Rhythmically Modulating Neural Entrainment during Exposure to Regularities Influences Statistical Learning

Laura J. Batterink, Jerrica Mulgrew, and Aaron Gibbings

Abstract

■ The ability to discover regularities in the environment, such as syllable patterns in speech, is known as statistical learning. Previous studies have shown that statistical learning is accompanied by neural entrainment, in which neural activity temporally aligns with repeating patterns over time. However, it is unclear whether these rhythmic neural dynamics play a functional role in statistical learning or whether they largely reflect the downstream consequences of learning, such as the enhanced perception of learned words in speech. To better understand this issue, we manipulated participants' neural entrainment during statistical learning using continuous rhythmic visual stimulation. Participants were exposed to a speech stream of repeating nonsense words while viewing either (1) a visual stimulus with a "congruent" rhythm that aligned with the word structure, (2) a visual stimulus with an incongruent rhythm, or (3) a static visual stimulus. Statistical learning was subsequently measured using both an explicit and implicit test. Participants in the congruent condition showed a significant increase in neural entrainment over auditory regions at the relevant word frequency, over and above effects of passive volume conduction, indicating that visual stimulation successfully altered neural entrainment within relevant neural substrates. Critically, during the subsequent implicit test, participants in the congruent condition showed an enhanced ability to predict upcoming syllables and stronger neural phase synchronization to component words, suggesting that they had gained greater sensitivity to the statistical structure of the speech stream relative to the incongruent and static groups. This learning benefit could not be attributed to strategic processes, as participants were largely unaware of the contingencies between the visual stimulation and embedded words. These results indicate that manipulating neural entrainment during exposure to regularities influences statistical learning outcomes, suggesting that neural entrainment may functionally contribute to statistical learning. Our findings encourage future studies using noninvasive brain stimulation methods to further understand the role of entrainment in statistical learning. ■

INTRODUCTION

Much of the input that hits our senses follows a predictable structure, with the same items or events co-occurring across repeated experiences. Humans are capable of extracting these patterns through mere exposure to environmental stimuli, without intention or effort—an ability known as statistical learning (Aslin, 2017). The first demonstration of statistical learning involved presenting infants with a continuous speech stream made up of repeating trisyllabic nonsense words (e.g., *bidakupadoti*…; Saffran, Aslin, & Newport, 1996). After only 2 min of exposure, infants were able to distinguish between words from the stream and recombined foil items, suggesting that they had extracted the temporal statistics of syllables in the stream. Subsequent research has extended these results, demonstrating that statistical learning is present across the life span (Choi, Batterink, Black, Paller, & Werker, 2020; Palmer, Hutson, & Mattys, 2018; Saffran & Kirkham, 2018; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Saffran et al., 1996) and operates across different modalities and types of stimuli (Conway & Christiansen, 2005; Fiser & Aslin, 2001; Saffran et al., 1999). Statistical learning also occurs in nonhuman animals, including cotton-top tamarins (Hauser, Newport, & Aslin, 2001) and dogs (Boros et al., 2021), and supports learning of a variety of different statistical structures (Saffran et al., 2008; Newport & Aslin, 2004; Gómez, 2002). Still, although statistical learning may be considered a ubiquitous learning mechanism, it remains particularly well-studied within the domain in which it was first investigated—speech segmentation (Siegelman, 2020; Aslin, 2017; Thiessen, Girard, & Erickson, 2016; Arciuli & Torkildsen, 2012; Romberg & Saffran, 2010; Estes, Evans, Alibali, & Saffran, 2007).

Although decades of behavioral work have characterized the contexts under which statistical learning operates (Isbilen & Christiansen, 2022), less is known about how the brain carries out the underlying computations. Neuroimaging studies have implicated a range of neural regions in statistical learning, such as modality-specific cortical areas (e.g., visual, auditory, and somatosensory), the left inferior frontal gyrus, and domain-general memory systems, including the striatum and the medial temporal lobe (for a review, see Batterink, Paller, & Reber, 2019).

Western University, London, Ontario, Canada

Alongside these neuroimaging data, EEG and MEG studies have characterized the neural response elicited by repeating patterns in input. For example, ERP studies have shown that statistical learning is reflected by an N400-like potential to nonsense words in continuous speech (Cunillera et al., 2009; De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007; Cunillera, Toro, Sebastián-Gallés, & Rodríguez-Fornells, 2006), which may represent the construction of a prelexical trace for new words (De Diego Balaguer et al., 2007).

More recently, a number of EEG studies have incorporated a frequency-tagging approach to show that statistical learning is also accompanied by *neural entrainment*, which can be defined broadly as the temporal alignment of neural activity with regularities in a stimulus stream (Obleser & Kayser, 2019). The typical approach in these studies (Benjamin et al., 2023; Batterink & Zhang, 2022; Fló, Benjamin, Palu, & Dehaene-Lambertz, 2022; Moreau, Joanisse, Mulgrew, & Batterink, 2022; Pinto, Prior, & Zion Golumbic, 2022; Smalle, Daikoku, Szmalec, Duyck, & Möttönen, 2022; Benjamin, Dehaene-Lambertz, & Fló, 2021; Elmer, Valizadeh, Cunillera, & Rodriguez-Fornells, 2021; Henin et al., 2021; Zhang, Riecke, & Bonte, 2021; Batterink, 2020; Choi et al., 2020; Ordin, Polyanskaya, Soto, & Molinaro, 2020; Batterink & Paller, 2017, 2019; Getz, Ding, Newport, & Poeppel, 2018; Kabdebon, Pena, Buiatti, & Dehaene-Lambertz, 2015; Buiatti, Peña, & Dehaene-Lambertz, 2009; see also Batterink & Choi, 2021; Benjamin et al., 2021) involves presenting individual syllables, organized into trisyllabic nonsense words, at a precise, fixed rate within a continuous sequence (e.g., "*tu-pi-ro-go-la-bu*…"). This produces well-defined peaks in the neural power and phase coherence spectrum at both the syllable and word frequencies, which may be taken as separate indices of syllable-related and word-related processing (e.g., Kabdebon et al., 2015; Buiatti et al., 2009). Many studies have reported that neural entrainment to words increases over the course of exposure, reflecting the gradual acquisition of the statistically defined units (Fló et al., 2022; Moreau et al., 2022; Elmer et al., 2021; Zhang et al., 2021; Batterink, 2020; Choi et al., 2020; Ordin et al., 2020; Batterink & Paller, 2017, 2019). Several results also suggest that stronger neural entrainment to words over the course of exposure is associated with superior learning outcomes, as assessed by correlating individuals' neural entrainment during learning with their performance on postlearning tasks (Batterink, 2020; Choi et al., 2020; Batterink & Paller, 2017, 2019; Kabdebon et al., 2015; Buiatti et al., 2009). On the basis of these results, neural entrainment to words has been interpreted to reflect various cognitive processes, such as the subjective perception of speech units (Buiatti et al., 2009), the perceptual binding of stimulus units into integrated composites (Batterink & Paller, 2017), and/or the segmentation process itself (Ordin et al., 2020).

As an aside, it is important to note that although these studies establish evidence of neural entrainment during statistical learning in the broad sense—that is, the alignment of neural signals with a rhythmic stimulus—they cannot provide evidence of neural entrainment as defined in the narrow sense, as the phenomenon in which endogenous neural oscillators adjust their frequency or phase to align with incoming stimuli (Bánki, Brzozowska, Hoehl, & Köster, 2022; Obleser & Kayser, 2019). In general, there is debate in the neural entrainment literature about the degree to which enhanced neural activity at stimulation frequencies is because of synchronization of neural oscillators versus the concatenation of evoked responses (Doelling, Assaneo, Bevilacqua, Pesaran, & Poeppel, 2019; Zoefel, ten Oever, & Sack, 2018; Keitel, Quigley, & Ruhnau, 2014; Capilla, Pazo-Alvarez, Darriba, Campo, & Gross, 2011). In the current article, unless otherwise specified, we define neural entrainment more broadly as a set of dynamic neural processes that track rhythmic sensory input and at least partially reflect internally generated predictions or top–down perception of the stimuli (e.g., Vanden Bosch der Nederlanden, Joanisse, Grahn, Snijders, & Schoffelen, 2022; Lu, Sheng, Liu, & Gao, 2021; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Nozaradan, Peretz, Missal, & Mouraux, 2011). In the context of statistical learning of speech sounds, neural entrainment could thereby serve to enhance key segments of the speech stream, such as word onsets. Although understanding whether neural entrainment in the narrow sense operates during statistical learning is certainly an important question for future research, this issue is highly complex (Zoefel, ten Oever, et al., 2018) and beyond the scope of the current study.

Outside the domain of statistical learning, neural entrainment mechanisms have also been linked to speech processing more broadly. Unlike the highly controlled, artificial speech streams used by statistical learning studies, natural speech is not perfectly regular, but it is quasiregular. A number of influential models have proposed that endogenous neural oscillations align with incoming rhythmic linguistic units in speech and that this alignment is a foundational mechanism for parsing and decoding connected speech (Meyer, 2018; Gross et al., 2013; Giraud & Poeppel, 2012; Peelle & Davis, 2012). More specifically, these models suggest that oscillations at different frequencies correspond to units of speech that also unfold at different time scales (e.g., phonemes, syllables, words, and phrases). Coupling between these signals at slower and faster frequencies then supports the hierarchical combination of smaller elements into larger units (e.g., syllables into words). These models have been supported by a number of findings demonstrating that neural entrainment and top–down comprehension of speech are correlated (e.g., Jin, Lu, & Ding, 2020; Luo & Ding, 2020; Ding et al., 2016, 2017; Gross et al., 2013; Peelle & Davis, 2012; Ahissar et al., 2001). For example, Ding and colleagues (Ding et al., 2016, 2017) presented participants with simple four-word phrases made up of monosyllabic words at an isochronous rate (e.g., "dry fur rubs skin").

Participants' neural response showed spectral peaks at frequencies corresponding to the syllable, word, and phrase presentation rates, providing evidence of concurrent neural tracking of hierarchical linguistic structures. Critically, peaks at the word and phrase rate were not observed when listeners were exposed to an unknown foreign language, ruling out contributions of low-level acoustic features.

In addition to correlational evidence linking neural entrainment and speech processing, there is also emerging evidence that neural entrainment plays a functional (causal) role in speech comprehension, as revealed by studies that have manipulated neural entrainment (van Bree, Sohoglu, Davis, & Zoefel, 2021; Kösem et al., 2018; Riecke, Formisano, Sorger, Başkent, & Gaudrain, 2018; Wilsch, Neuling, Obleser, & Herrmann, 2018; Zoefel, Archer-Boyd, & Davis, 2018). Wilsch and colleagues (2018) presented participants with speech in noise while applying transcranial electrical currents in the shape of the speech envelope, and found a systematic modulation of speech intelligibility as a function of stimulation lag. Similarly, Riecke and colleagues (2018) reported that transcranial stimulation with speech-shaped currents improved word recognition for speech that was presented in a two-talker, cocktail-party stream, as well as for speech that had been artificially stripped of critical rhythmic cues. In another recent study, participants were presented with sentences that always ended with an ambiguous word (e.g., "tak" or "taak" in Dutch, containing a vowel ambiguous between a short /a/ and a long /a:/; Kösem et al., 2018). The initial part of the sentence—with a speech envelope designed to contain a strong rhythmic component—was presented at either a slow or fast rate, thereby producing corresponding changes in neural entrainment. MEG results showed that participants' entrainment to the speech rhythm persisted after the initial sentence stem had been presented, and that this sustained entrainment systematically biased participants' perception of the ambiguous word. Overall, these studies show that the alignment of neural oscillations with external rhythms of speech influences comprehension, providing evidence for a functional relevance of neural entrainment in speech processing.

Currently, there is less evidence on the functional significance of neural entrainment as it operates during statistical learning specifically. However, some initial evidence related to this issue comes from a behavioral study by Wang, Zevin, and Mintz (2017). By leveraging the same paradigm used by Ding and colleagues (2016, 2017), these authors induced "grammatical entrainment" in participants by cyclically presenting four-word English sentences following the same syntactic structure (e.g., "Brian puts it down"; "John turns these in"), followed immediately by artificial language phrases containing nonadjacent (AXB) dependencies. By manipulating the alignment of the cyclic English structures and the nonadjacent dependencies in the artificial language, the authors tested whether this form of entrainment influenced the statistical learning of the artificial language. Critically, participants successfully learned the nonadjacent dependencies when they aligned, or occurred "in phase," with the cyclic English structures, but not when they occurred out of phase. These results indicate that cyclically presenting abstract, grammatical structures can induce a form of entrainment in learners that subsequently influences statistical learning. Combined with the results from Ding and colleagues (2016, 2017) demonstrating neurophysiological tracking of the repeated syntactic structures, it may be hypothesized that this paradigm entrained neural activity to the syntactic structures contained in the English sentences, resulting in behavioral facilitation of phase-aligned structures. However, because the study used only behavioral methods, the neural mechanisms underlying these behavioral effects are not yet known.

To summarize, although there is currently correlational evidence linking neural entrainment to statistical learning (Sherman et al., 2023; Moreau et al., 2022; Elmer et al., 2021; Henin et al., 2021; Moser et al., 2021; Zhang et al., 2021; Batterink, 2020; Choi et al., 2020; Ordin et al., 2020; Batterink & Paller, 2017, 2019) and emerging causal evidence showing that neural entrainment may play a functional role in speech comprehension more broadly (van Bree et al., 2021; Kösem et al., 2018; Riecke et al., 2018; Wilsch et al., 2018; Zoefel, Archer-Boyd, et al., 2018), there is little evidence on whether neural entrainment contributes causally to statistical learning. Although it has been repeatedly shown that neural entrainment increases over exposure to structured input as statistical learning progresses (Moreau et al., 2022; Elmer et al., 2021; Moser et al., 2021; Zhang et al., 2021; Batterink, 2020; Choi et al., 2020; Ordin et al., 2020; Batterink & Paller, 2017, 2019), an open question is whether this increased entrainment largely reflects a downstream consequence of the statistical learning process itself, such as the enhanced perception of individual words in the speech stream, or a mechanism that plays a functional (supporting) role in learning. Addressing this question requires an experimental approach, in which neural entrainment itself is manipulated. If enhancing neural entrainment at the word frequency during statistical learning facilitates performance on subsequent tests of learning, this would provide evidence that the neural process (or processes) indexed by this neural entrainment signal play a functional role in statistical learning. In contrast, if boosting neural entrainment at the word frequency has no impact on subsequent learning performance, this would suggest that the processes reflected by neural entrainment may largely reflect downstream perceptual or cognitive consequences of the learning processes itself, such as stronger word perception or an increase in lexical search processes to the newly learned words (Sanders, Newport, & Neville, 2002), rather than influencing the learning process itself.

The Current Study

The overall goal of the current study was to improve our understanding of what observed neural entrainment signals during statistical learning reflect, particularly whether such signals represent a cause or consequence of learning (or both). To shed light on this issue, we rhythmically modulated the neural dynamics during statistical learning using task-irrelevant, continuous visual stimulation, and then examined subsequent expressions of learning at both the behavioral and neural level. While participants' EEG was recorded, they listened to a continuous speech stream made up of repeating trisyllabic nonsense words while passively viewing one of three visual stimuli designed to influence neural entrainment: (1) a looping video showing a droplet of water falling from a leaf, with a rhythmic cycle that aligned with the repeating words (*congruent condition*); (2) the same looping video with an adjusted cycle that differed in duration from the repeating words (*incongruent condition*), or (3) a completely static image of the leaf (*static condition*; see Figure 1).

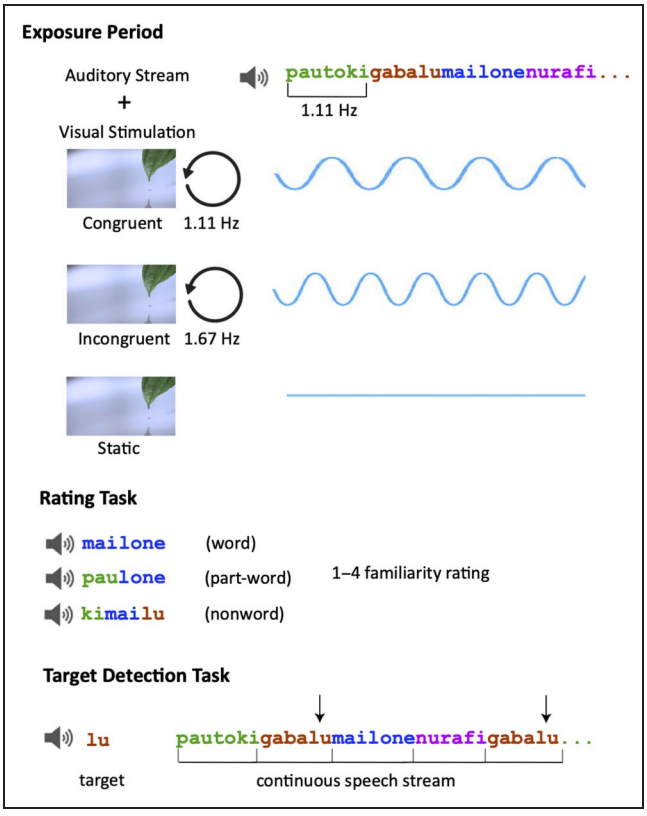

Figure 1. Summary of experimental tasks. During the exposure period, participants were exposed to a continuous stream of trisyllabic nonsense words while viewing one of three types of visual stimuli. In two conditions, the visual stimuli consisted of a looping video of a water droplet forming and then falling off a leaf, with a repetition cycle that either aligned (congruent condition) or misaligned (incongruent condition) with the word presentation rate of the speech stream. In the third condition, a static image was presented. Following the exposure period, participants completed two tests of statistical learning: the rating task, which provides an explicit measure of participants' knowledge, and the target detection task, which serves as an implicit test of knowledge.

Posttask interviews suggested that at most, only one participant correctly inferred that there was a synchronization in timing between the visual stimuli and the repeating words, such that any effects on entrainment could be generally attributed to implicit multisensory integration, as opposed to strategic or conscious processes on the part of the participant. After exposure to the speech stream, participants completed two behavioral tests of learning: a rating task, which required participants to provide explicit familiarity ratings for words from the stream and foil items, and a RT-based target detection task, which required participants to respond to target syllables embedded in continuous syllable streams. These tasks preferentially capture explicit and implicit aspects of statistical learning, respectively (Batterink, Reber, Neville, & Paller, 2015).

Given previous correlational evidence linking neural entrainment during statistical learning to subsequent learning outcomes (Batterink, 2020; Choi et al., 2020; Batterink & Paller, 2017, 2019; Kabdebon et al., 2015; Buiatti et al., 2009), we hypothesized that rhythmically enhancing neural entrainment at the relevant frequency would facilitate statistical learning. We therefore predicted that participants in the congruent group (in which neural entrainment at the word frequency should be enhanced by the visual stimuli) would show enhanced performance on the rating task and/or target detection task compared with participants assigned to the static condition. Furthermore, to the extent that the visual manipulation at the incongruent frequency disrupted entrainment to the word frequency, we expected that participants in the incongruent group should show the poorest performance on subsequent measures of learning.

METHODS

Participants

On the basis of previous studies (Batterink & Paller, 2017, 2019), we aimed to include data from 20 participants per group. Seventy participants were recruited at Western University to participate in the study. The experiment was undertaken with the understanding and written consent of each participant. All participants were proficient or native English speakers between 18–35 years old. Ten participants were excluded from the final sample because of technical issues, noisy data, or experiment ineligibility (e.g., on medications that impact brain functioning), resulting in a final sample of 60 participants (39 women; mean age = 19.2 years). Participants were compensated $14/hr or received course credit for their time. All participants in the current study had normal hearing and normal or corrected-to-normal vision, as well as no prior history of neurological disorders. Participants were randomly assigned to one of the three experimental conditions (congruent, incongruent, and static),

determined by the order run (n = 20 per group, with sequential order assignment, i.e., 1-2-3-1-2-3).

Stimuli

Auditory Stimuli

Twelve syllables were generated using an artificial speech synthesizer (Apple's text-to-speech application, voice "Samantha," set to "Fast" speaking rate) and recorded as separate sound files in Audacity with a sampling rate of 44100 Hz. Syllable sound files were 300 msec or shorter in duration. These syllables were concatenated to create 4 nonsense "words" (*pautoki*, *mailone*, *nurafi*, *gabalu*; syllables were taken from Batterink & Paller, 2017). To create the continuous speech stream, these nonsense words were presented at a rate of 300 msec per syllable (3.33 Hz) in a predefined pseudorandom order, with the constraint that the same word did not repeat consecutively. The speech stream contained 1200 syllables (400 words), with each word represented an equal number of times throughout the stream (100 presentations for each word).

Visual Stimuli

The visual stimuli were created by capturing frames (still images) from a slow-motion video, which depicts a looping cycle of a droplet of water falling off a leaf and hitting the surface of a body of water below. The video was obtained from an online stock image website called Deposit Photos (https://depositphotos.com/). Image sets were created by calculating the total number of images needed for a full video cycle at the desired frequency (i.e., the time between the appearance of one droplet on the leaf to the appearance of the next droplet), given the monitor refresh rate (60 Hz). For the 1.11 Hz "congruent" condition, a full cycle consisted of 54 images; for the 1.67 Hz "incongruent" condition, a full cycle consisted of 36 images. In the static condition, a single image of the leaf was presented. For all three conditions, images were presented at a rate of 60 images per second.

As mentioned previously, the congruent and incongruent conditions differed in terms of the temporal relationship between the video and the syllables in the speech stream. In the congruent condition, one cycle of the video was equivalent to the presentation rate of a trisyllabic nonsense word (video cycle frequency = 1.11 Hz, word presentation frequency 1.11 Hz). In the incongruent condition, one cycle of the video was equivalent to the presentation of two syllables in the stream, such that the onset of a video cycle coincided equally frequently with each of the three triplet positions (video cycle frequency = 1.67 Hz, word presentation frequency 1.11 Hz).

Procedure

After electrode setup, participants were seated in a sound-attenuated booth at a comfortable viewing distance (approximately 70 cm) from the monitor. Auditory stimuli were presented at a comfortable listening level over computer speakers positioned on either side of the monitor. A visual summary of the protocol is shown in Figure 1. Before beginning the statistical learning tasks, resting-state EEG data were collected for 6 min 15 sec (data not presented here). Participants were instructed to relax and maintain focus on a fixation cross in the center of the screen during this time.

Exposure Task

In all three conditions (congruent, incongruent, and static), participants were presented with the continuous speech stream while viewing visual stimuli presented on the computer monitor. The visual stimuli began first, with the auditory stimuli beginning 15 sec later, to establish neural entrainment effects before the onset of the speech stream. Participants in the congruent and incongruent conditions, who viewed looping videos, were instructed to visually fixate on the location in which the droplet of water hits the surface below. Participants in the static condition viewed a single static image taken from the video and were instructed to maintain fixation on the tip of the leaf. To disguise the true purpose of the entrainment manipulation, participants were told that the aim of the experiment was to test whether viewing videos (or images, in the static condition) from nature helps people to relax while listening to a nonsense language.

Rating Task

All tasks administered after the exposure task were identical across groups. First, participants completed a familiarity rating task to assess explicit knowledge of the nonsense words. For each trial, participants heard a trisyllabic item and had to indicate via a button press how familiar that item sounded based on the language they had just heard (with 1 being *very unfamiliar* and 4 being *very familiar*). The item was a word from the language (*pautoki*, *nurafi*, *mailone*, *gabalu*), a part-word consisting of a syllable pair from the language along with an additional syllable from another word (*nuralu*, *maitoki*, *gabafi*, *paulone*), or a nonword consisting of syllables from the language that never occurred together (*nuloto*, *kimailu*, *paraba*, *gafine*). As in the exposure task, the stimulus onset asynchrony between consecutive syllables within each item was 300 msec. In total, the task consisted of 12 trials, including 4 words, 4 part-words, and 4 nonwords. Because of this low trial number, which precludes meaningful EEG analysis, no EEG data were saved during this task.

Target Detection Task

Following the rating task, participants completed a speeded target detection task to assess implicit knowledge of the nonsense words. Each trial consisted of a short

speech stream containing the four trisyllabic words, each presented 4 times in pseudorandom order. Although we had intended to present the target detection streams at the same rate as the original exposure stream (300 msec per syllable), because of a technical issue, syllables were actually presented at a rate of ~284.5 msec per syllable (slightly faster than the original speech stream). For each trial, participants were instructed to press a button every time they heard a specific target syllable in the speech stream. Both speed and accuracy were emphasized. At the beginning of each trial, participants pressed a button to hear the target syllable and then started the speech stream via a second button press. Throughout the trial, the phonetic spelling of the syllable remained on the screen to remind participants which target syllable they were listening for. As in the exposure stream, the same word never repeated consecutively. In addition, target syllables never occurred within the first or last words of the stream. In total, each syllable in the language served as a target 3 times, resulting in 36 total streams. Each stream contained four target syllables, yielding 48 trials for each triplet-position condition (word-initial syllable, word-medial syllable, and word-final syllable). This task measures participants' ability to use their acquired statistical knowledge to optimize online processing, as reflected by faster responses to more predictable syllables (i.e., those occurring in later positions within a triplet). This facilitation reflects participants' ability to predict upcoming syllables based on the initial syllables already presented within a word, and can be considered an implicit measure of learning, capturing learning effects even in the absence of explicit word knowledge (Batterink et al., 2015).

Posttask Interview

Following the experimental tasks, participants in the congruent and incongruent groups completed an oral posttask interview to assess their explicit awareness of the relationship between the artificial language and the dynamic visual stimulus. To fully capture low levels of awareness, the interview began with questions that provided no specific information about the experimental manipulation, revealing more specific information as the interview went on. First, participants were asked to respond yes or no as to whether they had noticed any connection between the video and the sounds. Next, participants were asked (yes/no) more specifically whether they had noticed any association between the timing of the video and the syllables in the stream. If they responded yes, they were then asked to describe the association. Participants' verbal responses to this final open-ended question were transcribed by the experimenter. On the basis of their recorded responses to this final question, participants were then coded as being fully aware of the contingency between the video and word structure, partially or imprecisely aware of the contingency, or completely unaware of the contingency. As there was no actual 1:1 correspondence between the word structure and the video cycle in the incongruent condition, the responses from this group served as a basis of comparison.

For each of these three questions, a Pearson chi-square test was used to assess whether the proportion of participants who provided a given response (Q1 and Q2: yes/no; Q3: aware/partially aware/not aware) differed between the congruent and incongruent groups. Three participants out of 40 did not contribute data to this task, as we began administering the questionnaire only after a few initial participants had already completed the experiment.

On the open-ended response, only 3 out 37 of participants (1 participant from the congruent group, 2 from the incongruent group) offered statements indicating awareness of the relationship between the structure of the speech stream and the video (e.g., "when words were said, drops happened"; "every time the water dropped it would be a section of the word finishing"). However, note that there was no actual 1:1 correspondence between the words and video for the incongruent participants, such that the two incongruent participants' statements were inaccurate. Another 16 participants were coded as partially or imprecisely aware of some relationship between the speech stream and video (e.g., "seemed to follow a similar rhythm"; "every drop there was a syllable" [not actually the case]). The remaining 18 participants offered either no information or information that was irrelevant to the question (e.g., "the language was weird to listen to"; "it would repeat the same ones for a few times and then switch and do new ones") and were coded as unaware.

Behavioral Data Analyses

Familiarity Rating Task

Rating scores (1–4 scale) for each trial were assessed using an ordinal regression with mixed effects, with word type (word, part-word, nonword) as a within-subject factor, condition (congruent, incongruent, static) as a between-subjects factor, and subject intercept as a random effect. Both word type and condition were modeled as categorical factors, and rating score was treated as an ordinal variable. Treatment coding was used for condition, with the static group set as the reference condition, allowing us to evaluate whether both the congruent and incongruent groups individually differed from this reference. Polynomial coding was used for word type to test the hypothesis that familiarity ratings would be highest for words, intermediate for partword, and lowest for nonwords.

Target Detection Task

Adopting the same criterion as previous studies in our laboratory (Batterink & Paller, 2017; Batterink et al., 2015), responses that occurred before 0 msec or after 1200 msec of a target were considered false alarms and were not included in further analyses. Given the length of this task,

it is possible that performance may either improve (because of online learning) or decline (because of fatigue) as a function of trial number. Furthermore, previous work by Batterink (2017) found that RTs to target syllables increased for targets occurring later in a given stream. To account for these sources of variance, RTs were modeled using a linear mixed-effects model, with predictors including fixed effects of condition (congruent, incongruent, static), triplet position (word-initial, word-medial, word-final), stream within the task (1–36), target position within the stream (4–45, i.e., in which position the target occurred within a single stream; targets never occurred within the first or final word), and the interaction between triplet position and condition. Stream position and target position within the stream were not variables of direct interest but were included in the model as control variables. Following previous approaches and demonstrations of linear influences of triplet position on RTs (with later occurring targets within a triplet eliciting progressively faster RTs; see Liu, Forest, Duncan, & Finn, 2023; Wang, Köhler, & Batterink, 2023; Wang, Rosenbaum, et al., 2023; Moreau et al., 2022; Batterink & Paller, 2017, 2019; Batterink et al., 2015), all predictors were modeled as continuous predictors except for condition, which was categorical. Treatment coding was used, with the static group set as the reference condition, allowing us to evaluate whether both the congruent and incongruent groups individually differed from this reference. Subject intercept was included as a random effect to account for differences in baseline RTs between participants. We confirmed that all reported models successfully converged.

EEG Recording and Analysis

EEG was recorded at a sampling rate of 512 Hz using a 64-channel Active-Two Biosemi system, set up according to the 10/20 system. Additional electrodes were placed on the left and right mastoids, on the outer canthi of the left eye, and below the left eye. Signals were recorded relative to the Common Mode Sensor active electrode and then rereferenced offline to the average of the left and right mastoid electrodes. All EEG analyses were conducted using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). Before epoching, data were band-pass filtered from 0.1 to 30 Hz using an IIR Butterworth filter, as implemented by *pop_basicfilter* in ERPLAB.

Exposure Period

Nonoverlapping epochs of 10.8 sec were extracted, time-locked to the onset of every 12th word in the language and corresponding to a duration of 12 words, or 36 syllables. All epochs contained data that corresponded to continuous auditory presentation (i.e., epochs occurring before a break in recording were not extracted). Data were visually inspected to allow for manual rejection of noisy epochs as necessary, although in the current data set, data quality was high and all epochs were retained. Occasional bad electrodes were identified and interpolated (mean = 0.8, max = 4 electrodes per participant). Each participant's data set contained between 32 and 33 epochs. Following a previous study that used multimodal rhythmic stimuli (Bauer, van Ede, Quinn, & Nobre, 2021), a surface Laplacian transform was then applied to help separate contributions from auditory and visual areas, as implemented by Cohen (2014). The surface Laplacian transform is a spatial filter that minimizes volume conduction effects, increases spatial resolution, and provides a reference-free representation of underlying neural generators (Bauer et al., 2021; Cohen, 2014).

We quantified neural entrainment by measuring inter-trial coherence (ITC) across all epochs in each individual data set. ITC is a measure of event-related phase-locking or phase synchronization across trials, which ranges from 0 to 1, with 0 indicating purely non-phase-locked (i.e., random) activity at a given frequency band, and 1 indicating strictly phase-locked activity (i.e., oscillations perfectly in phase across all epochs). For each participant, the fast Fourier transform was applied to the individual EEG epochs. Given the epoch length, this yielded a frequency resolution of ~0.0926 Hz, which produces spectral estimates at frequency bins that include the word presentation rate (1.11 Hz) and syllable presentation rate (3.33 Hz). The phase component at each frequency was then used to compute ITC, which is the circular sum (absolute value) of the phases across trials at a certain point in time. This procedure was carried out for all 64 scalp channels. Given the auditory nature of the task and based on our prior results (Choi et al., 2020; Batterink & Paller, 2017, 2019), we expected maximum entrainment effects over auditory regions. We selected 12 bilateral frontocentral electrodes based on the frequency of the word distribution entrainment effect in the static condition, in which there was no dynamic visual stimulus (FC1, FC3, FC5, T7, C3, C5; FC2, FC4, FC6, C4, C6, T8). These electrodes correspond well with the observed distribution and electrode groups selected by previous studies of auditory entrainment that applied Laplacian transforms (Bauer et al., 2021; Jaeger, Bleichner, Bauer, Mirkovic, & Debener, 2018). For all subsequent analyses, ITC values for each electrode within this region were averaged together.

In a first stage analysis, we tested whether the visual manipulation significantly influenced neural entrainment as intended by conducting two separate ANOVAs on the ITC values at our frequencies of interest (word = 1.11 Hz; syllable = 3.33 Hz), with Group (congruent, incongruent, static) as a between-subjects factor. We predicted that neural entrainment at the word frequency should be higher in the congruent Group compared with the incongruent and static groups.

In a subsequent follow-up analysis, we assessed whether any increased neural entrainment at the word frequency in

the congruent group (if observed) was because of mere passive volume conduction from posterior-occipital regions or whether it exceeded the values that would be expected based on passive volume condition alone. Here, we applied the general logic of super-additive interactions in the multisensory integration field, which is commonly used litmus test for demonstrating multisensory convergence and integration (e.g., Stanford & Stein, 2007). The logic here is that, if the neural response to a multisensory stimulus is greater than the sum of the separate responses to modality-specific components, this cannot be readily explained by the recruitment of separate pools of unisensory neurons and thus indicates multisensory integration by multisensory neural populations (Stanford & Stein, 2007). Applying this logic to the current study, we reasoned that if the neural response at the word frequency in the congruent group (reflecting a combination of visual stimulation and repeating auditory patterns) exceeded the sum of the independent responses produced by the words alone and the visual stimuli alone (i.e., if A * V > A + V), this would demonstrate cross-modal influence and integration of the visual stimuli within our neural substrates of interest, over and above effects of passive volume conduction. To test this idea, we leveraged data from the incongruent and static groups to estimate the neural response produced by the auditory input and the visual rhythm, in the absence of alignment between the hidden words and the visual rhythmic cycle. We submitted ITC values at 1.11 Hz (word frequency) and 1.667 Hz (corresponding to syllable pairs and the visual stimulation cycle in the incongruent group) across all participants to a linear mixed-effects model, with frequency bin (1.11 Hz, 1.67 Hz), visual stimulation (with binary coding; 1 = frequency of visual stimulation cycle, applied to 1.11 Hz frequency in the congruent group and 1.667 Hz in the incongruent group; 0 = no alignment, applied to all other frequency and group combinations), and their interaction modeled as fixed factors, and participant intercept modeled as a random effect. Frequency and visual stimulation were not modeled as random slopes because of convergence issues. A significant interaction between frequency (1.11 Hz, 1.667 Hz) and visual stimulation-word alignment (1 or 0), supported by appropriate follow-up contrasts, would provide evidence of a super-additive cross-modal integration effect. This result would show that any increased neural entrainment at the word frequency in the congruent group exceeds the level attributable to spreading activation from visual regions and would provide evidence that the auditory stimulation successfully engaged the same relevant neural substrates that were involved in processing the speech stream.

Target Detection Task

Next, we quantified neural entrainment during the test phase to test whether the visual manipulation applied during the exposure phase influenced subsequent neural processing of the speech stream during the target detection task (here, in the absence of visual stimulation). We hypothesized that participants in the congruent group would show an increase in neural entrainment at the word frequency during the subsequent test, which would provide additional evidence of superior statistical learning in this group.

For each target detection stream in the task, we extracted an epoch with a duration of 12.8 sec, corresponding to the time from the second word presentation to the final word, yielding 36 epochs. We excluded data corresponding to the initial word presentation to avoid auditory onset effects triggered by the initial first few syllables. Data were visually inspected, and occasional bad electrodes were identified and interpolated (mean = 0.83 electrodes per participant). Any excessively noisy epochs were manually removed (mean = 35.3 epochs per participant remaining for analysis, min = 31 epochs remaining). ITC was then computed using the same general method as described above for the analysis of the EEG data from the exposure period, using the same electrode ROI. As the epoch length for this analysis was 12.8 sec, this yielded a frequency resolution of 0.0781 Hz, which produces spectral estimates at frequency bins that include the word presentation rate (1.17 Hz) and syllable presentation rate (3.515 Hz). A technical problem affected the data from one participant in the incongruent condition, resulting in a sample of only 19 participants (rather than 20) for this group.

Correlations between Measures

To assess correlations between EEG measures of neural entrainment and behavioral performance, a composite measure was derived for both the rating task and the target detection task. For the rating task, a "rating score" was calculated for each participant by subtracting the average score for part-words and nonwords from the average score for words. For this measure, a score of 3 would indicate perfect sensitivity, with all values above 0 providing evidence of learning (Batterink & Paller, 2017). For the target detection task, mean RTs for each syllable position (word-initial, word-medial, word-final) were calculated for each participant, and a "RT prediction effect" was computed as the proportion of RT decrease to third position targets relative to initial position targets $[(RT_1 - RT_3)/RT_1]$ (Batterink & Paller, 2019). Because decreases in RTs are not independent of the overall speed of response (cf. Siegelman, Bogaerts, Kronenfeld, & Frost, 2018), this computation adjusts for potential differences in baseline RTs between individuals, allowing us to compare statistical learning across individuals with different RT baselines. Larger positive values on the RT prediction effect indicate greater facilitation because of statistical learning. Pearson's correlations were then computed between neural entrainment to the word frequency and (1) the

rating score and (2) the RT prediction effect, across all participants.

RESULTS

Behavioral Results

Rating Task

Across all three groups, participants demonstrated significant evidence of statistical learning, with words rated as most familiar, followed by part-words, followed by nonwords as the least familiar (linear effect of word type: $z = -6.88$, $p < .001$; see Figure 2A). However, in contrast to our hypothesis, ratings did not significantly differ across the three visual condition groups (congruent > static × linear contrast for trial type: $z = -0.39$, $p = .69$; incongruent > static × linear contrast for trial type: $z = -0.087$, $p = .93$).

Target Detection Task

Across all groups, we observed the expected decline in RT as a function of syllable position, with progressively faster RTs for more later, more predictable syllables (see Figure 2B). In the initial model, stream position (4–45, i.e., in which position the target occurred within a single trial) was not found to be significant, $F(1, 7783) = 0.171$, $p = .68$. We thus conducted a follow-up model that included all the same factors as the original model, except for the nonsignificant factor of stream position. The second, simpler model revealed a significant effect of triplet position, triplet position effect: $F(1, 7783) = 496.20$, $p < .001$, and an interaction between condition and triplet position, Condition × Triplet Position: $F(2, 7783) = 3.42$, $p = .033$, with no main effect of condition, Condition: $F(2, 105) = 0.23$, $p = .79$. In line with our hypothesis, the congruent group showed a significantly stronger triplet position effect compared with the static group, $t(17784) = -2.51$, $p = .012$; parameter estimate syllable position: $M = -12.0$ msec, $SE = 4.78$. In contrast, the RT triplet position effect did not significantly differ between the incongruent and static groups, $t(17784) = -0.61$, $p = .54$.

Overall, accuracy for targets was high (mean = 91.3%, $SD = 6.0\%$), with approximately 10.5 ($SD = 8.82$) false alarms per participant, and no significant differences between groups, accuracy: $F(2, 59) = 0.91$, $p = .40$; false alarms: $F(2, 59) = 2.48$, $p = .093$.

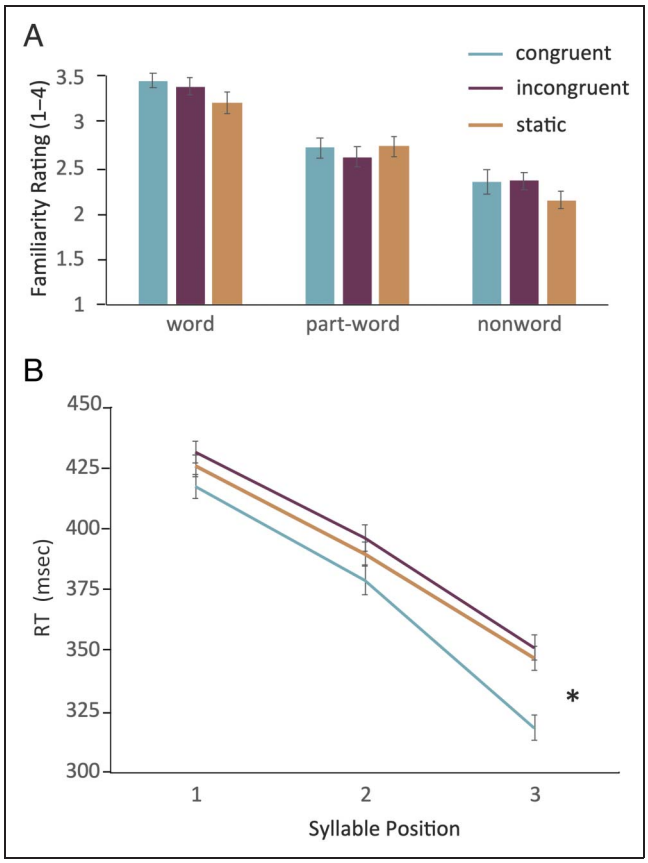

Figure 2. Behavioral results for the rating task (A) and target detection task (B). The three groups showed similar performance on the rating task. On the target detection task, the congruent group showed a significantly greater triplet position effect to more predictable (third syllable) targets, indicative of enhanced prediction and superior statistical learning. Error bars represent standard error of the mean.

Behavioral Task Correlations

There was a moderate, significant correlation between the rating score and the RT prediction effect, $r(60) = .32$, $p = .012$, indicating that participants who were more successful at discriminating words and foils on the rating task also showed a stronger prediction effect on the target detection task.

EEG Results

Exposure Phase

As shown in Figure 3, all groups showed clear peaks in neural entrainment at both the syllable and word frequencies in canonical auditory electrodes, as well as the frequency corresponding to the second harmonic of the word presentation rate (~2.22 Hz). As hypothesized, the three groups showed significant differences in neural entrainment at the word frequency within our independently selected auditory ROI, consisting of 12 bilateral frontocentral electrodes, $F(2, 57) = 20.9$, $p < .001$. Planned contrasts revealed that the congruent group showed significantly stronger word frequency entrainment than the other two groups (contrast estimate = 0.16, $SE = 0.024$, $p < .001$). In contrast, word entrainment in the static and incongruent group did not significantly differ from one another ($p = .83$). There were no significant differences in syllable-level entrainment between groups, $F(2, 57) = 0.24$, $p = .78$.

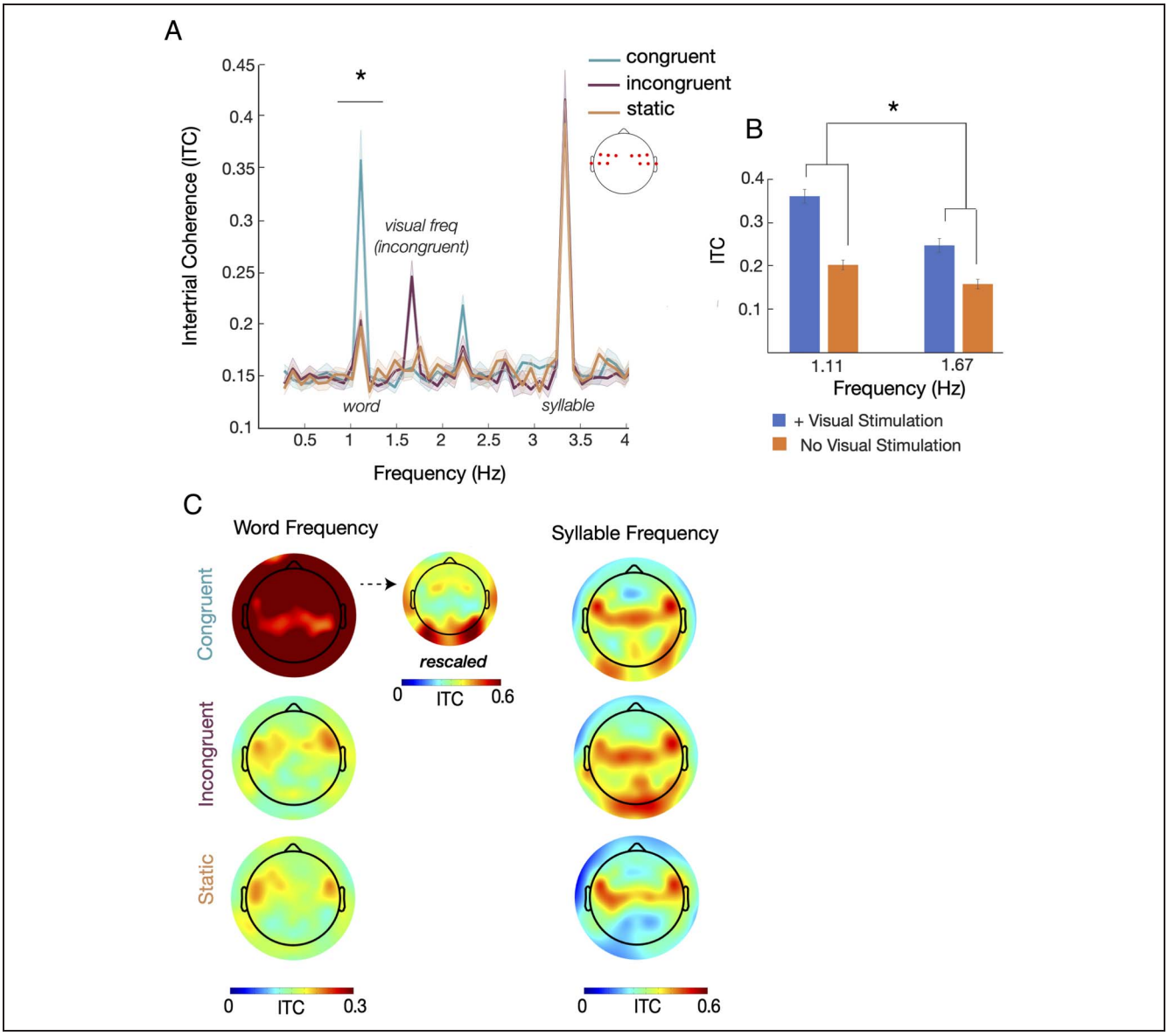

Figure 3. Neural entrainment across the exposure period. (A) EEG ITC as a function of frequency and condition (congruent, incongruent, static). Data in the line graph are averaged over 12 frontocentral bilateral electrodes, selected on the basis of the ITC distribution observed in the static condition (in which no visual stimulation occurred), whose location is indicated by the red dots. Shaded regions represent the standard error of the mean within each group. A significant group effect was observed at the word frequency; participants in the congruent condition showed significantly greater neural entrainment at the word frequency than the other two groups. (B) Across groups, a significant interaction was observed between frequency (1.11, 1.67) and visual stimulation, indicating the effect of visual stimulation was greater when stimulation occurred at the same frequency as the repeating words and demonstrating a super-additive cross-modal response to auditory structure and visual stimulation. (C) Distribution of ITC across the scalp, at the word frequency (left) and syllable frequency (right), within each group. ITC at the word frequency in the congruent group is plotted at a different scale in the inset for better resolution.

Next, we conducted a follow-up analysis to test whether phase-locking at the word frequency in the congruent group reflects true engagement of the same relevant neural substrates involved in processing the speech stream, over and above effects of mere passive volume conduction. To do this, we probed for evidence of a super-additive response in the congruent group at the word frequency (see Methods section). As expected, based on the presence of statistical structure at the trisyllabic frequency, overall greater ITC values were found at 1.11 Hz (trisyllabic frequency) than at 1.67 Hz, bisyllabic frequency; main effect of frequency: $F(1, 70) = 33.9$, $p < .001$. In addition, ITC values differed as a function of visual stimulation, main effect of visual stimulation: $F(1, 88) = 82.1$, $p < .001$. Critically, an interaction was found between frequency and visual stimulation, $F(1, 98) = 5.81$, $p = .018$, such that ITC values were significantly greater when the frequency of visual stimulation coincided with the hidden words in the speech stream (i.e., 1.1 Hz in the congruent group; parameter estimate for Word

Frequency × Visual Stimulation interaction = 0.069, SE = 0.028, 95% CI [0.012, 0.125]; see Figure 3B). This interaction indicates that the effect of the visual stimulation was not equal across the two frequencies (1.11 Hz and 1.67 Hz). This result suggests that participants in the congruent group showed an increase in neural entrainment at the word frequency that cannot be accounted for by mere passive volume conduction alone, as the enhancement is greater than simple additive effects of trisyllabic auditory structure and visual stimulation (as estimated under conditions and frequencies when visual stimulation and auditory structure do not align).

Finally, both the congruent and incongruent groups showed large ITC peaks as their respective visual frequencies over occipital electrodes, reflecting strong visual entrainment to the stimulus (Figure 4). We found no significant differences in ITC values between the two groups at their respective visual stimulation frequencies, $F(1, 38) = 0.34$, $p = .56$, indicating that the strength of the visual entrainment was similar between the two groups.

Target Detection Task

The three visual condition groups showed significantly different neural entrainment at the word frequency, group effect: $F(1, 56) = 5.29$, $p = .008$ (Figure 5A). Consistent with our hypothesis, planned contrasts indicated that the congruent group demonstrated significantly higher word-level entrainment compared with the incongruent and control groups, providing evidence of superior segmentation of the speech stream (contrast estimate = 0.047, $SE = 0.15$, $p = .002$). In contrast, neural entrainment at the word frequency did not differ between the incongruent and static image groups (contrast estimate = 0.008, $SE = 0.017$, $p = .66$). As expected, neural entrainment at the syllable frequency did not differ between groups, group: $F(1, 56) = 0.068$, $p = .93$.

Behavioral-neural Entrainment Correlations

Finally, we examined whether neural entrainment during exposure predicted behavioral performance on the two measures of statistical learning. Counter to our prediction, neural entrainment at the word frequency (during the exposure period, computed over our auditory ROI) did not significantly predict the RT prediction effect ($r = .21$, $p = .10$), nor did it significantly predict the rating score ($r = .16$, $p = .23$). Interestingly, however, neural entrainment at the word frequency during the target detection task did show a significantly correlation with the RT prediction effect ($r = .34$, $p = .008$; Figure 5B), indicating the stronger entrainment to words during the task itself was related to a stronger behavioral prediction effect.

Posttask Interview

Participants were initially asked whether they had noticed any connection between the video they had observed and the sounds (Q1), and then whether they had noticed any association between the timing of the video and the syllables in the stream (Q2). The proportion of participants who responded "yes" to these questions did not differ between groups, Q1: $\chi^2(1, N = 37) = 0.65$, $p = .42$; Q2:

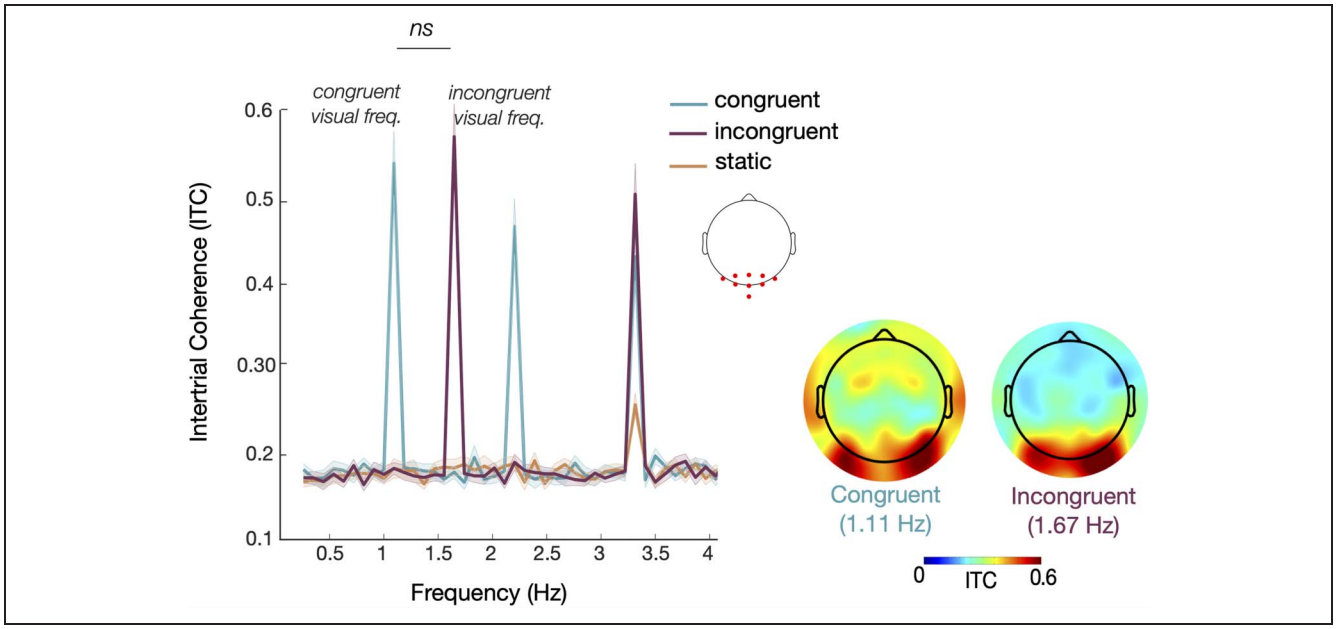

Figure 4. Neural entrainment over occipital electrodes during the exposure period. EEG ITC is plotted as a function of frequency and condition (congruent, incongruent, static). Red dots indicate the electrode locations for this analysis. Shaded regions represent the standard error of the mean within each group. The congruent and incongruent group showed similar ITC values at their respective visual stimulation frequencies.

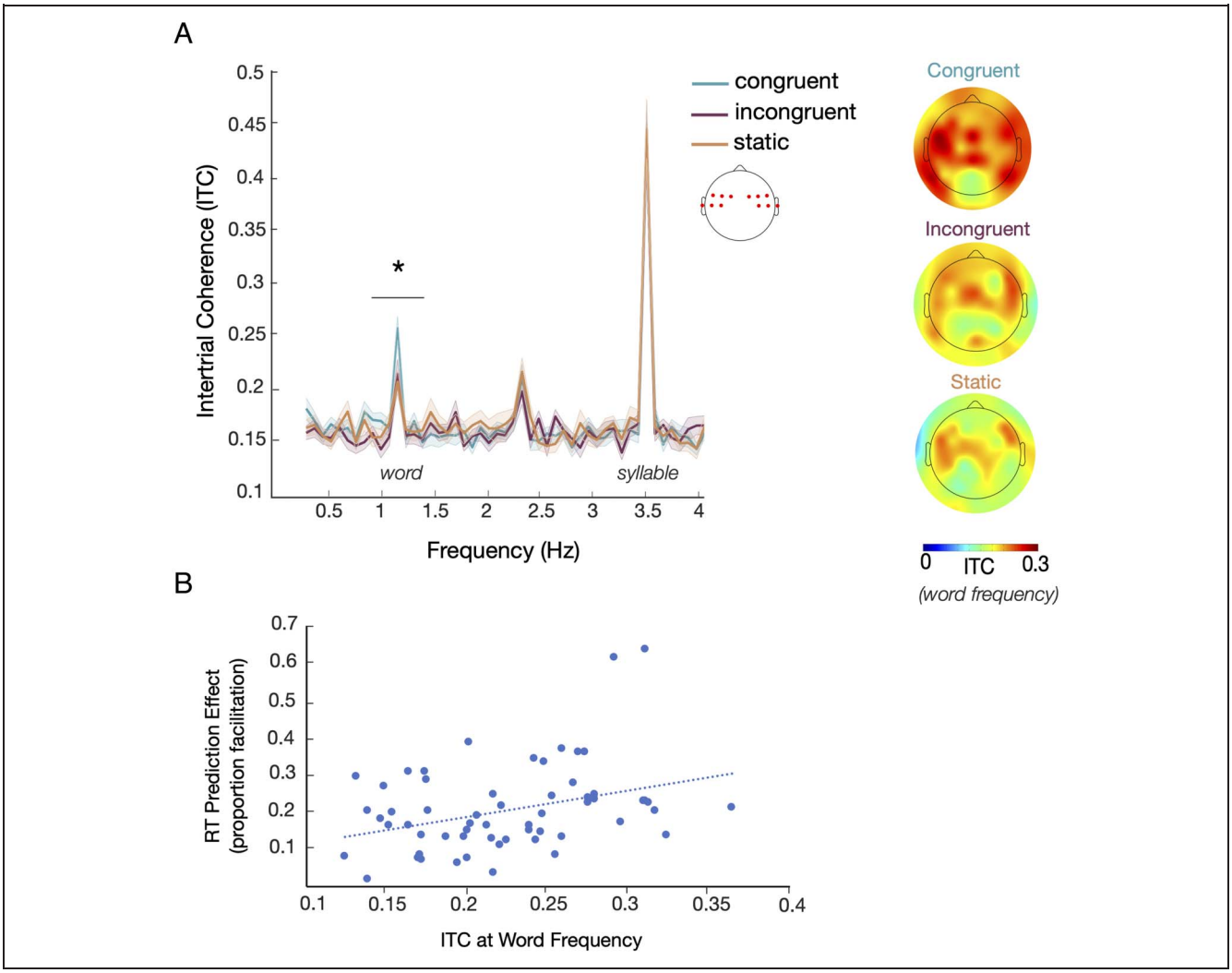

Figure 5. Neural entrainment during the target detection task. (A) EEG ITC values as a function of frequency and condition (congruent, incongruent, static). Data in the line graph are averaged across the same frontocentral region as for the exposure phase analysis, indicated by the red dots. Shaded regions represent the standard error of the mean within each group. Topographical plots show the distribution of ITC values across the scalp at the word frequency. Participants in the congruent condition showed a significant enhancement in neural entrainment at the word frequency. (B) A significant correlation was observed between neural entrainment at the word frequency during the target detection task and the RT prediction effect.

$\chi^2(1, N = 37) = 2.28, p = .13$; interestingly, a numerically higher proportion of participants from the incongruent group endorsed this statement. Participants who endorsed an association between the video and audio streams were then asked to describe this association. On this open-ended response, the proportion of participants who were coded as aware, partially aware, or unaware also did not differ by group, $\chi^2(2, N = 37) = 0.78, p = .78$. Overall, these results suggest that participants in the incongruent condition—who in reality had not been presented with words that aligned with the repeating video loop—were as likely as participants in the congruent condition to indicate that there was a relationship between the structure of the stream and the video. Although this measure is subjective as it relies on participants' verbal reports, these results suggest that very few (at most one) participants in the congruent group gained accurate, explicit knowledge of the speech structure-video contingency.

DISCUSSION

The goal of the current study was to better understand whether neural entrainment plays a functional role in statistical learning, or whether these recorded signals largely reflect the downstream consequences of learning. To shed light on this issue, we tested whether modulating neural entrainment during statistical learning impacts subsequent learning outcomes. While participants listened to a structured speech stream, we attempted to manipulate neural entrainment using continuous rhythmic visual stimulation. Our initial set of analyses served as a manipulation check, demonstrating that this cross-modal manipulation was successful. Relative to participants in incongruent

stimulation and control conditions, participants in the congruent group—who viewed a visual stimulus that was consistent with the embedded words in the speech stream—showed stronger neural entrainment at the word frequency within our auditory electrode ROI. Importantly, this entrainment response at the word frequency was significantly greater than the additive responses to structured speech and visual stimulation alone, reflecting a cross-modal integration response over and above effects of mere volume conduction. These results demonstrate that the visual rhythm successfully engaged auditory-relevant neural substrates involved in processing the speech stream, confirming that our visual manipulation successfully altered neural dynamics within relevant neural populations as intended. Posttask interviews revealed that only one participant in the congruent group correctly reported the temporal alignment between the visual rhythm and the hidden words in the speech steam, indicating that the increase in neural entrainment occurred largely outside of participants' awareness and strategic control, and thus reflects implicit cross-modal integration at the neural level.

After the exposure phase, participants' statistical learning was assessed using explicit and implicit tests of word knowledge. Critically, on the target detection task, we found that participants in the congruent condition showed a significantly stronger priming effect, responding more quickly to predictable (later positioned) syllables than participants in the other two conditions. Furthermore, participants in the congruent condition also showed stronger neural entrainment at the word frequency over our auditory ROI, which we interpret to reflect enhanced segmentation and/or perception of the word units during re-exposure to the same artificial language. These findings indicate that participants in the congruent group gained greater sensitivity to the hidden word structure of the speech stream, enabling them to better perceive embedded words and predict upcoming syllables. In addition, the neural entrainment response to words (during the implicit test) significantly correlated with participants' prediction effect at the behavioral level. In contrast, on the explicit word rating task, participants showed similar levels of learning, with no significant group effects. These results indicate that boosting neural entrainment at the relevant frequency during exposure to regularities facilitates statistical learning as measured implicitly, at both the neural and behavioral levels. Taken together, these findings suggest that neural entrainment is a facilitative mechanism and not simply an outcome of statistical learning.

Visual Stimulation Modulated Neural Entrainment within Targeted Auditory Regions

As expected, our rhythmic visual stimulus produced a maximal entrainment response over posterior-occipital scalp regions, consistent with the well-known general topography of visual EEG responses (e.g., Müller et al., 1998; Clark, Fan, & Hillyard, 1994). However, in the case of the congruent condition—in which the words in the speech stream aligned with the visual rhythmic stimulation—effects of the visual stimulation were observed well beyond posterior regions of the scalp, including within our more frontocentral auditory ROI. For several reason, we believe that this boost in neural entrainment at the word frequency in the congruent group over auditory electrodes reflects more than just passive volume conduction. First, we applied a Laplacian filter to separate auditory and visual sources. The topography observed in the congruent group (as well across the other two groups more generally) suggests that this transform effectively separated these two stimulus sources, revealing a clear frontocentral bilateral distribution for auditory stimuli and a posterior-occipital distribution for visual stimuli, as has been shown in previous studies using a Laplacian transform (Bauer et al., 2021; Jaeger et al., 2018; Kayser & Tenke, 2015). More importantly, participants in the congruent group showed a super-additive neural entrainment response at the word frequency over our auditory electrode ROI. That is, the neural entrainment response in the congruent group exceeded the estimated independent contributions from word processing and visual processing alone, as modeled from data in the incongruent and static groups. This result indicates that multisensory integration occurred between the auditory stream and visual rhythm at some level, producing an enhanced neural response that cannot be attributed merely to the independent sensory effects of each separate stimulus stream. These EEG results can be thought of as a manipulation check, showing that our visual manipulation influenced neural entrainment to the auditory stream as intended.

Previous studies of cross-modal interactions have shown that sensory input in one modality (e.g., vision) can influence neural activity in the sensory cortex of another modality (e.g., audition; Mégevand et al., 2020; Atilgan et al., 2018; Luo, Liu, & Poeppel, 2010; Kayser, Petkov, & Logothetis, 2008; see Bauer, Debener, & Nobre, 2020, for a review). For example, in the ferret, Atilgan and colleagues (2018) showed that visual stimuli shape how auditory cortical neurons respond to sound mixtures, enhancing the representation of the sound that is temporally coherent with the visual stimuli. The authors also demonstrated that visual information shifted the phase of ongoing oscillations in auditory cortex, supporting the role of neural oscillations in the integration and enhancement of coherent multisensory input. Similarly, in humans, the phase of low-frequency neural activity in auditory cortex tracks unisensory visual speech, as reflected by mouth movements (Mégevand et al., 2020). This mechanism may underlie the well-known benefit of visual speech cues on speech comprehension (Sumby & Pollack, 1954). Interestingly, temporally coherent visual input can benefit auditory processing even when the coherence occurs between stimulus features that are

completely task-irrelevant (Maddox, Atilgan, Bizley, & Lee, 2015). Extending these results, in the current study, the temporally coherent visual stimulus (in the congruent condition) appears to have directly influenced regions related to auditory processing, which we speculate may have occurred by phase-shifting or aligning ongoing oscillations with the trisyllabic word structure.

Participants in the Congruent Group Showed Enhanced Prediction and Word Entrainment on the Subsequent Target Detection Task

Our key behavioral finding was that participants in the congruent condition—who experienced a boost in neural entrainment at the word frequency during exposure—showed an enhanced RT prediction effect on the subsequent target detection task, responding more quickly to predictable, later positioned syllables compared with the other two groups. Participants in the congruent group also showed significantly greater neural entrainment at the word frequency during this same task. In addition, a significant correlation was found between participants' neural entrainment to words during the target detection task and their RT prediction effect, suggesting that participants whose perception of the stream was more biased toward the word-like units were also better able to predict upcoming syllables. Taken together, these results indicate that participants in the congruent group gained greater sensitivity to the hidden structure of the speech stream during exposure. Because of their superior statistical learning, they were then better able to segment the component words and to predict upcoming syllables during re-exposure to the speech stream (e.g., Ordin et al., 2020; Batterink & Paller, 2017; Buiatti et al., 2009).

These results converge with previous findings that grammatical entrainment to repeated English phrases facilitates statistical learning of compatible, aligned structures in an artificial language (Wang et al., 2017). Although Wang and colleagues relied only on behavioral methods, it is possible that their experimental manipulation may have had similar effects on neural entrainment, producing neurophysiological tracking of the syntactic structures (e.g., Ding et al., 2016, 2017) that persisted long enough to influence parsing of the subsequent artificial language structures. Taken together, these findings suggest that alignment of neural processes to relevant structures has measurable effects on statistical learning at the behavioral level, and provide support for our original hypothesis that enhancing neural entrainment at the frequency of the hidden word structure in the incoming auditory stream facilitates statistical learning.

As mentioned previously in the Introduction, the current results (as well as previous studies on neural entrainment during statistical learning) do not provide direct evidence of neural entrainment in the narrow sense, which invokes the presence of an endogenous neural oscillator that adjusts its frequency and/or phase to align with rhythmic sensory input (Obleser & Kayser, 2019). Disentangling true endogenous oscillatory activity from neural responses that are simply evoked by rhythmic input is an issue that is beyond the scope of the current study and will require additional studies incorporating careful and clever experimental design (Zoefel, ten Oever, et al., 2018). Although it is not yet known whether the neural mechanism(s) reflected by the observed entrainment signal in this kind of artificial word segmentation paradigm is inherently oscillatory in nature, we have shown that reinforcing this dynamic through task-free cross-modal stimulation impacts learning outcomes. These results provide insight into the nature of the measured neural entrainment response to words during statistical learning, suggesting that the neural process (or processes) indexed by this neural entrainment signal play a functional role in statistical learning.

Although we cannot make claims about the existence of endogenous oscillators based on the current data, our findings can be speculatively interpreted in the context of general neural entrainment models (Obleser & Kayser, 2019; Gross et al., 2013; Thut, Miniussi, & Gross, 2012; Schroeder & Lakatos, 2009). These models propose that the encoding of a given stimulus relies on a neural population with its own preferred, intrinsic firing rates. When confronted with a quasirhythmic stimulus, the neural population may slightly shift its firing rate, and/or the phase of its firing, such that an incoming stimulus will arrive during a phase of high excitability. In turn, this leads to amplification of the stimulus and more efficient processing. This general framework has been used to explain the benefit of multisensory cues on processing (Bauer et al., 2020; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007). For example, Schroeder, Lakatos, Kajikawa, Partan, and Puce (2008) have proposed that viewing a speaker's face improves speech intelligibility because the predictive visual input resets or shifts the phase of ongoing neural oscillations in auditory cortex so that linked auditory input arrives at the optimal oscillatory phase. By extension, in the congruent condition of the current study, the compatible visual rhythmic cue may have shifted the phase or firing rate of some proportion of the relevant neural population to align with the word structure of the speech stream. This in turn could align peaks of neural excitability with the hidden word structure, potentially facilitating processing at key moments of the signal (e.g., word onsets) and leading to enhanced perception of the component words. We reiterate that this explanation remains speculative and must be tested by future research that is specifically designed to disentangle contributions of endogenous oscillatory processes and from the reinforcement of discrete neural responses more generally.

Unexpected Findings

We expected (and intended) that the incongruent visual stimulation would interfere with neural entrainment to

the words, resulting in reduced entrainment at the word frequency relative to the static group. However, participants in the incongruent group showed very similar levels of neural entrainment as participants in the static group. Thus, although congruent visual signals enhanced neural entrainment to words, incongruent visual signals did not reduce entrainment. Perhaps unsurprisingly given the null effects observed at the neural level, we also found no behavioral differences between the incongruent and static groups on subsequent learning tests. Overall, these results suggest that participants may downweigh or discard competing sensory input that is uninformative or nonpredictive of statistical structure in the environment, instead prioritizing the signal that contains to-be-learned regularities. Our results converge with the findings of Cunillera, Càmara, Laine, and Rodríguez-Fornells (2010), who manipulated the presentation of line drawings relative to an ongoing continuous speech stream. Although synchronous, rhythmic visual input that temporally coincided with word boundaries improved statistical learning relative to an audio-only condition, arrhythmic or asynchronous visual information did not interfere with learning. This conclusion is also consistent with previous evidence that statistically structured streams of information receive attentional priority over random or noisy streams (Forest, Siegelman, & Finn, 2022; Yu & Zhao, 2015; Zhao, Al-Aidroos, & Turk-Browne, 2013). Taken together, these prior findings support our interpretation that learners simply discard visual information that is noninformative and serve to highlight the relative automaticity or obligatory nature of linguistic statistical learning, in line with previous results (Batterink & Paller, 2019; Fernandes, Kolinsky, & Ventura, 2010; Saffran et al., 1997).

Another unexpected finding was that congruent visual stimulation did not result in better performance on the familiarity rating task, our explicit measure of learning. Although participants in all three groups showed significant evidence of learning on this task, performance did not differ among the groups. Thus, boosting neural entrainment at the word frequency during exposure enhanced the subsequent implicit expression of statistical learning, but not explicit memory for the learned words. There are a few possible explanations for this finding. First, the target detection task has been previously found to be a more sensitive measure of learning compared with explicit learning measures; it reveals learning effects in a greater number of individuals compared with the classic explicit force-choice recognition task (Pinto et al., 2022; Batterink et al., 2015), is more sensitive to a between-participants attentional manipulation compared with the explicit rating task (Batterink & Paller, 2019), and correlates more robustly with neural entrainment compared with other behavioral measures (Pinto et al., 2022; Batterink & Paller, 2017, 2019). Therefore, one possibility is that the familiarity rating is simply insufficiently sensitive to capture group differences, at least with the current power. An alternative explanation concerns the nature of the learned representations that are expressed on the two tasks. As described in the Results section, the vast majority of participants were largely unaware of the temporal alignment between the visual rhythm and the embedded words, suggesting that the observed learning advantages occurred through neural processes outside of their conscious strategy and control. Thus, enhancing neural entrainment at the relevant word frequency may have more strongly facilitated moment-by-moment perception, segmentation, and prediction processes—as expressed most easily through online performance—rather than directly enhancing explicit memory for the learned words, as assessed on the rating task. Related to this point, on the rating task, a consistent finding is that the group average typically shows less than a 1-point difference between words and nonwords (where a difference of 3 would indicate optimal performance on the 4-point rating scale; Wang, Köhler, et al., 2023; Wang, Rosenbaum, et al., 2023; Moreau et al., 2022; Batterink & Paller, 2017, 2019). This moderate level of performance suggests that participants do not generally acquire strong or precise explicit memory representations of the component words. Given the general difficulty of this task, it may be less likely to capture the effects of subtle shifts in neural entrainment patterns during statistical learning. Future studies could consider incorporating additional measures of statistical learning—both implicit and explicit—to further test these ideas.

Finally, we did not find a significant correlation between neural entrainment during the exposure period and subsequent prediction effects on the target detection task, as we had initially expected based on our prior studies. This correlation was in the positive direction, as predicted, but failed to reach significance ($p = .10$). Although neural entrainment has often been found to correlate with subsequent performance on postlearning tests (Batterink, 2020; Choi et al., 2020; Batterink & Paller, 2017, 2019; Kabdebon et al., 2015; Buiatti et al., 2009), this is not always the case (Moreau et al., 2022; Smalle et al., 2022; Zhang et al., 2021; Ordin et al., 2020). Part of this inconsistency across studies may stem from the field's lack of a "ground truth" indication of statistical learning performance (Pinto et al., 2022); even at the behavioral level, different measures of statistical learning often do not correlate across participants and have been theorized to reflect different aspects of statistical learning (Isbilen, McCauley, Kidd, & Christiansen, 2020; Batterink et al., 2015; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015). In addition, in the current study, neural entrainment at the time of exposure may have been affected (or obscured) by individual differences in cross-modal integration of the visual stimulus and auditory stream. That is, independently of the statistical learning task, the magnitude of the super-additive effect of cross-modal integration may not have been consistent across participants, which would weaken the overall relation between neural entrainment at exposure and performance on later tasks of learning. However, as previously described, we did find a correlation between neural

entrainment to words during the target detection task itself and the behavioral prediction effect, an effect that was not tested by previous studies. This finding supports the general correspondence between neural entrainment to words and statistical learning effects as assessed at the behavioral level.

Limitations and Future Directions

Neural entrainment in the present study was manipulated cross-modally using a visual stimulus, rather than directly using brain stimulation methods (cf. Riecke et al., 2018; Wilsch et al., 2018), which limits the mechanistic insights that can be drawn. Specifically, an argument could be made that neural entrainment was enhanced because the congruent visual rhythm provided a cue to learn word boundaries, by visually reinforcing the statistical regularities. We consider this argument carefully below.

We introduced our visual stimulus to participants as a task-irrelevant nature video designed to encourage relaxation, with the intention of disguising the temporal alignment between the visual stimuli and speech stream structure. In addition, the visual stimulus was smoothly continuous, consisting of a repeating cycle in which each frame slightly differed from the last, rather than discrete pulses that would provide clear onset cues to word boundaries. We hoped that these features would disguise the temporal alignment between the visual stimuli and speech stream structure in the congruent condition. The results of the posttask interview suggest that this disguise was successful; as previously described, very few, if any, participants in the congruent condition reported awareness of the congruency between the speech structure and visual rhythm. The inference that participants in the congruent condition did not use the visual stimulus as an explicit cue to strategically discover word boundaries in the stream is further supported by the finding that they achieved similar performance on the rating task, our explicit measure of word knowledge. If participants in the congruent group had indeed strategically used the visual stimulus as an explicit cue for decoding words, we would expect them to correspondingly have higher levels of explicit word knowledge. Thus, the boost in neural entrainment at the word frequency cannot be readily attributed to conscious, strategic processes on the part of the participant, but instead reflects multisensory integration between the audio and visual stimuli, occurring outside of participants' awareness.

Nonetheless, the visual stimulus may still have provided an implicit cue for word boundaries, reinforcing the relevant word rhythm and facilitating learning outside of participants' conscious awareness. However, we would consider the idea of implicit cueing and increased neural entrainment at the word frequency to be mutually compatible with one another, representing two sides of the same coin. As an analogy, a well-documented finding is that speech signals are consistently more intelligible when the listener has access to both audio and visual input (e.g., the speaker's mouth; Van Engen, Dey, Sommers, & Peelle, 2022; Sommers, Tye-Murray, & Spehar, 2005; Arnold & Hill, 2001; Erber, 1975; Sumby & Pollack, 1954). In particular, the opening of the mouth is associated with louder amplitudes, which provides clues about the rhythmic structure of speech and helps listeners predict incoming information (Van Engen et al., 2022; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). This benefit of visual information on speech processing occurs implicitly, without requiring conscious judgments or decision making about the two sources of input. At the behavioral level, access to visual speech information may thus be thought of as providing an implicit, convergent cue to incoming speech sounds, facilitating speech comprehension. A proposed neural mechanism for this effect is neural entrainment: Visual information has been shown to result in enhanced entrainment of neural oscillatory activity in auditory cortex to the amplitude envelope of speech (Crosse, Butler, & Lalor, 2015; Zion Golumbic et al., 2013; Luo et al., 2010; Schroeder et al., 2008). In other words, visual information may serve as an implicit cue by enhancing neural activity in relevant networks, with these two concepts reflecting different levels of description (cognitive vs. neural). In the current study, we have shown that congruent visual stimulation results in both increased neural entrainment at the target (word) frequency as well as stronger performance on our implicit measure of statistical learning. We interpret this increase in neural entrainment as a potential mechanistic explanation for how the congruent visual stimulation drives the observed learning advantage.

Of course, as mentioned previously, our understanding of what precisely this increase in the neural entrainment signals reflect—and whether such effects stem from a true adjustment of endogenous oscillators in line with the repeating regularities—is currently limited. Thus, to call this "entrainment" in the narrower sense (Obleser & Kayser, 2019), future studies are needed to disentangle the potential contribution of stimulus-driven cues from the contributions of ongoing, endogenous oscillatory processes. One promising approach would involve manipulating neural entrainment directly using transcranial electrical stimulation, following previous studies in the speech comprehension literature (Riecke et al., 2018; Wilsch et al., 2018).

Conclusions

The current findings show that boosting neural entrainment at the same frequency as the underlying statistical regularities during exposure to structured input facilitates subsequent expression of statistical knowledge, suggesting that entrainment to rhythmic input plays a functional role in statistical learning. Although the speech stream in the current study was isochronous, neural entrainment does not depend on a stimulus being perfect rhythmic,

but has also been implicated in the processing of natural (quasirhythmic) speech, as we described earlier (e.g., Meyer, 2018; Gross et al., 2013; Giraud & Poeppel, 2012; Peelle & Davis, 2012). Entrainment at delta and theta frequencies is related to an individual's ability to understand speech (Park, Ince, Schyns, Thut, & Gross, 2015; Doelling, Arnal, Ghitza, & Poeppel, 2014; Peelle, Gross, & Davis, 2013) and is thought to support the segmentation of the continuous speech stream into timescales corresponding to linguistic units, including words and phrases (Meyer, 2018; Gross et al., 2013; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Ghitza, 2011; Poeppel, 2003).

Given this framework, our results may set the stage for future work aimed at revealing novel ways to boost statistical learning and related aspects of language acquisition. For example, to aid in the discovery of words in continuous speech, second language learners or children with language impairments may benefit from viewing predictive visual cues to word onsets while listening to target speech. Through cross-modal integration, this type of manipulation could allow neural activity in auditory-sensitive regions to optimally entrain to word onsets, facilitating language learning. Furthermore, neural entrainment has been found to be enhanced to song compared with speech (Vanden Bosch der Nederlanden, Joanisse, & Grahn, 2020), and time-compressed speech is understood more easily when silent gaps are inserted into the signal periodically (Ghitza & Greenberg, 2009). Thus, another possible route to facilitating speech segmentation of a target language would be by manipulating the speech signal itself, through alterations that increase the rhythmicity or predictability of word units. Exposing learners to an optimized, more rhythmic speech signal at early learning stages may help to scaffold language learning by enhancing neural entrainment, similar to known benefits of infant-directed speech (Nencheva & Lew-Williams, 2022).

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Corresponding author: Laura J. Batterink, Department of Psychology, Western Institute for Neuroscience, Western University, 1151 Richmond Street, London, ON, Canada, or via e-mail: lbatter@uwo.ca.

Data Availability Statement

Study data and materials are available at the following link: https://osf.io/xk87m/.

Author Contributions

Laura J. Batterink: Conceptualization; Formal analysis; Funding acquisition; Methodology; Supervision; Visualization; Writing—Original draft. Jerrica Mulgrew: Formal analysis; Investigation; Methodology; Visualization; Writing—Review & editing. Aaron Gibbings: Writing—Review & editing.

Funding Information

Natural Sciences and Engineering Research Council of Canada (https://dx.doi.org/10.13039/501100000038), grant number RGPIN-2019-05132 to Laura Batterink.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 13367–13372. https://doi.org/10.1073/pnas.201400998, PubMed: 11698688

Arciuli, J., & Torkildsen, J. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Psychology*, *3*, 324. https://doi.org/10.3389/fpsyg.2012.00324, PubMed: 22969746

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*, 339–355. https://doi.org/10.1348/000712601162220, PubMed: 11417785

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*, e1373. https://doi.org/10.1002/wcs.1373, PubMed: 27906526

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., et al. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron*, *97*, 640–655. https://doi.org/10.1016/j.neuron.2017.12.034, PubMed: 29395914

Bánki, A., Brzozowska, A., Hoehl, S., & Köster, M. (2022). Neural entrainment vs. stimulus-tracking: A conceptual challenge for rhythmic perceptual stimulation in developmental neuroscience. *Frontiers in Psychology*, *13*, 878984. https://doi.org/10.3389/fpsyg.2022.878984, PubMed: 35602682

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological Science*, *28*, 921–928. https://doi.org/10.1177/0956797617698226, PubMed: 28493810

Batterink, L. J. (2020). Syllables in sync form a link: Neural phase-locking reflects word knowledge during language learning. *Journal of Cognitive Neuroscience*, *32*, 1735–1748. https://doi.org/10.1162/jocn_a_01581, PubMed: 32427066

Batterink, L. J., & Choi, D. (2021). Optimizing steady-state responses to index statistical learning: Response to Benjamin and colleagues. *Cortex*, *142*, 379–388. https://doi.org/10.1016/j.cortex.2021.06.008, PubMed: 34321154

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004, PubMed: 28324696

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013, PubMed: 30771622

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*, 482–503. https://doi.org/10.1111/tops.12420, PubMed: 30942536

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004, PubMed: 26034344

Batterink, L. J., & Zhang, S. (2022). Simple statistical regularities presented during sleep are detected but not retained. *Neuropsychologia*, *164*, 108106. https://doi.org/10.1016/j.neuropsychologia.2021.108106, PubMed: 34864052

Bauer, A.-K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of neural oscillations and cross-modal influences. *Trends in Cognitive Sciences*, *24*, 481–495. https://doi.org/10.1016/j.tics.2020.03.003, PubMed: 32317142

Bauer, A.-K. R., van Ede, F., Quinn, A. J., & Nobre, A. C. (2021). Rhythmic modulation of visual perception by continuous rhythmic auditory stimulation. *Journal of Neuroscience*, *41*, 7065–7075. https://doi.org/10.1523/JNEUROSCI.2980-20.2021, PubMed: 34261698

Benjamin, L., Dehaene-Lambertz, G., & Fló, A. (2021). Remarks on the analysis of steady-state responses: Spurious artifacts introduced by overlapping epochs. *Cortex*, *142*, 370–378. https://doi.org/10.1016/j.cortex.2021.05.023, PubMed: 34311971

Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., & Dehaene-Lambertz, G. (2023). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. *Developmental Science*, *26*, e13300. https://doi.org/10.1111/desc.13300, PubMed: 35772033

Boros, M., Magyari, L., Török, D., Bozsik, A., Deme, A., & Andics, A. (2021). Neural processes underlying statistical learning for speech segmentation in dogs. *Current Biology*, *31*, 5512–5521. https://doi.org/10.1016/j.cub.2021.10.017, PubMed: 34717832

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. https://doi.org/10.1016/j.neuroimage.2008.09.015, PubMed: 18929668

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., & Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One*, *6*, e14543. https://doi.org/10.1371/journal.pone.0014543, PubMed: 21267081

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*, e1000436. https://doi.org/10.1371/journal.pcbi.1000436, PubMed: 19609344

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment. *Psychological Science*, *31*, 1161–1173. https://doi.org/10.1177/0956797620933237, PubMed: 32865487

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*, 170–187. https://doi.org/10.1002/hbm.460020306

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9609.001.0001

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24–39. https://doi.org/10.1037/0278-7393.31.1.24, PubMed: 15641902

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, *35*, 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015, PubMed: 26490860

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology*, *63*, 260–274. https://doi.org/10.1080/17470210902888809, PubMed: 19526435

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., et al. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*, *48*, 541–553. https://doi.org/10.1016/j.neuroimage.2009.06.069, PubMed: 19580874

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, *1123*, 168–178. https://doi.org/10.1016/j.brainres.2006.09.046, PubMed: 17064672

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One*, *2*, e1175. https://doi.org/10.1371/journal.pone.0001175, PubMed: 18000546

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. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

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. https://doi.org/10.3389/fnhum.2017.00481, PubMed: 29033809

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. https://doi.org/10.1038/nn.4186, PubMed: 26642090

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035, PubMed: 23791839

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences, U.S.A.*, *116*, 10113–10121. https://doi.org/10.1073/pnas.1816414116, PubMed: 31019082

Elmer, S., Valizadeh, S. A., Cunillera, T., & Rodriguez-Fornells, A. (2021). Statistical learning and prosodic bootstrapping differentially affect neural synchronization during speech segmentation. *Neuroimage*, *235*, 118051. https://doi.org/10.1016/j.neuroimage.2021.118051, PubMed: 33848624

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders*, *40*, 481–492. https://doi.org/10.1044/jshd.4004.481, PubMed: 1234963
Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science*, *18*, 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x, PubMed: 17444923
Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, *72*, 1522–1532. https://doi.org/10.3758/APP.72.6.1522, PubMed: 20675798
Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504. https://doi.org/10.1111/1467-9280.00392, PubMed: 11760138
Fló, A., Benjamin, L., Palu, M., & Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. *Scientific Reports*, *12*, 4391. https://doi.org/10.1038/s41598-022-08411-w, PubMed: 35292694
Forest, T. A., Siegelman, N., & Finn, A. S. (2022). Attention shifts to more complex structures with experience. *Psychological Science*, *33*, 2059–2072. https://doi.org/10.1177/09567976221114055, PubMed: 36219721
Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid serial auditory presentation: A new measure of statistical learning in speech segmentation. *Experimental Psychology*, *62*, 346–351. https://doi.org/10.1027/1618-3169/a000295, PubMed: 26592534
Getz, H., Ding, N., Newport, E. L., & Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition*, *181*, 135–140. https://doi.org/10.1016/j.cognition.2018.08.019, PubMed: 30195135
Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. https://doi.org/10.3389/fpsyg.2011.00130, PubMed: 21743809
Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*, 113–126. https://doi.org/10.1159/000208934, PubMed: 19390234
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517. https://doi.org/10.1038/nn.3063, PubMed: 22426255
Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436. https://doi.org/10.1111/1467-9280.00476, PubMed: 12219809
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. https://doi.org/10.1371/journal.pbio.1001752, PubMed: 24391472
Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3, PubMed: 11124355
Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., et al. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*, eabc4530. https://doi.org/10.1126/sciadv.abc4530, PubMed: 33608265
Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis into 25 years of research. *Cognitive Science*, *46*, e13198. https://doi.org/10.1111/cogs.13198, PubMed: 36121309
Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*, e12848. https://doi.org/10.1111/cogs.12848, PubMed: 32608077
Jaeger, M., Bleichner, M. G., Bauer, A.-K. R., Mirkovic, B., & Debener, S. (2018). Did you listen to the beat? Auditory steady-state responses in the human electroencephalogram at 4 and 7 Hz modulation rates reflect selective attention. *Brain Topography*, *31*, 811–826. https://doi.org/10.1007/s10548-018-0637-8, PubMed: 29488040
Jin, P., Lu, Y., & Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *eLife*, *9*, e55613. https://doi.org/10.7554/eLife.55613, PubMed: 32310082
Kabdebon, C., Pena, M., Buiatti, M., & Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. *Brain and Language*, *148*, 25–36. https://doi.org/10.1016/j.bandl.2015.03.005, PubMed: 25865749
Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, *18*, 1560–1574. https://doi.org/10.1093/cercor/bhm187, PubMed: 18180245
Kayser, J., & Tenke, C. E. (2015). Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. *International Journal of Psychophysiology*, *97*, 189–209. https://doi.org/10.1016/j.ijpsycho.2015.04.012, PubMed: 25920962
Keitel, C., Quigley, C., & Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: Entrainment of intrinsic rhythms or frequency-following response? *Journal of Neuroscience*, *34*, 10137–10140. https://doi.org/10.1523/JNEUROSCI.1904-14.2014, PubMed: 25080577
Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*, 2867–2875. https://doi.org/10.1016/j.cub.2018.07.023, PubMed: 30197083
Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*, 279–292. https://doi.org/10.1016/j.neuron.2006.12.011, PubMed: 17224408
Liu, H., Forest, T. A., Duncan, K., & Finn, A. S. (2023). What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition*, *236*, 105439. https://doi.org/10.1016/j.cognition.2023.105439, PubMed: 36934685
Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213, PubMed: 24782741
Lu, L., Sheng, J., Liu, Z., & Gao, J.-H. (2021). Neural representations of imagined speech revealed by frequency-tagged magnetoencephalography responses. *Neuroimage*, *229*, 117724. https://doi.org/10.1016/j.neuroimage.2021.117724, PubMed: 33421593
Luo, C., & Ding, N. (2020). Cortical encoding of acoustic and linguistic rhythms in spoken narratives. *eLife*, *9*, e60433. https://doi.org/10.7554/eLife.60433, PubMed: 33345775
Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*, e1000445. https://doi.org/10.1371/journal.pbio.1000445, PubMed: 20711473
Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. C. (2015). Auditory selective attention is enhanced by a task-irrelevant

temporally coherent visual stimulus in human listeners. *eLife*, *4*, e04995. https://doi.org/10.7554/eLife.04995, PubMed: 25654748
Mégevand, P., Mercier, M. R., Groppe, D. M., Zion Golumbic, E., Mesgarani, N., Beauchamp, M. S., et al. (2020). Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *Journal of Neuroscience*, *40*, 8530–8542. https://doi.org/10.1523/JNEUROSCI.0555-20.2020, PubMed: 33023923
Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*, 2609–2621. https://doi.org/10.1111/ejn.13748, PubMed: 29055058
Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience*, *57*, 101154. https://doi.org/10.1016/j.dcn.2022.101154, PubMed: 36155415
Moser, J., Batterink, L., Li Hegner, Y., Schleger, F., Braun, C., Paller, K. A., et al. (2021). Dynamics of nonlinguistic statistical learning: From neural entrainment to the emergence of explicit knowledge. *Neuroimage*, *240*, 118378. https://doi.org/10.1016/j.neuroimage.2021.118378, PubMed: 34246769
Müller, M. M., Picton, T. W., Valdes-Sosa, P., Riera, J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (1998). Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range. *Cognitive Brain Research*, *6*, 249–261. https://doi.org/10.1016/S0926-6410(97)00036-0, PubMed: 9593922
Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047. https://doi.org/10.1016/j.dr.2022.101047
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*, 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2, PubMed: 14732409
Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*, 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011, PubMed: 21753000
Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*, 913–926. https://doi.org/10.1016/j.tics.2019.08.004, PubMed: 31606386
Ordin, M., Polyanskaya, L., Soto, D., & Molinaro, N. (2020). Electrophysiology of statistical learning: Exploring the online learning process and offline learning product. *European Journal of Neuroscience*, *51*, 2008–2022. https://doi.org/10.1111/ejn.14657, PubMed: 31872926
Palmer, S. D., Hutson, J., & Mattys, S. L. (2018). Statistical learning for speech segmentation: Age-related changes and underlying mechanisms. *Psychology and Aging*, *33*, 1035–1044. https://doi.org/10.1037/pag0000292, PubMed: 30247045
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. https://doi.org/10.1016/j.cub.2015.04.049, PubMed: 26028433
Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320, PubMed: 22973251
Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*, 1378–1387. https://doi.org/10.1093/cercor/bhs118, PubMed: 22610394
Pinto, D., Prior, A., & Zion Golumbic, E. (2022). Assessing the sensitivity of EEG-based frequency-tagging as a metric for statistical learning. *Neurobiology of Language*, *3*, 214–234. https://doi.org/10.1162/nol_a_00061, PubMed: 37215560
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time.' *Speech Communication*, *41*, 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3
Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*, 161–169. https://doi.org/10.1016/j.cub.2017.11.033, PubMed: 29290557
Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 906–914. https://doi.org/10.1002/wcs.78, PubMed: 21666883
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926, PubMed: 8943209
Saffran, J. R., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*, 479–500. https://doi.org/10.1016/j.cognition.2007.10.010, PubMed: 18082676
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4, PubMed: 10193055
Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805, PubMed: 28793812
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. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x
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. https://doi.org/10.1038/nn873, PubMed: 12068301
Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*, 9–18. https://doi.org/10.1016/j.tins.2008.09.012, PubMed: 19012975
Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*, 106–113. https://doi.org/10.1016/j.tics.2008.01.002, PubMed: 18280772
Sherman, B. E., Aljishi, A., Graves, K. N., Quraishi, I. H., Sivaraju, A., Damisah, E. C., et al. (2023). Intracranial entrainment reveals statistical learning across levels of abstraction. *Journal of Cognitive Neuroscience*, *35*, 1312–1328. https://doi.org/10.1162/jocn_a_02012, PubMed: 37262357
Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*, e12365. https://doi.org/10.1111/lnc3.12365
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(Suppl. 3)*, 692–727. https://doi.org/10.1111/cogs.12556, PubMed: 28986971
Smalle, E. H. M., Daikoku, T., Szmalec, A., Duyck, W., & Möttönen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. *Proceedings of the National*

Academy of Sciences, U.S.A., *119*, e2026011119. https://doi.org/10.1073/pnas.2026011119, PubMed: 34983868

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*, 263–275. https://doi.org/10.1097/00003446-200506000-00003, PubMed: 15937408

Stanford, T., & Stein, B. E. (2007). Superadditivity in multisensory integration: Putting the computation in context. *NeuroReport*, *18*, 787–792. https://doi.org/10.1097/WNR.0b013e3280c1e315, PubMed: 17471067

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215. https://doi.org/10.1121/1.1907309

Thiessen, E. D., Girard, S., & Erickson, L. C. (2016). Statistical learning and the critical period: How a continuous learning mechanism can give rise to discontinuous learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*, 276–288. https://doi.org/10.1002/wcs.1394, PubMed: 27239798

Thut, G., Miniussi, C., & Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Current Biology*, *22*, R658–R663. https://doi.org/10.1016/j.cub.2012.06.061, PubMed: 22917517

van Bree, S., Sohoglu, E., Davis, M. H., & Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biology*, *19*, e3001142. https://doi.org/10.1371/journal.pbio.3001142, PubMed: 33635855

Vanden Bosch der Nederlanden, C. M., Joanisse, M. F., & Grahn, J. A. (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *Neuroimage*, *214*, 116767. https://doi.org/10.1016/j.neuroimage.2020.116767, PubMed: 32217165

Vanden Bosch der Nederlanden, C. M., Joanisse, M. F., Grahn, J. A., Snijders, T. M., & Schoffelen, J.-M. (2022). Familiarity modulates neural tracking of sung and spoken utterances. *Neuroimage*, *252*, 119049. https://doi.org/10.1016/j.neuroimage.2022.119049, PubMed: 35248707

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. *Journal of the Acoustical Society of America*, *152*, 3216–3225. https://doi.org/10.1121/10.0015262, PubMed: 36586857

Wang, F. H., Zevin, J. D., & Mintz, T. H. (2017). Top–down structure influences learning of nonadjacent dependencies in an artificial language. *Journal of Experimental Psychology: General*, *146*, 1738–1748. https://doi.org/10.1037/xge0000384, PubMed: 29251987

Wang, H. S., Köhler, S., & Batterink, L. J. (2023). Separate but not independent: Behavioral pattern separation and statistical learning are differentially affected by aging. *Cognition*, *239*, 105564. https://doi.org/10.1016/j.cognition.2023.105564, PubMed: 37467624

Wang, H. S., Rosenbaum, R. S., Baker, S., Lauzon, C., Batterink, L. J., & Köhler, S. (2023). Dentate gyrus integrity is necessary for behavioral pattern separation but not statistical learning. *Journal of Cognitive Neuroscience*, *35*, 900–917. https://doi.org/10.1162/jocn_a_01981, PubMed: 36877071

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *Neuroimage*, *172*, 766–774. https://doi.org/10.1016/j.neuroimage.2018.01.038, PubMed: 29355765

Yu, R. Q., & Zhao, J. (2015). The persistence of the attentional bias to regularities in a changing environment. *Attention, Perception, & Psychophysics*, *77*, 2217–2228. https://doi.org/10.3758/s13414-015-0930-5, PubMed: 26037211

Zhang, M., Riecke, L., & Bonte, M. (2021). Neurophysiological tracking of speech-structure learning in typical and dyslexic readers. *Neuropsychologia*, *158*, 107889. https://doi.org/10.1016/j.neuropsychologia.2021.107889, PubMed: 33991561

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*, 667–677. https://doi.org/10.1177/0956797612460407, PubMed: 23558552

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. https://doi.org/10.1016/j.neuron.2012.12.037, PubMed: 23473326

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. https://doi.org/10.1016/j.cub.2017.11.071, PubMed: 29358073

Zoefel, B., ten Oever, S., & Sack, A. T. (2018). The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Frontiers in Neuroscience*, *12*, 95. https://doi.org/10.3389/fnins.2018.00095, PubMed: 29563860