; Frost et al., 2015).

These two alternatives have not been previously explored, as previous statistical learning studies generally have not drawn a conceptual distinction between these two components of learning, nor attempted to disentangle them empirically. In large part, this may be due to the experimental approach that has been used to investigate statistical learning. With some notable exceptions (Turk-Browne et al., 2008; Cunillera et al., 2006, 2009; Karuza et al., 2013; McNealy et al., 2006, 2010), the vast majority of studies have followed the same general approach: an initial learning phase involving exposure to structured input, followed by an offline test. In infants, this test generally comprises assessing visual fixation times to test items (e.g., Saffran et al., 1996a, 1999; Aslin et al., 1998), and in adults, responses on a forced-choice recognition measure (Saffran et al., 1996b, 1997; Turk-Browne et al., 2005). These tasks require learners to retrieve previously encoded knowledge after statistical learning is completed, focusing on the final outcome of learning rather than on the word identification component, the central process of online statistical learning. Poor performance on such tasks may be driven by a failure of the word identification component of statistical learning, or of long-term memory storage. Another limitation of such end-state outcome measures is that they are unable to assess the time course of learning.

According to our conceptualization of statistical learning, a shift in the perception and encoding of raw individual stimulus units towards that of larger integrated items is a critical component of statistical learning, and a prerequisite for above-chance performance on subsequent statistical learning tests. The goal of the present study was to characterize this word identification component of statistical learning, including its time course, its variability among individuals, and its relation to performance on post-exposure learning tasks. First, we hypothesized that a shift in perception of integrated items over individual units should increase as a function of exposure to structured input. Second, we hypothesized that individuals would show measurable differences in this word-identification component, with some individuals showing evidence of more effective segmentation than others. Third, we hypothesized that individual variability in the word-identification component would predict performance on post-exposure statistical learning tasks. Such a result would provide evidence that the perceptual shift from syllables to words is a key component of statistical

learning, giving rise to long-term memory storage of segmented representations that can then be measured on post-exposure tasks.

To test these hypotheses, we used an online EEG frequency-based measure of learning that allowed us to track a potential shift in learners' perception of integrated items over individual stimuli in an ongoing sensory stream. Following previous auditory statistical learning studies (e.g., Saffran et al., 1996a, 1996b), participants were presented with a stream of repeating trisyllabic nonsense words, in which transitional probabilities served as the only cue to word boundaries. EEG was recorded throughout the exposure period. We quantified learners' perception of integrated word items using an EEG measure that takes advantage of the neural steady-state response— a property of the electromagnetic activity of the brain to resonate at the same frequency as an ongoing rhythmic stimulus (Buiatti et al., 2009; Picton et al., 2003). When EEG signal power is computed across frequencies, this effect appears as an increase in power at the stimulus frequency and/or its harmonics. The steady-state response corresponds to the presentation frequency of basic perceptual units and thus should be sensitive to statistical learning. Our prediction was that when a learner is successful in perceptually grouping the syllables into words in the speech stream, then the concurrently recorded steady-state response will show a decrease at the frequency of the individual syllables and an increase at the frequency of the trisyllabic words (Figure 1).

A similar approach has previously been used to investigate the effects of acoustic cues such as subliminal pauses on speech segmentation (Buiatti et al., 2009). In this study, concatenated syllables were presented at a regular rate, either in random order or structured as trisyllabic items with nonadjacent dependencies, in which the first syllable of each item predicted the identity of the third syllable (i.e., AXC). The syllable streams were generated in two versions, with and without the addition of a subliminal (25 ms) pause after every third syllable. Successful speech segmentation occurred only in the structured streams with pauses, resulting in an increase in EEG power corresponding to the frequency of the underlying word units and a corresponding decrease in power corresponding to the raw syllable presentation frequency, relative to the other three conditions. The peak in power at the word frequency was correlated with the number of correctly reported words, as assessed at regular intervals throughout the learning phase. Critically, this result suggests that the neural steady-state response is not dominated by low-level sensory processing but is modulated by higher-level perception and integration processes.

Because of the continuous nature of the statistical speech stream, the neural steady-state response is ideally suited as a measure of statistical learning. In contrast, the computation of event-related potentials (ERPs) to individual syllables within rapidly presented speech is complicated by baseline issues as well as a dampening in amplitude of the evoked response to each syllable, affecting the signal-to-noise ratio (cf. Buiatti et al., 2009). Nonetheless, several ERP components have been implicated as potential indices of statistical learning across a number of different studies (Cunillera et al., 2006, 2009; Sanders, Newport & Neville, 2002; De Diego Balaguer et al, 2007). Of these effects, the N400 component appears to be the most robust ERP index of speech segmentation (Cunillera et al., 2006, 2009; Sanders et al., 2002; De Diego Balaguer et al., 2007), though N100 (Sanders et al., 2002) and P200 (De Diego Balaguer et al, 2007) effects have also been reported. In this

context, the N400 may reflect lexical search (Sanders et al., 2002), which occurs only after learners have segmented the continuous stream of syllables into individual words. To align with this prior literature, in the present study we also computed ERPs to word onsets, though our main measure of interest was the neural steady-state response in the frequency domain.

To summarize our design, EEG was recorded while learners were exposed to a speech stream of repeating trisyllabic nonsense words. As a control, participants were also exposed to a speech stream of pseudorandomly concatenated syllables. After exposure to the structured stream, participants performed two post-exposure learning tasks designed to assess statistical learning knowledge. Knowledge of words in the structured stream was assessed directly using a rating task, in which participants provided familiarity ratings to words from the stream and foil items made up of recombined syllables. Statistical learning was also assessed indirectly using a target detection task, an online, speeded, performance-based measure (Batterink et al., 2015a, 2015b, Franco et al., 2015). This task requires participants to respond to target syllables occurring in a continuous syllable stream and assesses the extent to which learners use their acquired statistical knowledge to optimize online processing. We expected to observe faster reaction times (RTs) to relatively predictable targets (i.e., those occurring in later syllable positions compared to the first position), indexing more efficient processing (e.g., Turk-Browne et al. 2005; Kim et al. 2009; Batterink et al. 2015a, 2015b; Franco et al. 2015). Whereas the rating task is presumably most sensitive to explicit memory, the target detection task has the potential to capture knowledge above and beyond what is reflected by explicit memory and may partially reflect contributions from implicit memory (Batterink et al., 2015a). Together, these tasks allowed us to assess both explicit and implicit memory following statistical learning, and to relate these measures to the online perception of word units as assessed through our neural entrainment measure.

Our central prediction was for greater neural entrainment to the underlying word structure in the structured condition compared to the random control condition. We also expected to observe increased evidence of word learning in the structured condition as a function of exposure, with progress following a learning curve. Finally, we hypothesized that individual differences in online neural entrainment would predict performance on post-exposure learning tasks, providing evidence that the online perception and encoding of underlying words in continuous speech is an important component of statistical learning that can be reliably measured at the neural level.

2. Materials and Methods

2.1 Participants

Two groups of participants were run. A primary group of participants (n = 24; 13 women; mean age = 20.8 y, SD = 1.5 y) completed an online statistical learning exposure task followed by several post-exposure learning tests. A secondary group of participants (n = 21; 13 women; mean age = 21.1 y, SD = 3.1 y) completed the same online statistical learning exposure task as the primary group, but did not complete the same post-exposure learning tasks. Because the experimental protocol during the exposure task was identical for both groups of participants, data from both groups (n = 45) were included in all EEG analyses

that did not involve behavioral data, thereby increasing statistical power. Analyses that involved behavioral data were conducted only on the primary group of participants (n = 24). Two additional participants (one from the primary group and one from the secondary group) completed the protocol but their data were subsequently excluded from all analyses due to technical issues with EEG recording.

All participants were fluent English speakers and had no history of neurological problems. Experiments were undertaken with the understanding and written consent of each participant. Participants were compensated $10/h for their time.

2.2 Stimuli

Syllables contributing to the speech streams were individually generated using an artificial speech synthesizer and recorded as separate sound files in Audacity with a sampling rate of 44100 Hz. Two separate sets of syllable inventories were created, corresponding to the two streams and consisting of 12 individual syllables. Each syllable was unique and belonged only to one stream. The two syllable inventory sets were recorded using different synthesizer voices in order to minimize interference between the two streams. For each participant, one inventory of syllables was assigned to the structured condition and the other inventory of syllables was assigned to the random condition. Assignment of the syllable inventories to the structured and random conditions was counterbalanced across participants. The continuous speech streams were created by concatenating the individual syllables together in a predefined order, at a rate of 300 ms per syllable. Syllable sound files were 300 ms or shorter in duration.

Because some syllables could have been easier to perceive than others, in our primary group (n = 24) assignment of individual syllables to the first, second and third positions of each word was counterbalanced across participants, resulting in three different counterbalancing versions. This counterbalancing procedure was carried out in order to minimize any stimulus-driven effects on our reaction time measure in the target detection task. The different counterbalancing versions (1–3) included 10, 7, and 7 participants, respectively. We confirmed that reaction time results were unchanged when 3 participants were randomly excluded from the first counterbalancing group, yielding exactly 7 participants in each counterbalancing version.

2.3 Procedure

A visual summary of the experimental design is shown in Figure 2. Auditory stimuli were presented at a comfortable listening level (approximately 70–75 dB) from two speakers (Dell) placed approximately 120 cm in front of the participant.

2.3.1 Exposure task—Each participant was presented with both a structured and a random syllable stream. In the structured condition, syllables were grouped into 4 repeating trisyllabic words (e.g., *tupiro, golabu, bidaku*, and *padoti*), with the transitional probability between neighboring syllables higher within words (1.0) than between words (0.33). For example, a transitional probability of 1.0 for a word, like *tupiro*, means that every *tu* in the stream was followed by *pi* and every *pi* was followed by *ro*; in contrast, *ro* was equally likely

to be followed by *go*, *bi*, or *pa* (words were not allowed to repeat). In the random condition, syllables were concatenated pseudorandomly, without any higher-order structure; the only constraint was that syllables could not repeat. A total of 2400 syllables (corresponding to 800 "words" in the structured condition) were presented in each stream, at a rate of 300 ms per syllable (i.e., 3.3 Hz). Each stream was broken up into three blocks, each block approximately 4 min in duration. Participants were given a brief break after each block. During the break, they were asked to complete a questionnaire in which they estimated the number of unique words in the stream containing a specified number of syllables (1, 2, 3, 4, 5, or 6 or more).

Condition order (structured before random or random before structured) was counterbalanced across participants. Participants in the primary group completed all subsequent behavioral learning tasks immediately after exposure to the structured stream. All post-exposure learning tasks referred exclusively to the structured stream, and no post-exposure learning tasks were conducted on the random stream.

2.3.2. Rating task—Following exposure to the structured stream, participants completed a rating task designed to assess explicit memory of the nonsense words. On each trial, participants were presented with one of three types of auditory stimuli: a word from the language that had been previously presented 800 times during the Exposure task (e.g., *tupiro*), a part-word that consisted of a syllable pair from a word from the language plus an additional syllable (e.g., *gopiro*), or a non-word that consisted of three syllables from the language that were never paired together within a word (e.g., *godapi*). As in the Exposure task, the stimulus onset asynchrony (SOA) between consecutive syllables within each word, part-word, or non-word was 300 ms. A prompt ("Please give familiarity rating") was presented 770 ms after the onset of the final syllable. Participants were asked to rate on a 1–4 scale how familiar the stimulus sounded based on the language that they had just heard, with 1 indicating "very unfamiliar" and 4 indicating "very familiar." A total of 12 trials were presented, consisting of 4 words, 4 part-words, and 4 non-words. The next trial began approximately 1500 ms after the participant entered his or her response.

2.3.3. Additional tasks of explicit memory—The rating task represented our primary test of explicit memory, as it included the largest number of test items and was performed prior to the other tasks, thus avoiding potential interference caused by repeated testing. However, we also included two additional tests of explicit memory following the rating task.

The first task was a comparison task, designed to assess whether explicit, strategic processing improved discrimination between words from the language and foils. The comparison task consisted of two phases. In the first phase, participants listened to eight auditory stimuli (four words and four non-words) sequentially. On each trial, they were asked to circle a number from 1–10 on a piece of paper indicating the likelihood that the word was present in the language that they had just heard, with 10 being highly likely. In the second phase, participants were informed that four of the words were in fact from the language and that four were not from the language. Again, they were asked to provide a rating from 1–10 indicating the likelihood that each stimulus was found in the language. Before giving their final responses, they were allowed to listen to each stimulus as many

times as they liked, and in any order, by pressing a button corresponding to each of the eight stimuli. This procedure was designed to encourage participants to explicitly compare the auditory stimuli, providing a measure of whether this type of strategy resulted in better or poorer discrimination between words and foils.

Secondly, we also included a forced-choice recognition task, as this task has traditionally been the most common way of assessing statistical learning. Each trial included a word and either a part-word or non-word foil, separated by a 1500-ms ISI. 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. The four words were paired exhaustively with four foils, resulting in a total of 8 test items and 16 trials. The next trial began approximately 150 ms after the participant's second response. As in prior tasks, the SOA between consecutive syllables within words for both the comparison and recognition tasks was 300 ms.

2.3.4 Target detection task—Finally, as in our previous statistical learning studies (Batterink et al., 2015a, 2015b), participants completed a speeded target detection task as a final measure of statistical learning. On each trial, participants were presented with a speech stream containing the four words from the structured language repeated four times each, which was shorter but otherwise similar to the speech stream presented during the Exposure task. For each stream, participants were required to detect a specific target syllable. Both RT and accuracy were emphasized. Each of the 12 syllables of the structured syllable inventory served as the target syllable three times, for a total of 36 streams. The order of the 36 streams was randomized for each participant. Each stream contained 4 target syllables, providing a total of 48 trials in each of the three-syllable conditions (word-initial, word-medial, and word-final). At the beginning of each trial, participants pressed "Enter" to listen to a sample of the target syllable. The stimulus stream was then initiated. Stimulus timing parameters were identical to those in the Exposure task. Based on our previous findings (Batterink et al., 2015a, 2015b), we expect graded reaction time (RT) effects as a function of syllable position. Syllable targets that occur in the final position of a word should elicit faster RTs, indexing facilitation due to statistical learning.

2.4. Behavioral Data Analysis

For each individual, we computed two measures of performance on the Rating task. "Rating accuracy" was computed as the percentage of trials that were rated correctly, defined as a rating of 3 or 4 for words and 1 or 2 for part-words and non-words. In addition, a "rating score" was computed by subtracting the average score given for part-words and non-words from the average score given for words. Perfect sensitivity on this measure would be a score of 3, with values above 0 providing evidence of learning.

For the target-detection task, median RTs to detected targets ("hits") were calculated for each syllable condition (word-initial, word-medial, and word-final) for each participant.

Responses that did not occur within 0 – 1200 msec of a target were considered to be false alarms. It was noted empirically that some syllables were more difficult to perceive and detect than others, resulting in undesirable stimulus-driven influences on reaction times that varied between the three different counterbalancing conditions. To correct for these differences, we computed adjusted reaction time scores for each participant at the individual level. First, across all participants, we computed median reaction times for each of the 24 distinct syllables. We then conducted a repeated-measures ANOVA with syllable (1–24) and syllable position (1–3) as within-participants factors, yielding predicted and residual RTs for each individual syllable based on the effect of syllable position. For each participant, we then subtracted the residual RT value from the observed value for each syllable, co-varying out the effects of physical stimulus factors and yielding a corrected RT effect. Corrected RTs were analyzed using a repeated-measures ANOVA with syllable position (initial, medial, final) as a within-participants factor. Planned contrasts were used to examine whether RTs decreased linearly as a function of syllable position. Finally, an "RT score" was computed for each individual participant by subtracting the corrected median RT to third syllable targets from the corrected median RT to first syllable targets, with larger values indicating greater facilitation.

2.5. EEG Recording and Analysis

During both the Exposure Task and the Target Detection task, EEG 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, Amsterdam, The Netherlands). Additional electrodes were placed on the left and right mastoid, at the outer canthi of both eyes, and below both eyes. Scalp signals were recorded relative to the Common Mode Sense (CMS) active electrode and then re-referenced off-line to the algebraic average of the left and right mastoid.

EEG analyses were carried out using EEGLAB (Delorme and Makeig 2004). Our analysis followed the same general procedure used by Buiatti and colleagues (2009) and Kabdebon and colleagues (2015). EEG data acquired during the Exposure task were band-pass filtered from 0.1 to 30 Hz. Data from each block were timelocked to the onset of each word (or every third syllable, in the random condition) and extracted into epochs of 10.8 s, corresponding to the duration of 12 trisyllabic words or 36 syllables (with no pre-stimulus interval). This procedure yielded epochs overlapping for 5/6 of their length. We employed an automatic artifact rejection procedure designed to remove only data containing large artifacts, based on threshold amplitude values adjusted individually for each participant (average threshold value = 210 μV; range = 200–350 μV). Data containing stereotypical eye movements were retained, as eye artifacts have a broad power spectrum and do not affect narrow-band steady-state responses (Srinivasan and Petrovic, 2006). An average of 90 (SD = 46) trials per participant were rejected in the structured condition and 104 (SD = 53) trials in the random condition, yielding averages of 708 (SD = 47) remaining structured trials and 694 (SD = 53) remaining random trials for analyses. Bad channels were identified and interpolated when necessary (mean number of interpolated channels per participant = 2.9 (SD = 3.4).

We quantified neural entrainment at the syllabic and word frequencies by measuring inter-trial coherence (ITC) within each condition (structured/random). ITC, also known as phase-locking value, is a measure of event-related phase locking. ITC values range from 0, indicating purely non-phase-locked activity, to 1, indicating strictly phase-locked activity. A significant ITC 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. As described by Kabdebon and colleagues (2015), phase-locking is a better suited measure of neural entrainment than power spectrum peak estimation, as it (1) is much more robust to background low frequency fluctuations (Forget, Buiatti, & Dehaene, 2009) and (2) has been shown to reliably track speech comprehension in other paradigms (Ahissar et al., 2001; Kerlin, Shahin, & Miller, 2010; Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2013).

ITC was computed using a continuous Morlet wavelet transformation from 0.2 to 20.2 Hz via the *newtimef* function of EEGLAB. Wavelet transformations were computed in 0.1 Hz steps with 1 cycle at the lowest frequency (0.2 Hz) and increasing by a scaling factor of 0.5, reaching 45 cycles at the highest frequency (20.2 Hz). This approach was selected to optimize the tradeoff between temporal resolution at lower frequencies and frequency resolution at high frequencies (Delorme and Makeig, 2004).

We hypothesized that the word identification component of statistical learning would be indexed as relatively higher ITC at the word frequency and lower ITC at the syllable frequency in the structured condition. That is, if participants become sensitive to the underlying trisyllabic structure of the speech stream, they should show a preferential shift in the entrainment of neural oscillations to underlying words relative to individual syllables. Within each condition, we quantified sensitivity to the trisyllabic structure according to the following simple formula, subsequently referred to as the *Word Learning Index* (WLI):

$$\mathrm{WLI}=\frac{\mathrm{ITC}_{\text{word frequency}}}{\mathrm{ITC}_{\text{syllable frequency}}}$$

Higher WLI values indicate greater neural entrainment towards the triplet frequency relative to the raw syllable frequency, indicative of statistical learning in the structured condition. The Word frequency corresponded to 1.1 Hz, which is the presentation frequency of the trisyllabic words, whereas the Syllable frequency was computed at 3.3 Hz, corresponding to the base presentation frequency of individual syllables. The WLI was computed across 6 centro-frontal midline electrodes where ITC at the word and syllable frequencies showed the strongest values (FC1, C1, FCz, Cz, FC2, and C2).

Our statistical analyses focused on testing two main hypotheses. First, we hypothesized that the WLI in the structured condition should be higher than in the random condition. This result would indicate that participants extracted the structure from the structured stream, perceiving or processing the stimuli as trisyllabic units rather than individual syllables, providing evidence of statistical learning. Second, we hypothesized that the WLI in the structured condition should increase as a function of exposure, but should show no change as in the random condition. This result would provide evidence that participants became increasingly sensitive to the trisyllabic word structure as exposure increased. To test these

hypotheses, we divided data from the exposure period into three equal blocks in each condition and computed the WLI within each block. A repeated-measures ANOVA was then conducted with condition (structured, random) and block (1–3) as within-participants factors. Planned contrasts were used to examine whether the WLI increased linearly as a function of block. Data from both the primary and secondary group of participants were combined for these analyses to increase power.

We also calculated correlations between the WLI in the structured condition and the WLI in the random condition, computed across blocks. It was noted that the distributions of WLI values in both conditions were significantly positively skewed (Structured WLI: skewness = 2.69, SE = 0.47, $W(24) = 0.610$; $p < 0.001$; Random WLI: skewness = 3.33, SE = 0.47, $W(24) = 0.53$, $p < 0.001$). Therefore, for all correlational analyses, we used log-transformed WLI values in order to increase sensitivity of these tests. We examined whether structured–random WLI correlations differed as a function of task order (structured first versus random first) by using the web utility provided by Lee and Preacher (2013), which converts correlation coefficients into *z*-scores using Fisher's *r*- to *z*-transformation.

To relate our results to those in previous ERP studies, we also computed ERPs to word onsets. After applying the same filter settings and artifact-rejection parameters as in the ITC analysis, data were timelocked to the onset of each word in the structured condition or every third syllable in the random condition, and extracted into epochs of 1200 ms, including a 300 ms baseline. Epochs were baseline corrected from −300 to 0 ms relative to stimulus onset, corresponding to the duration of the prior syllable. Statistical analyses focused on the N400, as only this component showed differences between the structured and random conditions, based on visual inspection of the waveform. The N400 was analyzed statistically by averaging amplitudes from 300 to 500 ms post-stimulus across neighboring electrodes to form nine channel groups of interest (left anterior region: AF7, AF3, F7, F5, F3; left central region: FT7, FC5, FC3, T7, C5, C3; left posterior region: TP7, CP5, CP3, P7, P5, P3, PO7, PO3; midline anterior region: AFZ, F1, FZ, F2; midline central region: FC1, FCZ, FC2, C1, CZ, C2; midline posterior region: CP1, CPZ, CP2, P1, PZ, P2, POZ; right anterior region: AF4, AF8, F4, F6, F8; Right central region: FC4, FC6, FT8, C4, C6, T8; right posterior region: CP4, CP6, TP8, P4, P6, P8, PO4, PO8). These mean amplitudes were analyzed using a repeated-measures ANOVA with condition (structured, random), left/mid/right (left, middle, right), and anterior/posterior (anterior, central, posterior) as within-subjects factors. Greenhouse–Geisser corrections were applied for factors with more than two levels. We also examined whether the WLI and ITC at the word frequency in the structured condition were related to the N400 effect by calculating correlations between these measures. In correlational analyses, the N400 effect was computed by subtracting N400 amplitude in the random condition from N400 amplitude in the structured condition, across the same 6 centro-frontal midline electrodes as used for the WLI analyses.

2.6. Correlations Between WLI and Performance on Post-Exposure Tasks

Finally, we examined whether the WLI predicted performance on post-exposure statistical learning tasks, namely the Rating task and the Target Detection task. Pearson's correlations were computed between individual participants' log-transformed WLI values in the

structured and random conditions and Rating scores and Rating accuracy (from the Rating task) and RT scores (from the Target Detection task).

3. Results

3.1. Behavioral Results

3.1.1 Exposure questionnaire—Across the three blocks, participants greatly overestimated the number of unique words in both the structured and random streams (overall number of words estimated in structured condition = 26.3, *SD* = 37.0; random condition = 37.1, *SD* = 38.5). Although far from accurate, this measure of perception showed significant differences between conditions. Participants estimated that there was a greater overall number of unique words in the random block compared to the structured block (Condition effect: $F(1,23) = 5.60$, $p = 0.027$). This measure also showed significant differences in the time course over exposure between conditions (Condition x Block: $F(2,46) = 4.52$, $p = 0.034$). In the structured condition, the overall number of estimated words did not significantly change as a function of block (Block: $F(2,46) = 0.20$, $p = 0.72$). In contrast, in the random condition, the number of estimated words significantly increased as a function of exposure (Block: $F(2,46) = 6.40$, $p = 0.014$; linear contrast: $F(1,23) = 6.86$, $p = 0.015$).

3.1.2. Rating task—Participants demonstrated significant evidence of statistical learning on the rating task (Word Category effect: $F(2,46) = 24.5$, $p < 0.001$; linear effect of word category: $F(1,23) = 40.3$, $p < 0.001$). Words were rated as most familiar, followed by part-words, with non-words rated as least familiar (Figure 3A). Rating accuracy was 62.1% (*SD* = 14.3%), significantly above chance ($t(23) = 21.3$, $p < 0.001$). Mean rating score across participants was 0.78 (*SD* = 0.61), significantly above chance ($t(23) = 6.28$, $p < 0.001$).

3.1.3. Target detection task—Participants performed well on the target detection task, responding to 89.1% (*SD* = 9.2%) of targets within 1200 ms and making an average of 12.3 false alarms (*SD* = 10.5). As hypothesized, RTs were significantly modulated as a function of syllable position (Uncorrected RTs: Syllable Position effect: $F(2,46) = 14.3$, $p < 0.001$; linear effect of Syllable Position: $F(1,23) = 24.5$, $p < 0.001$; Corrected RTs: Syllable Position effect: $F(2,46) = 30.6$, $p < 0.001$; linear effect of Syllable Position: $F(1,23) = 42.1$, $p < 0.001$), with slowest responses to initial-position targets, intermediate responses to second-position targets, and fastest responses to final-position targets (Figure 3B). These results indicate the processing was progressively facilitated for more predictable syllables, providing evidence of statistical learning. Mean RT score across participant (computed as the RT difference between first syllable targets and third syllable targets) was 79.1 ms (*SD* = 59.7), significantly above chance ($t(23) = 6.49$, $p < 0.001$).

3.1.4. Relation between performance on rating task and target detection task —Performance on the rating task and target detection task significantly correlated across participants (rating score: $r = 0.51$, $p = 0.010$; rating accuracy: $r = 0.42$, $p = 0.044$), indicating that learners who performed better on the rating task also showed larger facilitation effects on the target detection task. Given that the rating task provides a measure of explicit memory, this correlation suggests that the target detection task is at least

somewhat sensitive to explicit memory as well. Nonetheless, a subgroup of participants (n = 7) who did not achieve above 50% accuracy on the rating task (mean = 45%, SD = 4.4%) still showed a significant reaction time effect (Corrected RTs: Syllable Position effect: $F(2,12) = 6.11$, $p = 0.030$; linear effect of Syllable Position: $F(1,6) = 28.7$, $p = 0.002$). In keeping with our prior studies of target detection after SL (Batterink et al., 2015), this result indicates that performance on the target detection task cannot be entirely accounted for by explicit memory, and also reflects contributions from implicit memory.

3.1.5. Additional tests of explicit memory—On the comparison task, across both the first and second round, participants rated the 4 words from the language as being more likely to have been presented in the language (M = 7.04, SD = 1.33) compared to the 4 non-words (M = 4.84, SD = 1.35), providing additional evidence of statistical learning (Word Category effect: $F(1,23) = 21.3$, $p < 0.001$). Overall ratings of familiarity declined from round 1 (mean familiarity score across items = 6.2) to round 2 (mean score across items = 5.7; Round effect: $F(1,23) = 7.72$, $p = 0.011$), suggesting that the additional round of testing may have caused some interference and made all items seem somewhat less familiar.

On the recognition task, mean accuracy was 61.7% (SD = 19.1%), significantly above chance ($t(23) = 15.8$, $p < 0.001$). Qualitatively, “remember” responses were the most accurate (M = 64.9%, SD = 34.3%), followed by “familiar” responses (M = 55.1%, SD = 26.6%), with “guess” judgments showing the lowest degree of accuracy (M = 53.7%, SD = 33.7%). However, effect of memory judgment on accuracy was not significant, likely due to the variable and low number of trials in each subdivided condition ($F(2,42) = 0.84$, p = 0.43). Mean proportion of trials in each condition was 28.4% (range = 0% – 100%) for “remember,” 50.0% for “familiar” (range = 0% – 94%), and 21.6% for “guess” (range = 0% – 50%).

Performance on the comparison task in both rounds correlated strongly across participants with rating score (round 1: $r = 0.50$, $p = 0.013$; round 2: $r = 0.45$, $p = 0.027$). Therefore, to minimize the number of comparisons, we included only our main measure of explicit memory, the rating task, in subsequent correlational analyses with the WLI.

3.2. EEG Results

ITC as a function of condition (structured, random), block (1–3), and frequency, computed across the combined participant groups (n = 45), is plotted in Figure 4A. Consistent with our predictions, the structured condition showed an increase in ITC at the word frequency and a decrease in ITC at the syllable frequency, relative to the random condition. This relative shift in ITC towards the word frequency and away from the syllable frequency appeared to increase as exposure progressed. The distribution of ITC across the scalp is shown in Figure 4B.

These observations were quantified and statistically tested using the WLI, which is plotted as a function of condition and learning block in Figure 4C. As we predicted, the structured condition showed a significantly higher WLI than the random condition across the three blocks (Condition effect: $F(1,44) = 17.3$, $p < 0.001$). This WLI in the structured condition compared to the random condition differed significantly as a function of block (Condition x

Block: $F(2,88) = 3.72$, $p = 0.029$; linear contrast: $F(1,44) = 6.62$, $p = 0.014$). In the structured condition, the WLI increased linearly as exposure progressed (Block: $F(2,88) = 3.11$, $p = 0.056$; linear contrast: $F(1,44) = 4.89$, $p = 0.032$). This increase in the WLI corresponded to a significant interaction in ITC between frequency and block (Frequency x Block: $F(2,88) = 4.14$, $p = 0.021$), reflecting an increase in ITC at the word frequency and a decrease in ITC at the syllable frequency as a function of block. In contrast, there was no significant change over blocks in the random-condition WLI (Block: $F(2,88) = 0.049$, $p = 0.92$). In sum, as hypothesized, EEG oscillatory phase-locking to triplet or word units was enhanced in the structured condition relative to the random condition, and increased as a function of exposure only in the structured condition. These results provide evidence of online statistical learning of the underlying word units in the structured condition.

The structured WLI and random WLI were strongly and significantly correlated ($r = 0.63$, $p = 0.001$; WLI values log-transformed). Task order (structured versus random) had a marginal effect on these correlations. Although both groups of participants showed significant correlations between the structured WLI and random WLI ("structured first": $r = 0.74$, $p < 0.001$; "random first": $r = 0.46$, $p = 0.024$), correlations were marginally stronger in participants who were exposed to the structured stream first ($z = 1.43$, $p = 0.076$, one-tailed). Task order did not have a significant effect on either the overall WLI condition effect or the condition by block interaction (all p values > 0.1).

3.3. EEG–Behavioral Correlations

We tested whether the WLI in the structured condition, computed across the entire exposure period, predicted subsequent learning effects on our two main post-learning behavioral measures, the rating task and the target detection task. Critically, across all participants ($n = 24$), the WLI in the structured condition significantly predicted RT scores at the individual level (Table 1; Figure 5). Correlations between the WLI (in both conditions) and performance on the rating task (both rating accuracy and rating score) were also positive but did not reach statistical significance. These results suggest that the tendency to segment the speech streams into triplet chunks, as assessed online through the WLI, predicts subsequent facilitation on the post-exposure target detection task. Unexpectedly, the WLI in the random condition also significantly predicted task performance on the target detection task (Table 1).

3.4. ERP Results

Word onsets in the structured condition elicited a significantly larger N400 effect compared to triplet onsets in the random condition (Condition: ($F(1,44) = 5.96$, $p = 0.019$; Figure 6). This N400 effect was maximal over midline sites (Condition x left/mid/right: $F(2,88) = 11.88$, $p < 0.001$). Visual inspection of the ERP waveforms did not reveal any other ERP components showing differences between the structured and random conditions.

Across subjects, the N400 effect (i.e., the difference in N400 amplitude between words and random triplets) did not significantly correlate with the WLI in the structured condition ($p = 0.15$), or with ITC at the word frequency ($p = 0.30$). Thus these measures may reflect different aspects of statistical learning and word segmentation.

4. Discussion

Our results provide support for the idea that the identification of word-like items in continuous speech is a critical component of statistical learning, and is conceptually dissociable from effective memory storage of the extracted representations for later use. The EEG frequency tagging approach (cf. Buiatti et al., 2009; Kabdebon et al., 2015), which was used here to generate a Word-Learning Index or WLI, is a powerful tool to study the progressive shift of word units over syllable units as the basic perceptual unit in continuous speech. In the present study, using a standard auditory statistical learning task, we showed that the WLI was enhanced in the structured condition relative to the random control condition. Using the WLI as an index of the word identification component of statistical learning, we also confirmed all three of our major hypotheses: (1) the WLI increased as a function of block in the structured condition only, (2) the WLI showed observable individual variability, and (3) variability in the WLI systematically predicted performance on post-exposure learning tasks.

First, the finding that the WLI increased as a function of block in the structured condition, reflecting a relative increase in neural entrainment to the trisyllabic structure, demonstrates that learners showed an overall shift in their perception from individual stimuli units to more integrated items as exposure increased. This result indicates that perception of underlying word units is built upon and shaped by previous knowledge. It follows a learning curve rather than occurring instantaneously, a hallmark of learning. Second, the perception and encoding of word units shows quantifiable differences at the individual level. Learners demonstrated substantial differences in terms of their overall perception of the underlying word units, with some individuals showing relatively higher entrainment at the word frequency and others showing relatively higher entrainment at the syllable frequency. Differences were also found with respect to individual learning curves. Finally, and perhaps most critically, the online perception of integrated units predicted individual performance on offline measures of statistical learning, most notably the target detection task. Learners who showed a greater tendency to perceive trisyllabic items, as measured through our online neural measure, also showed a larger reaction time effect, reflecting greater facilitation in processing from knowledge acquired through statistical learning. This result indicates that individual differences in online perception are measurable and relate systematically to other measures of statistical learning. It also indicates that long-term memory storage processes depend upon and perhaps interact with processes involved in perceptual segmentation.

We had originally hypothesized that the WLI in the structured condition, and not the random condition, would predict performance on post-exposure learning measures. However, somewhat surprisingly, we found that the WLI in *both* the structured and random conditions predicted performance on post-exposure measures, namely rating accuracy and RT score. In other words, the correlation between the WLI and post-exposure task performance was not specific to the structured condition. In addition, the WLI in the structured and random conditions were highly correlated across learners ($r = 0.63$). This result indicates that learners who showed high neural entrainment to the trisyllabic structure in the structured stream also tended to show high entrainment to every third syllable in the random stream, despite the absence of an underlying trisyllabic structure in the latter condition. Together,

these findings suggest that the WLI in both conditions may be tapping into the general tendency of an individual to seek out underlying patterns in the environment, particularly at the triplet level. Some individuals may tend to naturally process incoming stimuli in groups of three and thus may also process even purely random sequences as potential triplets. On average, these individuals would show better statistical learning in the structured condition, showing superior performance on post-exposure measures. In contrast, other individuals may tend to naturally process input in bundles of 2 or 4 stimuli at a time, rather than in triplets. Still others may tend to show more bottom-up processing, being less likely to impose or organize input according to any overarching structure at all. Both of these latter groups on average would show lower WLI values as well as poorer statistical learning performance in the structured condition.

The observed correlation between the WLI in the structured and random conditions also appears to be partially influenced by which speech stream was delivered first. Learners who were exposed to the structured stream first showed marginally significantly higher correlations between structured and random WLI values than learners who were exposed to the random stream first. This result suggests that initial exposure to the structured speech stream may have induced participants to process the subsequent random stream in a similar way to the structured stream, using trisyllabic grouping.

4.1. Sources of Individual Differences in Statistical Learning Performance

Our data may be understood in the context of recent theoretical work on individual differences in statistical learning conducted by Frost and colleagues (2015). This model proposes that there are two major sources that influence variance in statistical learning performance: (1) variance in encoding representations of individual elements in a stream, within the presentation modality, and (2) variance in detecting the distributional properties of the encoded representations (e.g., the transitional probabilities between syllables). Binding of temporal or spatial contingencies may occur in both modality-specific brain areas (such as higher-level visual areas for visual stimuli, and higher-level auditory areas for auditory stimuli) as well as domain-general areas that are involved regardless of the stimulus modality (such as the medial temporal lobe system). A recent study provided support for this model, showing that these two factors can be dissociated and that they interact to jointly determine statistical learning performance (Bogaerts et al., 2016). This study also demonstrated that these two mechanisms are not independent or additive, but interact with one another. For example, sensitivity to the distributional properties of input can facilitate encoding of individual elements, and conversely, better encoding of elements can enhance the extraction of underlying statistics (Bogaerts et al., 2016). In the context of the present study, both of these mechanisms—encoding and binding—should presumably lead to changes at the perceptual level and influence the WLI in the structured condition. For example, learners in our study who encode the clearest representations of individual syllables and/or who are superior at computing the transitional probabilities between syllables will ultimately have the greatest success at uncovering the hidden trisyllabic structure, showing greater neural entrainment to the word frequency

Another potential mechanism contributing to statistical learning performance, which is related to, yet distinct from, the encoding of individual elements within a sensory modality, involves contributions from “echoic” (Neisser, 1967) or auditory sensory memory. Analogous to the rapidly decaying sensory memory mechanisms that underlie the mismatch negativity (MMN; see Naatanen et al., 2007 for a review), syllables that were recently heard should be represented as short-lived memory traces in auditory sensory memory. If the same syllable is presented again while its sensory memory representation is still active, it will be processed differently from syllables that have not recently been encountered, just as deviant tones are processed differently from congruent tones in MMN paradigms. For example, if two words in the structured stream were presented in close succession, separated by only one other word (e.g., tupiro-golabu-tupiro), the individual syllables occurring in the second word (e.g., tu-pi-ro) would all be processed as “recent repeats.” This differential processing of neighboring syllables within a triplet may potentially lead learners to perceive these syllables as an underlying integrated unit, rather than as individual syllables. Although not commonly discussed in the context of statistical learning, such a mechanism may allow learners to discover patterns in sensory input, independent from the computation of transitional probabilities. In the present study, learners who can maintain a greater number of syllables in auditory sensory memory and/or clearer representations in sensory memory may be more likely to “recognize” if the same triplet has occurred twice within a given time interval, ultimately leading to word identification and faster statistical learning.

Finally, our data from the random condition additionally implicate a fourth source of variance as contributing to the perceptual component of statistical learning, which is the extent to which individuals spontaneously or naturally process incoming stimuli in groups or bundles, even in the absence of structure. Together, all these mechanisms should lead to corresponding changes in the online perception of input, as captured by our neural-frequency-tagging method.

In addition to these mechanisms that contribute to the online word identification component, we propose that the storage of segmented representations should also logically influence performance on statistical learning tasks. Participants may successfully encode the individual elements in a stream and compute their distributional properties, allowing them to perceive integrated units rather than the raw stimuli. However, if they are unable to effectively store the extracted representations into long-term memory, they will subsequently perform poorly on offline statistical learning tasks. Rather than influencing statistical learning performance independently, storage of the extracted representations is likely to interact with and influence word-identification mechanisms. For example, having access to strong representation of a single unitized word stored in long-term memory may facilitate further segmentation, by providing a means of segmenting adjacent words in the continuous speech stream. Consistent with this idea, one study found that infants were able to discriminate high probability and low probability items only when a subset of words in the speech stream (nontargets not included in the test) were also presented in isolation, suggesting that knowledge of discrete items may enhance statistical learning (Lew-Williams et al., 2011). Thus, we propose that the storage of extracted representations also influences perception and correspondingly the frequency-tagging index, but through a less direct path than online perceptual mechanisms.

Finally, our results also provide support for the idea that statistical learning shows reliable individual differences (Siegelman & Frost, 2015). The strong correlation between the WLI in the structured and random conditions is consistent with previous evidence showing that statistical learning is a stable individual capacity, as measured by test-retest reliability on a number of different statistical learning tasks (Siegelman & Frost, 2015). Future studies may address whether the WLI within individuals varies across different statistical learning tasks. This question is interesting in light of recent research suggesting that an individual's statistical learning ability is highly task specific. Siegelman and Frost (2015) demonstrated a lack of correlation across a wide range of statistical learning tasks, suggesting that statistical learning is characterized by both modality- and stimulus-specificity. In addition, performance on statistical learning tasks is largely independent of general cognitive abilities such as intelligence, working memory, and executive function. These findings suggest that individuals in our study may exhibit a very different WLI pattern if assessed on a different statistical learning task (e.g., visual stimuli instead of auditory, or tones instead of syllables).

4.1.2. Relation of the WLI to Subsequent Learning Measures—The WLI showed the strongest correlation with the RT score, derived from the target detection task. The WLI did not significantly predict performance on the rating task when all participants were considered. This finding suggests that the target detection task, as an indirect, processing-based test of memory, may be a more sensitive measure of statistical learning than direct tests of memory used in statistical learning studies, such as the recognition task or the rating task used in the present study. This conclusion converges with one of our previous studies of statistical learning, in which we also found that the target detection task was more sensitive than the recognition task (Batterink et al., 2015a). In particular, the rating and recognition tasks primarily reflect contributions from explicit memory, whereas the target detection task may reflect contributions from both implicit and explicit memory (Batterink et al., 2015a). The WLI, as an index of the word identification component of statistical learning, is dissociable from subsequent explicit memory storage. Better online word identification, as reflected by a higher WLI, gives rise to representations that can potentially contribute to both implicit and explicit memory. Thus, even if explicit memory storage is ineffective (as in a subset of participants in the present study who showed chance-level performance on the rating task), the WLI would correlate with performance on the target detection task, reflecting implicit memory.

4.1.3. Frequency Trade-Off in Neural Entrainment—Relative to the random condition, the structured condition showed both an increase in neural phase-locking at the word frequency and a decrease in phase-locking at the syllable frequency (Figure 4A & B). Similarly, Buiatti and colleagues (2009) observed that the presence of the trisyllabic structure in both the pause-present and pause-absent conditions induced the suppression of entrainment at the syllable frequency, relative to conditions in which no structure was present. Thus, our results support Buiatti and colleagues' suggestion that "word learning induces both an enhancement in power at the frequency of the discovered word, and an inhibition of power at frequencies associated to alternative words with different lengths" (*p.* 516–517). At the neural level, this trade-off effect may represent a perceptual sharpening

mechanism, simultaneously enhancing the signal at the relevant item frequency while reducing entrainment at nontarget or irrelevant frequencies.

Perceptually, this frequency trade-off in neural entrainment may underlie the subjective experience of incomprehensible continuous speech in an unknown language. During early stages of language acquisition prior to mastery of speech segmentation, language learners may perceive a continuous speech stream as a sequence of syllables confusingly blending into each other. However, once segmentation has occurred, the same physical stimulus will be perceived as a sequence of words rather than individual syllables. In fact, the learner will no longer be capable of experiencing the same continuous speech stream in the now-familiar language as he or she once did, as individual syllables. This shift in perception may be driven by the suppression in neural entrainment at the syllable frequency that accompanies word learning. Binocular rivalry and bistable representations represent similar phenomena, in which an individual's subjective perceptual experience fluctuates in the absence of any change in the physical stimulus. Consistent with the present data, one binocular rivalry study that presented two rivalrous stimuli at different flicker rates found an increase in power at the flicker frequency of the consciously perceived stimulus relative to the unperceived stimulus (Tononi et al., 1998).

4.1.4. Conclusion—The neural WLI tracks the perceptual binding of individual discrete stimuli into integrated items. Our results indicate that this perceptual process is a critical component of statistical learning and predicts subsequent performance on post-exposure learning tasks. In addition to providing insight into the underlying mechanisms involved in statistical learning, use of this measure also has the potential to address a number of unresolved questions in this area. For example, future work may focus on using this measure to investigate the optimal learning conditions under which statistical learning occurs, or to directly compare learning in different populations where behavioral responses may not be easily acquired or readily comparable (e.g., infants and patients).

Acknowledgments

This work was supported by grants NIH grants T32 NS047987 and F32 HD 078223.

References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences USA. 2001; 98:13367–13372.

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month- old infants. Psychological Science. 1998; 9:321–324.

Batterink LJ, Reber PJ, Neville HJ, Paller KA. Implicit and explicit contributions to statistical learning. Journal of Memory and Language. 2015a; 83:62–78. [PubMed: 26034344]

Batterink LJ, Reber PJ, Paller KA. Functional differences between statistical learning with and without explicit training. Learning and Memory. 2015b; 22:544–556. [PubMed: 26472644]

Bertoncini J, Mehler J. Syllables as units in infant speech perception. Infant Behavior and Development. 1981; 4:247–260.

Bogaerts L, Siegelman N, Frost R. Splitting the variance of statistical learning performance: A parametric investigation of exposure duration and transitional probabilities. Psychonomic Bulletin & Review. 2016 (Epub ahead of print).

Bors, DA., MacLeod, CM. Individual differences in memory. In: Bjork, EL., Bjork, RA., editors. Handbook of perception and cognition Vol. 10: Memory. San Diego CA: Academic Press; 1996. p. 411-441.

Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. Cognition. 2011; 121:127–132. [PubMed: 21745660]

Buiatti M, Pena M, Dehaene-Lambertz G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. NeuroImage. 2009; 44:509–519. [PubMed: 18929668]

Conway CM, Christiansen MH. Modality-constrained statistical learning of tactile visual and auditory sequences. Journal of Experimental Psychology: Learning Memory and Cognition. 2005; 31:24–39.

Creel SC, Newport EL, Aslin RN. Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. Journal of Experimental Psychology: Learning Memory and Cognition. 2004; 30:1119–1130.

Cunillera T, Camara E, Toro JM, Marco-Pallares J, Sebastian-Galles N, Ortiz H, Pujol J, Rodriguez-Fornells A. Time course and functional neuroanatomy of speech segmentation in adults. NeuroImage. 2009; 48:541–553. [PubMed: 19580874]

Cunillera T, Toro JM, Sebastian-Galles N, Rodriguez-Fornells A. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. Brain Research. 2006; 1123:168–178. [PubMed: 17064672]

De Diego Balaguer R, Toro JM, Rodriguez-Fornells A, Bachoud-Levi A. Different neurophysiological mechanisms underlying word and rule extraction from speech. PLoS One. 2007:e1175. [PubMed: 18000546]

Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004; 134:9–21. [PubMed: 15102499]

Forget J, Buiatti M, Dehaene S. Temporal integration in visual word recognition. Journal of Cognitive Neuroscience. 2009; 22:1054–1068.

Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. Trends in Cognitive Sciences. 2015; 19:117–125. [PubMed: 25631249]

Franco A, Eberlen J, Destrebecqz A, Cleeremans A, Bertels J. Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. Experimental Psychology. 2015; 62:346–351. [PubMed: 26592534]

Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychological Science. 2001; 12:499–504. [PubMed: 11760138]

Graf-Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 2007; 18:254–260. [PubMed: 17444923]

Goujon A, Fagot J. Learning of Spatial Statistics in Nonhuman Primates: Contextual Cueing in Baboons (Papio papio). Behavioural Brain Research. 2013; 247:101–109. [PubMed: 23499707]

Hunt RH, Aslin RN. Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. Journal of Experimental Psychology: General. 2001; 4:658–680.

Kabdebon C, Pena M, Buiatti M, Lambertz-Dehaene G. Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. Brain & Language. 2015; 148:25–36. [PubMed: 25865749]

Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. The neural correlates of statistical learning in a word segmentation task: An fMRI study. Brain and Language. 2013; 127:46–54. [PubMed: 23312790]

Kerlin JR, Shahin AJ, Miller LM. Attentional gain control of ongoing cortical speech representations in a “cocktail party”. Journal of Neuroscience. 2010; 30:620–628. [PubMed: 20071526]

Kim R, Seitz A, Feenstra H, Shams L. Testing assumptions of statistical learning: Is it long-term and implicit? Neuroscience Letters. 2009; 461:145–149. [PubMed: 19539701]

Lee, IA., Preacher, KJ. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. 2013. http://quantpsy.org

Lehiste I. An acoustic phonetic study of open juncture. Phonetica Supplementum ad. 1960; 5:1–54.

Lew-Williams CB, Pelucchi B, Saffran JR. Isolated words enhance statistical language learning in infancy. Developmental Science. 2011; 14:1323–1329. [PubMed: 22010892]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54:1001–1010. [PubMed: 17582338]

McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. The Journal of Neuroscience. 2006; 26:7629–7639. [PubMed: 16855090]

McNealy K, Mazziotta JC, Dapretto M. The neural basis of speech parsing in children and adults. Developmental Science. 2010; 13:385–406. [PubMed: 20136936]

Mehler J, Dommergues JY, Frauenfelder U, Segui J. The syllable's role in speech segmentation. Journal of Verbal Learning and Verbal Behavior. 1981; 20:298–305.

Mirman D, Magnuson JS, Estes KG, Dixon JA. The link between statistical segmentation and word learning in adults. Cognition. 2008; 108:271–280. [PubMed: 18355803]

Naatanen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. Clinical Neurophysiology. 2007; 118:2544–2590. [PubMed: 17931964]

Neisser, U. Cognitive psychology. Englewood Cliffs: Prentice-Hall; 1967.

Pellegrino F, Coupe C, Marsico. A cross-language perspective on speech information rate. Language. 2011; 87:539–558.

Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral Cortex. 2013; 23:1378–1387. [PubMed: 22610394]

Picton TW, Sasha JM, Dimitrijevic A, Purcell D. Human auditory SSRs. International Journal of Audiology. 2003; 42:177–219. [PubMed: 12790346]

Reber AS, Walkenfeld FF, Henstad R. Implicit and explicit learning: Individual differences and IQ. Journal of Experimental Psychology: Learning Memory and Cognition. 1991; 17:888–896.

Reber, AS. Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford: Oxford University Press; 1993.

Saffran JR. Statistical language learning: Mechanisms and constraints. Current Directions of Psychological Science. 2003; 12:110–114.

Saffran JR, Aslin RN, Newport EL. Statistical Learning by 8-Month-Old Infants. Science. 1996a; 274:1926–1928. [PubMed: 8943209]

Saffran JR, Newport EL, Aslin R. Word segmentation: The role of distributional cues. Journal of Memory and Language. 1996b; 35:606–621.

Saffran JE, Newport R, Aslin R, Tunick RA, Barrueco S. Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science. 1997; 8:101–105.

Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999; 70:27–52. [PubMed: 10193055]

Saffran JR, Thiessen ED. Pattern induction by infant language learners. Developmental Psychology. 2003; 39:484. [PubMed: 12760517]

Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: An event- related potential index of perceived onsets in continuous speech. Nature Neuroscience. 2002; 5:700–703. [PubMed: 12068301]

Siegelman N, Frost R. Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. Journal of Memory and Language. 2015; 81:105–120. [PubMed: 25821343]

Srinivasan R, Petrovic S. MEG phase follows conscious perception during binocular rivalry induced by visual stream segregation. Cerebral Cortex. 2006; 16:597–608. [PubMed: 16107587]

Tononi G, Srinivasan R, Russell DP, Edelman GM. Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. Proceedings of the National Academy of Sciences USA. 1998; 95:3198–3203.

Turk-Browne NB, Junge JA, Scholl BJ. The Automaticity of Visual Statistical Learning. Journal of Experimental Psychology: General. 2005; 134:552–564. [PubMed: 16316291]

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. Journal of Cognitive Neuroscience. 2008; 21:1934–1945.

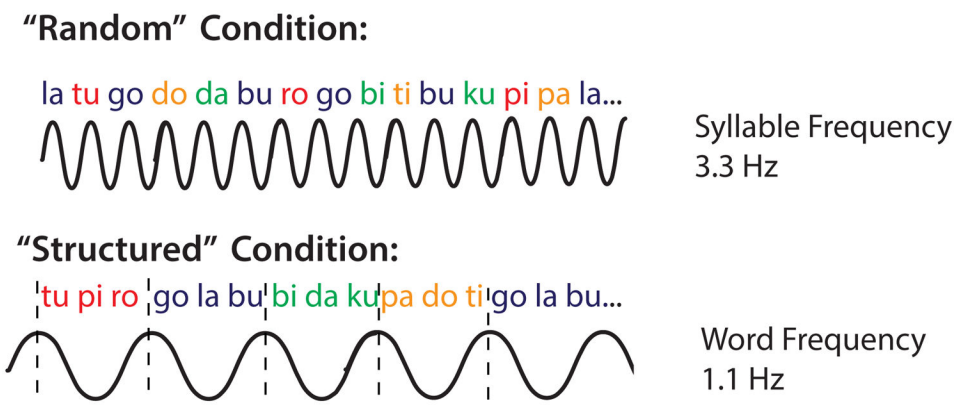

Figure 1.

EEG-based entrainment measure of learning. If perceptual grouping of individual syllables into trisyllabic words occurs during statistical learning, the steady-state response should show a decrease at the frequency of the individual syllables and an increase at the frequency of the trisyllabic words.

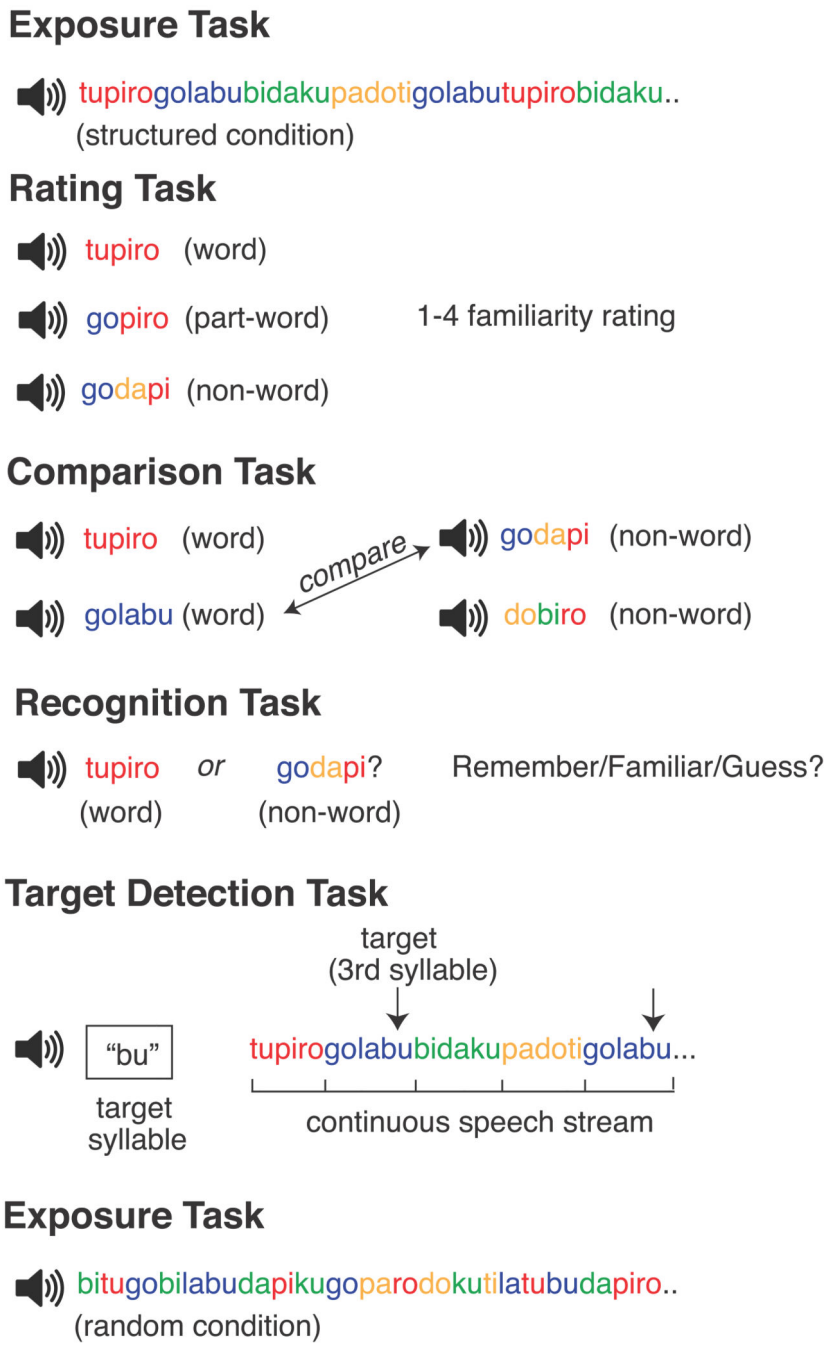

Figure 2.

Summary of experimental design. The exposure task in the structured condition consisted of 12 min of continuous auditory exposure to four repeating nonsense words. The exposure task in the random condition consisted of exposure to pseudorandomly repeating syllables. The main test of explicit memory was the rating task, which required participants to provide words and foil items with a familiarity rating. This task was followed by two additional tests of explicit memory, the comparison task and the recognition task. Finally, the target-detection task was a reaction-time-based measure of statistical learning, in which

participants detected target syllables embedded in a continuous auditory speech stream composed of the four nonsense words. The syllable assigned as the target was rotated across trials.

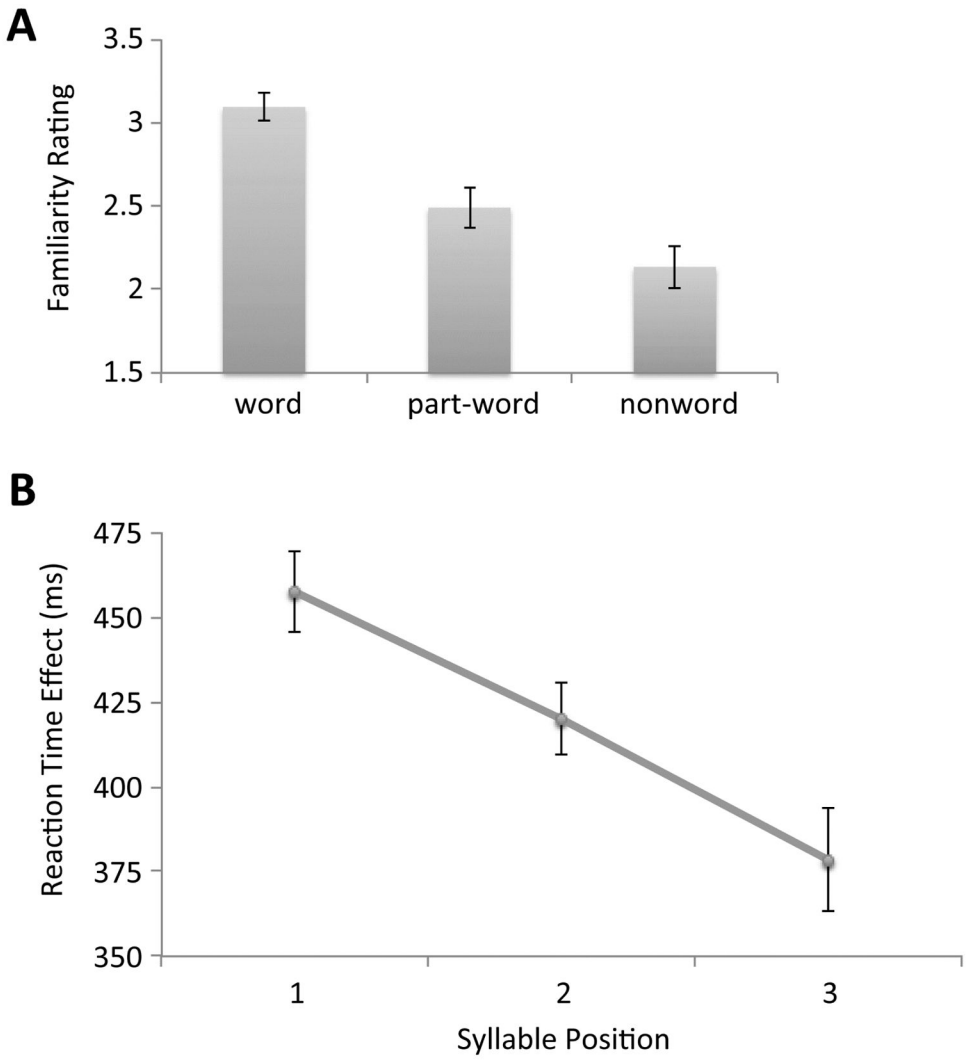

Figure 3.

Behavioral results reflecting statistical learning. (A) Familiarity ratings provided on the rating task. (B) Corrected reaction times as a function of syllable position on the target detection task.

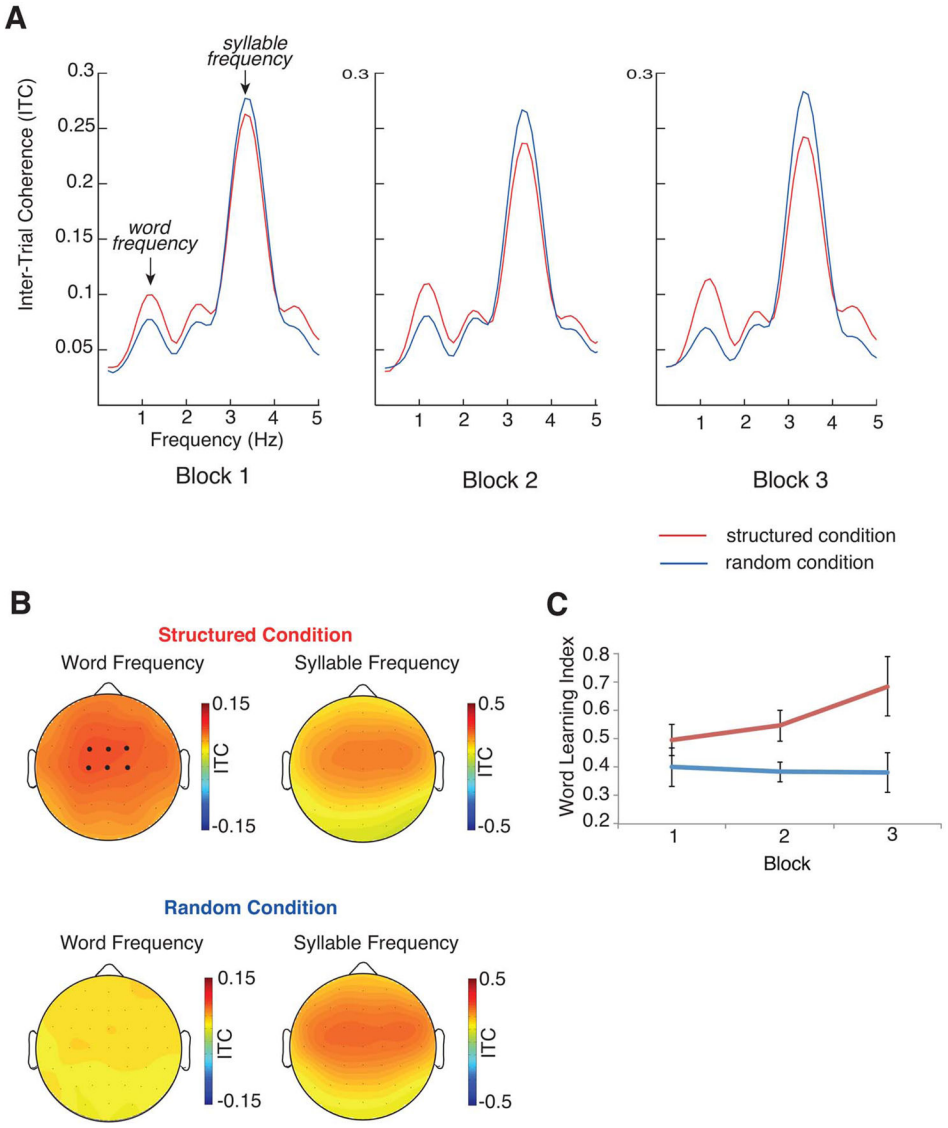

Figure 4.

EEG results. (A) ITC as a function of condition (structured, random), block (1–3), and frequency. ITC values were used to compute the WLI, as described in Methods. (B) Topographical plots showing distribution of ITC across the scalp, as a function of condition and frequency (word, syllable). Note that different scales are used for word versus syllable frequencies. The six darker dots on the upper left scalp plot denote the approximate locations of the six centro-frontal electrodes used for WLI computations, where ITC was generally maximal at both the word and syllable frequencies. (C) The WLI as a function of condition and block.

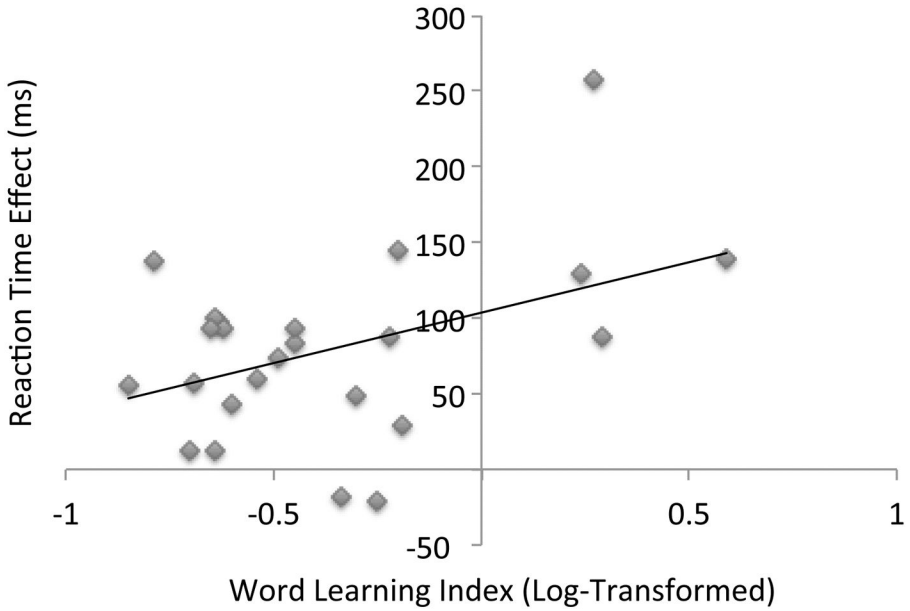

Figure 5.
Scatterplot showing the relation between the corrected reaction time effect and WLI in the structured stream (log-transformed scale).

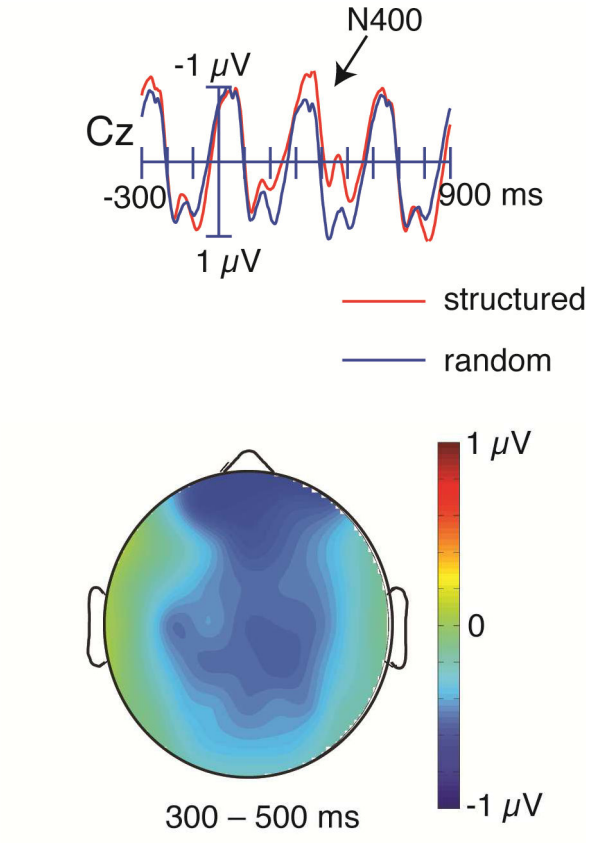

Figure 6.

ERP analysis of word/triplet onsets in the structured (red) and random (blue) conditions. Data are timelocked to the onset of each word in the structured condition, and the onset of each random triplet (i.e., every third syllable) in the random condition. Words in the structured condition differed from triplets in the random condition in the 300–500 ms range, showing a significantly larger N400 amplitude. The topographical voltage plot shows the distribution of this effect, computed by subtracting the mean amplitude in the random condition from the mean amplitude in the structured condition, from 300–500 ms.

Table 1

$N = 24$. Pearson's correlations between log-transformed WLI values in the structured and random condition and post-exposure behavioral measures of learning. Significant correlations are bolded.

	Structured WLI	Random WLI
Rating Accuracy	$r = 0.30$, $p = 0.16$	$r = 0.22$, $p = 0.29$
Rating Score	$r = 0.32$, $p = 0.12$	$r = 0.34$, $p = 0.11$
RT Score	**$r = 0.42$, $p = 0.039$**	**$r = 0.59$, $p = 0.003$**