# Sentential Contextual Facilitation of Auditory Word Processing Builds Up during Sentence Tracking

**Min Wu[1], Hans Rutger Bosker[2,3], and Lars Riecke[1]**

## Abstract

■ While listening to meaningful speech, auditory input is processed more rapidly near the end (vs. beginning) of sentences. Although several studies have shown such word-to-word changes in auditory input processing, it is still unclear from which processing level these word-to-word dynamics originate. We investigated whether predictions derived from sentential context can result in auditory word-processing dynamics during sentence tracking. We presented healthy human participants with auditory stimuli consisting of word sequences, arranged into either predictable (coherent sentences) or less predictable (unstructured, random word sequences) 42-Hz amplitude-modulated speech, and a continuous 25-Hz amplitude-modulated distractor tone. We recorded RTs and frequency-tagged neuroelectric responses (auditory steady-state responses) to individual words at multiple temporal positions within the sentences, and quantified sentential context effects at each position while controlling for individual word characteristics (i.e., phonetics, frequency, and familiarity). We found that sentential context increasingly facilitates auditory word processing as evidenced by accelerated RTs and increased auditory steady-state responses to later-occurring words within sentences. These purely top–down contextually driven auditory word-processing dynamics occurred only when listeners focused their attention on the speech and did not transfer to the auditory processing of the concurrent distractor tone. These findings indicate that auditory word-processing dynamics during sentence tracking can originate from sentential predictions. The predictions depend on the listeners' attention to the speech, and affect only the processing of the parsed speech, not that of concurrently presented auditory streams. ■

## INTRODUCTION

Speech comprehension is crucial for social communication in everyday life. It is thought to emerge from multiple auditory and linguistic processes including the identification of individual words in auditory input and the integration of words into syntactic and semantic structure (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Friederici, 2002; MacDonald, Pearlmutter, & Seidenberg, 1994). Besides such contributions from auditory word processing to the formation of higher-level syntactic and semantic information, a contribution in the opposite direction—that is, a top–down effect of syntactic and semantic processing on lower-level auditory word processing—also exists (Kaufeld et al., 2020; Guediche, Reilly, Santiago, Laurent, & Blumstein, 2016; Friederici, 2012; Davis, Ford, Kherif, & Johnsrude, 2011; Vigliocco et al., 2007). A striking illustration is temporal position effects in unambiguous speech, where auditory processing evolves incrementally as the meaning of the ongoing speech unfolds. For example, detection of brief acoustic events (e.g., clicks) has been shown to be quicker when these events occur during words late (vs. early) in a sentence (Lobina, Demestre, & Garcia-Albea, 2018; Holmes & Forster, 1970). This facilitation of auditory processing across word positions has also been observed in linguistic tasks requiring detection of target words in syntactically and/or semantically intact sentences (Oliver, Gullberg, Hellwig, Mitterer, & Indefrey, 2012; Marslen-Wilson & Tyler, 1975). These behavioral benefits may be enabled by corresponding across-word increases in neural oscillations in the gamma and theta bands, as suggested by neurophysiological studies on written word processing (Fedorenko et al., 2016; Lam, Schoffelen, Udden, Hulten, & Hagoort, 2016). In summary, the processing of acoustic events and words accelerates in a word-by-word manner during sentence processing, here referred to as "word-position effect."

A common interpretation of the word-position effect on auditory processing is that it originates from top–down processes involving semantic and syntactic predictions. The grouping of words into phrases can create semantic context from which the meaning of upcoming words can be predicted (Kuperberg & Jaeger, 2016; Staub, 2015). This semantic prediction has been demonstrated in many neurophysiological studies showing an enlarged N400 scalp potential evoked by unexpected words compared with expected words (Grisoni, Tomasello, & Pulvermüller, 2021; Kutas & Federmeier, 2011). Apart from semantic prediction, listeners may also employ syntactic prediction to anticipate the syntactic properties of upcoming words in grammatical sentences (Ferreira & Qiu, 2021; Demberg,

[1]Maastricht University, The Netherlands, [2]Max Planck Institute for Psycholinguistics, The Netherlands, [3]Radboud University, Nijmegen, The Netherlands

*Journal of Cognitive Neuroscience 35:8, pp. 1262–1278*
*https://doi.org/10.1162/jocn_a_02007*

Keller, & Koller, 2013). Based on the anticipation of meaning of upcoming words, the processing of later-occurring words in a sentence may be facilitated. A potential factor for this facilitation may be top–down attention: Even shorter RTs to later words have been observed when listeners are semantically primed to focus their attention on the later-occurring words in sentences (Cutler & Fodor, 1979). However, the precise role of attention in the word-position effect remains to be determined.

Another interpretation of the word-position effect puts more emphasis on auditory predictions derived from the overt prosody carried by the acoustic signal (Foltz, 2021; Ito & Speer, 2008). This view is supported by behavioral findings demonstrating the word-position effect within pseudosentences that contain no valid syntax or meaning but normal intonation (Tyler & Warren, 1987). However, interpretation of these previous observations in terms of top–down predictions is hampered by the fact that words at different positions differed not only in their lexical meaning and/or intonation but also in other individual word characteristics such as their phonetic detail, frequency of occurrence, and familiarity. Thus, it is still unclear from which processing level these dynamics in auditory processing originate. According to the current body of evidence, they may be driven by variations in word meaning, word phonetics, word frequency, word familiarity, or any of their combinations.

The present study aimed to disambiguate these possibilities. To this end, it tested the hypothesis that the dynamics in auditory processing during sentence processing can be driven exclusively by top–down factors, in particular sentential predictions derived from the speech.[1] Under our hypothesis, the word-position effect should (i) be observable during sentence processing even when individual word characteristics are held constant, (ii) depend on the listener's focus of endogenous top–down attention, and (iii) affect selectively the processing of the speech stream; that is, it should not transfer to the processing of a concurrent auditory stream.

We presented participants with either semantically coherent sentences or acoustically matched word sequences. To evaluate participants' ability to track sentence structure and to rule out potentially confounding prosodic cues, we used artificial isochronous speech (Ding et al., 2016). We recorded responses to individual words at multiple positions within the sentences. From these responses within sentential context, we subtracted responses to the acoustically identical words in randomly ordered word sequences to measure *sentential contextual facilitation* (referred to as SCF) at each position. To test our second and third predictions, we directed participants' attention to either the speech or a simultaneously ongoing distractor tone and simultaneously assessed cortical responses to speech or tone using a frequency-tagging technique. Previous neural studies of auditory speech processing have used frequency tagging to, for example, disentangle unisensory brain responses to simultaneously presented audiovisual speech (Drijvers, Jensen, & Spaak, 2021; Giani et al., 2012) or assess the internal integration of speech units (Ding et al., 2018; Ding et al., 2016) and implicit statistical learning (Pinto, Prior, & Zion Golumbic, 2022; Zhang, Riecke, & Bonte, 2021; Batterink & Paller, 2017). Frequency tagging involves the periodic modulation of a given stimulus feature at a specific frequency ("tagging frequency"), which results in brain activity that is phase-locked to the modulation and observable in neural recordings as a spectral peak at the tagging frequency.

We predicted that (i) SCF varies systematically across word positions, and that this putative word-position effect is modulated by (ii) the variation in the participants' focus of attention, and (iii) does not affect the processing of the simultaneous distractor tone.

## METHODS

### Participants

Twenty-eight fluent speakers of English (20 women, mean age = 23.8 years, age range = 19–44 years) participated in the experiment. The number of participants was chosen based on previous studies (Ding et al., 2018; Oliver et al., 2012). To ensure that all participants were fluent speakers of English, we included only native English speakers ($n$ = 3) and Maastricht University students enrolled in programs for which formal proof of English proficiency (Common European Framework of Reference level C1 or a similar accredited certification) is an admission requirement. All participants reported normal hearing and no history of neurological or psychiatric disorders. Written informed consent was obtained before the experiment. Participants were compensated with study credits or monetary reward for their participation. The experimental procedure was approved by the local research ethics committee.

### Stimuli

*Speech Stimuli*

Auditory word processing was elicited by presenting auditory stimuli consisting of 86 unique semantically valid four-word English sentences. Half of the sentences were composed of the syntactic structure [noun—verb—adjective—noun] (e.g., *Cats eat fresh fish*) whereas the other half had the syntactic structure [adjective—adjective—noun—verb] (e.g., *Two big dogs bark*); these two syntactic structures are respectively referred to as NVAN and AANV for simplicity. Although both structures contained a noun phrase and a verb phrase, the two structures differed strongly in the order and complexity of these phrases and words of the same category (noun, adjective, or verb) never occurred at the same position. Each trial consisted of eight different consecutive sentences with a fixed, predictable syntactic structure. The first sentence was the same in all trials of a given syntactic structure and served as a preparation of the participants for the upcoming speech syntax (this preparatory sentence was excluded from data analysis). The remaining 84 unique sentences (86 minus the two syntax-specific preparatory sentences) were distributed randomly across 12 consecutive trials. This set of trials was presented six times within each condition, each time in a different random order. No word was immediately repeated. All words were monosyllabic and synthesized independently using text-to-speech software in MacOS (male voice USA Alex) to avoid systematic overt prosody differences across words. Every word spanned an interval of 360 msec; this was achieved by applying time compression with a stretch factor ranging from 0.49 to 0.92 to the original sound waveform (speech acoustics are presented in the Appendix). Values of stretch factor below 1 represent compression; thus, only compression, but no expansion, was applied to the words. The speech stimuli were presented monaurally to participants' right ear (Figure 1A).

### *Acoustic Control Stimuli*

To allow ruling out confounding effects of word phonetics, frequency, or familiarity, appropriate control speech stimuli were included. These stimuli contained semantically anomalous versions of the 86 sentences and were constructed by shuffling all words across sentences (separately for each syntactic structure) except for target words (for details see section *Tasks*), which were always kept at their original positions (Figure 1B). The resulting random word sequences (or "unstructured speech") contained less predictable syntactic or semantic structure, and they were verified that no structured sentence was accidentally created. Analogous to the structured speech stimuli, the first four words were the same in all trials and no word was immediately repeated.

### *Distractor Tone*

To allow investigating effects of endogenous selective attention on speech processing (for details see below, section Tasks) and a word-position effect on a concurrent auditory stream, an auditory distractor stimulus was presented simultaneously with the speech stimuli. The distractor was an ongoing 250-Hz pure tone presented monaurally to participants' left ear, contralateral to the speech. The choice for a tonal distractor was motivated by previous studies showing sentential context effects on the processing of nonlinguistic auditory stimuli such as clicks (reviewed in Introduction section) and environmental sounds (Uddin, Heald, Van Hedger, Klos, & Nusbaum, 2018).

### *Frequency Tagging*

To allow assessing separately the auditory processing of the speech and the distractor, a frequency-tagging technique was used. Amplitude modulations (sinusoidal waveform, modulation depth: 100%, fixed phase at the start of each word interval) with different rates were applied to the speech stimuli ($41 + \frac{2}{3}$ Hz, hereafter called 42 Hz for simplicity) and the distractor (25 Hz). These modulation rates have been shown to evoke strong phase-locked auditory cortical responses (Gransier, van Wieringen, & Wouters, 2017; Schoonhoven, Boden, Verbunt, & De Munck, 2003; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000). As the eight four-word sentences within each trial were presented continuously at a fixed rate of $(4 \times 360\ \text{msec})^{-1}$ (i.e., 0.69 Hz), neural responses at this rate tagged participants' detection of the sentence structure.

## Tasks

To allow assessing effects of endogenous selective attention, participants' attention was experimentally drawn toward the speech stimuli, or away from them, using two detection tasks with acoustically matched stimulation (Figure 1A). In the word-detection task (hereafter called "speech task"), the target was a specific word embedded in the continuous auditory speech. To obtain a measure of auditory word processing at the linguistic level, the target word was defined to participants through the visual sensory modality (Figure 1C), which we deemed to encourage linguistic processing in the auditory speech task. In the tone-loudness change detection task (hereafter called "distractor task"), the target was a temporary loudness decrease (−5.1 dB) in the ongoing tone (target duration: 360 msec including 10-msec on/off ramps, corresponding to a single word interval).

In each task, participants were visually instructed to pay attention to the relevant auditory stream (speech or distractor tone), ignore the other stream, and detect and report the target in the attended stream as quickly as possible by pressing a key with the index finger of their right hand. On the basis of previous related studies (Jin, Zou, Zhou, & Ding, 2018; Oliver et al., 2012), only keypresses that fell into a 1440-msec window (corresponding to four consecutive word intervals) starting from the onset of the target were analyzed. Limiting the data analysis to responses within shorter windows (e.g., 720 msec, corresponding to two word intervals) did not alter the results qualitatively. In each task, targets occurred between 1 and 3 times per trial with approximately equal probability at each word position. Targets never occurred within the first sentence of a trial. As mentioned above, that sentence was presented only to inform participants of the upcoming speech syntax and never analyzed. Targets from different tasks (i.e., a given target word and a tone-loudness decrease) never concurred during the same interval.

## Procedure

The experimental procedure involved the following steps: First, before the experimental session, participants were

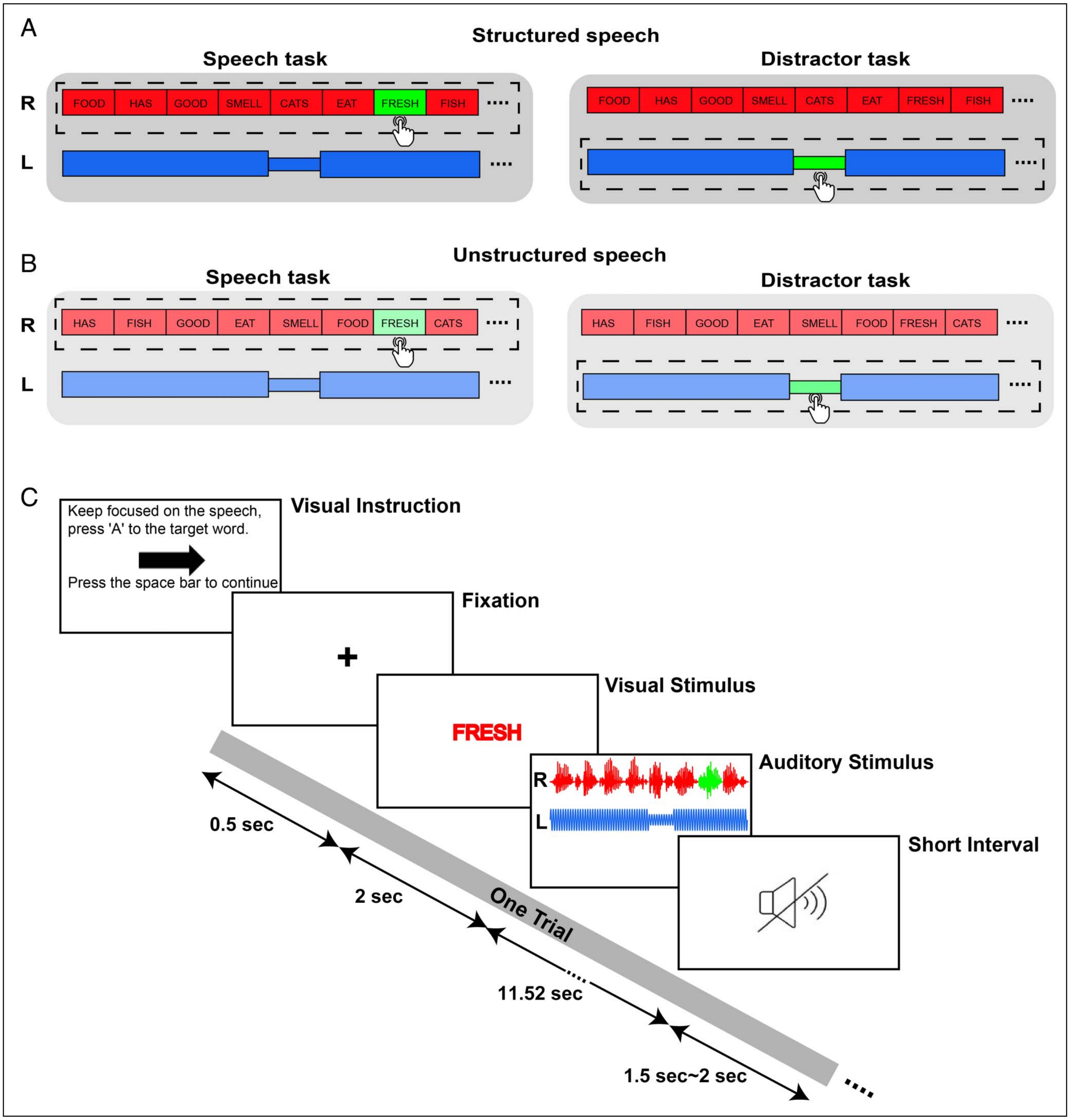

**Figure 1.** Auditory stimuli, experimental design, and trial design. (A) Illustration of the two main experimental conditions (gray boxes) involving structured speech. Sequences of red rectangles represent the speech stream (in the right ear, R), and sequences of blue rectangles represent the simultaneously presented distractor tone (in the left ear, L). Targets are depicted in green. The left shows conditions involving the speech task; the right illustrates the distractor task in which acoustically matched stimuli were presented. Large dashed contours indicate participants' focus of selective attention. (B) Same as (A), but for the control condition. The control condition was identical to the main experimental condition, except that word order was shuffled, which rendered the continuous speech less predictable. (C) Schematic of an exemplary trial. First, a task instruction was presented on the screen, followed by a fixation cross at the center of the screen. Next, a word (in this example: "fresh") was shown, starting 2 sec before the onset of the auditory stimulation (speech stream and distractor tone); this word was treated as a target only in the speech task. Participants had to give a response as soon as they heard the target (in this example, the word "fresh"). Consecutive trials were separated by silent intervals lasting between 1.5 sec and 2 sec.

screened for hearing impairments and neurological or psychiatric disorders using a questionnaire. Second, participants were seated in a sound-attenuated and electrically shielded chamber. To reduce potential adaptation or learning effects in the main EEG experiment, participants were first familiarized with the stimuli and tasks during four practice blocks, each consisting of 12 trials, after which they received feedback on their performance. This practice was repeated until participants achieved a target detection accuracy of 80% or higher in each block. Four participants who failed to meet the criterion were excluded from the main experiment.

In the main experiment, four blocks, each containing 72 trials (i.e., 576 sentences) and lasting 18 min, were presented in individually randomized order while EEG was recorded. In one block, participants performed the speech task on the structured speech stimuli (Figure 1A left), which were presented in halves representing the two syntactic structures (presentation order was randomized individually). In another block, the same stimuli were presented, but participants performed the distractor task (Figure 1A right). These two blocks represented the two main experimental conditions. The remaining two blocks were identical to the main ones above, except that the control stimuli (random word sequences) instead of the structured speech stimuli were presented (Figure 1B).

Each block started with the presentation of a task instruction on the screen, followed by a central fixation cross and a word that remained visible during the subsequent auditory stimuli, which started 2 sec later. Participants were instructed to pay attention to the visual word and treat it as a target only when they had received the instruction to perform the speech task. The intertrial interval was randomized between 1.5 sec and 2 sec, and no feedback on task performance was given. Participants could take a break after 36 trials (half a block, after which the syntactic structure switched) for as long as they needed. All stimuli were presented using Presentation software (Version 16.0, Neurobehavioral Systems, Inc) and insert earphones at ~69 $\text{dB}_{\text{SPL}}$ (speech stimuli) and ~67 $\text{dB}_{\text{SPL}}$ (distractor tone).

## EEG Recording

EEG was recorded using a 64-channel active BrainCap (Brain Products) in the standard 10–20 system. All EEG electrodes were referenced online to scalp position FCz. Electrode impedances were kept below 10 kΩ. The EEG recordings were bandpass-filtered (cutoffs: 0.01 and 200 Hz, analog filter) and digitized with a sampling rate of 1000 Hz.

## Data Analysis

### *Behavioral Data Analysis*

Behavioral performance was assessed based on response accuracy (percentage of correctly recognized targets) and the average RTs associated with correct responses.

### *EEG Preprocessing*

EEG data preprocessing was performed offline using EEGLAB 2019.1 (Delorme & Makeig, 2004) and MATLAB 9.4 as follows. First, bad channels with a leptokurtic voltage distribution (i.e., kurtosis higher than five) were replaced by interpolating between the surrounding channels (spherical spline interpolation; percentage of interpolated channels: 2.9 ± 1.8, mean ± *SD* across participants). Second, the interpolated channel data were rereferenced to an average reference. Third, independent component analysis was applied to the channel data to reduce artifacts. For this analysis, the data were first band-pass filtered between 1 Hz and 40 Hz using a linear-phase finite impulse response filter (zero phase shift, filter order: 3300). Artifactual components were identified using the EEGLAB plugin *ICLables* (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) and discarded (artifactual components: 15.8 ± 6.7; mean ± *SD* across participants). The weights of the nonartifactual components were reapplied to the original unfiltered channel data (Jaeger, Bleichner, Bauer, Mirkovic, & Debener, 2018). Three participants were excluded from further analysis because of excessive artifactual components.

### *Analysis of Frequency-tagged Neural Responses*

To assess sentence- and word-rate responses, the continuous EEG data were segmented into 10.08-sec epochs resembling single trials excluding the first sentence interval (which was discarded to avoid onset effects). Epochs were averaged across trials in the time domain separately for each condition and then submitted to a discrete Fourier transform (10080 points, resulting in a spectral resolution of approximately 0.1 Hz). The resulting spectra were averaged across all EEG channels.

Auditory steady-state responses (ASSRs) to the individual words and the distractor were assessed as described above for sentence- and word-rate responses, except for the two following differences: The 10.08-sec epochs were further segmented into 360-msec epochs corresponding to word intervals, and consequently, a reduced number of discrete Fourier transform points was used, resulting in a spectral resolution of 2.78 Hz. To investigate cortical processes involved in sentence tracking, the data analysis was focused on scalp sites showing strong sentence-tracking responses. The strongest responses were found distributed mainly over the central–frontal scalp area (channels: AFz, Fz, F1, F2, F3, and FC1; Figure 2A); thus, these EEG channels were selected and averaged in the frequency domain before statistical analysis.

### *Extraction of Word-position Effects*

To allow assessing effects of word position on auditory processing, neural and behavioral responses were analyzed according to the position (first, second, third, or fourth word interval) at which the stimulus or target occurred within the (pseudo-)sentences.

To obtain a measure of SCF that is unrelated to word phonetics, frequency, or familiarity, the responses to the control speech stimuli were unshuffled, averaged across trials in the time domain (i.e., the constituent words of the unstructured speech were rearranged so that their

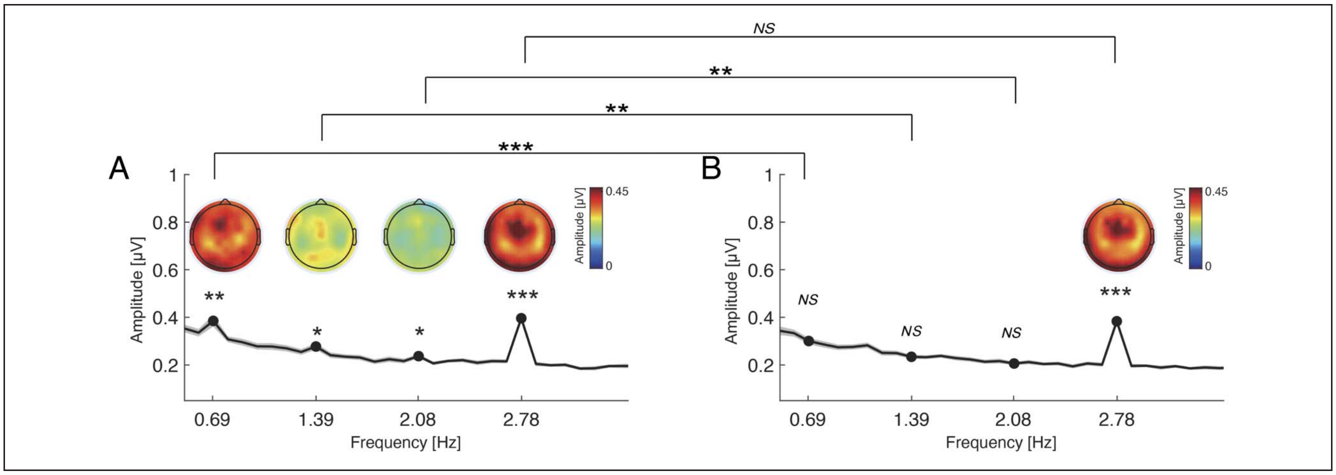

**Figure 2.** Neural responses to hierarchical linguistic structures. (A) The plot shows spectral responses averaged across all EEG channels to structured speech in the speech task as a function of frequency. Asterisks indicate significant responses at the sentence rate (0.69 Hz), word rate (2.78 Hz), and the second and third harmonics (1.39 Hz and 2.08 Hz) of the sentence rate relative to the average response at the six neighboring frequency bins, reflecting participants' detection of the sentence and word structures. The strongest sentence tracking was observed in a central–frontal scalp area. (B) Same as (A), but for unstructured speech in the speech task. A significant response was observed only at the word rate (2.78 Hz), reflecting tracking of the word structure. Responses at the sentence rate and its harmonics were nonsignificant and significantly smaller than in structured speech (A). Black lines and shaded areas represent mean ± *SEM* across participants. N.S. = nonsignificant. **p* < .05, ***p* < .01, ****p* < .001, FDR-corrected.

order matched that of the original, semantically intact sentences), and subtracted from the responses to the matching intact sentences before statistical analysis. We refer to the resulting unbiased, acoustically controlled behavioral and neural measure as SCF (see Introduction section). Note that within-trial positional differences between shuffled words and their unshuffled counterparts varied randomly across trials with a mean close to zero; therefore, the distance between shuffled and unshuffled words could not induce any systematic bias in SCF. Note further that we reversed the sign of the behavioral SCF; thus, positive SCF values represent accelerated RTs and negative values represent decelerated RTs.

The aforementioned analysis steps (stratification according to position and subtraction from unshuffled responses) were applied analogously to responses to the speech (RTs to target words and 42-Hz ASSRs during a given word interval) and to responses to the distractor (RTs to tone-loudness decreases and 25-Hz ASSRs during a given word interval).

### *Analysis of Monotonicity*

To assess monotonous increases of SCF across word positions, Sen's slope (β) was calculated for each participant. Sen's slope is a robust, nonparametric estimate of slope (for more details, see Sen, 1968) and is calculated as

$$\beta = Median\left(\frac{x_j - x_i}{j - i}\right)$$

where $x_j$ and $x_i$ are the data values at positions $j$ and $i$, respectively ($j > i$). Positive values of β indicate upward trends across positions.

## Statistical Analysis

The single-subject estimates from the 21 participants were submitted to second-level (random-effect) statistical analyses. A significance criterion $\alpha = .05$ was used, and Type I error probabilities inflated by multiple comparisons were corrected by false discovery rate (FDR).

### *Spectral Peak*

Spectral peaks at tagging frequencies were compared with the average spectral response at the six neighboring frequency bins (noise floor, excluded the 50-Hz line noise) using a paired *t* test. The statistical test was applied to each tagging frequency: 0.69 Hz (sentence rate), 1.39 Hz (second harmonic), 2.08 Hz (third harmonic), 2.78 Hz (word rate), 25 Hz (tone tag), and 42 Hz (speech tag). The spectral peaks at speech rates were further compared between the structured speech condition and control condition with a paired *t* test.

### *SCF*

Statistical analysis of the word-position effect involved a one-way repeated-measures ANOVA including the factor Position (word interval: first, second, third, or fourth) and two-way repeated-measures ANOVAs including an additional factor: either Attention (speech task/attended vs. distractor task/unattended) or Sound Stream (speech stream [RTs to target words; 42-Hz ASSR] vs. tone stream [RTs to tone-loudness decreases; 25-Hz ASSR]). The two syntaxes were pooled. To assess the SCF at each word position and its across-word trend, paired *t* tests were applied that respectively compared SCF at each position to zero and Sen's slope to zero. The assumption of normality was verified with Kolmogorov–Smirnov tests, which did not detect any significant deviation from normality (all $p > .05$). The assumption of sphericity was assessed with Mauchly's tests, and Greenhouse–Geisser correction was applied to adjust the degrees of freedom when the assumption was violated.

To further seek evidence of alternative hypotheses (H1) versus the null hypotheses (H0), Bayesian repeated-measures ANOVAs including the factor Position and, where applicable, the additional factor Attention or Sound Stream were carried out in JASP (Love et al., 2019). Evidence was inferred from the Bayes Factor ($BF_{10}$), defined as the ratio of the likelihood of the data fitting the alternative hypothesis to the likelihood of the data fitting the null hypothesis. $BF_{10}$ values higher than 10 or lower than 1/10 represent strong evidence for the alternative and null hypotheses, respectively. A $BF_{10}$ higher than 3 and lower than 1/3 represents moderate evidence for the alternative and null hypotheses, respectively. A $BF_{10}$ between 3 and 1/3 reflects weak or anecdotal evidence for either hypothesis (Keysers, Gazzola, & Wagenmakers, 2020).

### *Correlation Analysis*

To test for a linear association between behavioral and neural SCF, repeated-measures correlation was assessed using the *rmcorr* R package (Bakdash & Marusich, 2017). Repeated-measures correlation fits a linear model between two variables while controlling for non-independence within participants. In the present study, it was used to estimate the common intra-individual regression slope between SCF of RTs and SCF of 42-Hz ASSR, while allowing the intercepts to vary across participants.

## RESULTS

As expected after the task training, participants were able to perform both tasks successfully. On average, they correctly detected between 94% and 86% of the targets. Higher accuracies were observed in the speech task (on average 93%) than the distractor task (on average 87%). To assess neural tracking of hierarchical linguistic structures, we analyzed neural responses at the sentence rate (0.69 Hz), its harmonics (1.39 Hz and 2.08 Hz), and word rate (2.78 Hz). As shown in Figure 2A, participants reliably showed spectral peaks at these rates (compared with the average spectral amplitude of the six neighboring frequency bins: $p < .05$, paired $t$ test, FDR-corrected), indicating participants could successfully track the sentence and word structures. In the unstructured speech condition, participants showed a spectral peak only at the word rate (Figure 2B), and the ASSR at the sentence rate and its harmonics were significantly smaller than in structured speech condition (paired $t$ test, FDR-corrected).

Figure 3 shows participants' neural responses for each condition and each word position, averaged across channels showing the strongest sentence tracking. Prominent ASSRs at each tagging frequency (25 Hz and 42 Hz) were observed for all conditions and positions (spectral peaks at tagging frequency > average spectral amplitude of the six neighboring frequency bins: $p < .05$; paired $t$ test, FDR-corrected). These results indicate that auditory processing of the speech and distractor stimuli could be reliably assessed and separated. Analysis of the 42-Hz ASSR at each individual electrode revealed that neural responses to individual words were most prominent at central–frontal scalp regions, consistently across conditions and word positions (Figure 4). Thus, the central–frontal channels that showed the strongest sentence tracking and were selected for further analysis (Figure 2A) also showed strong 42-Hz ASSRs.

### Buildup of SCF of Auditory Word Processing during Sentence Tracking

To assess the word-position effect during sentence tracking at the linguistic level, we first compared behavioral responses to target words presented at the different word positions. We found that SCF (a measure of sentential context effects controlled for word phonetics, frequency, and familiarity, see Methods section) increased monotonically across word positions, reflecting a gradual decrease of RTs toward the end of sentences (Figure 5A). This observation of an across-word pattern in SCF was statistically confirmed by a one-way ANOVA including Word Position as a four-level factor, which revealed a main effect of Word Position on SCF, $F(3, 60) = 30.86, p < .001, \eta_p^2 = .61$. In line with this, a corresponding Bayesian ANOVA revealed a $BF_{10}$ of $5.03 \times 10^{10}$, thus providing decisive evidence in favor of the main effect model compared with the null model (i.e., no word-position effect). This word-position effect shows that sentential context gradually accelerated the perceptual processing of words at later positions in the sentences.

Applying the same analysis (one-way repeated-measures ANOVA, Bayesian ANOVA) to the neural responses to individual words revealed a similar across-word pattern. More specifically, SCF of the 42-Hz ASSR monotonically increased across word positions (Figure 5B). This observation of an across-word pattern in SCF was supported by strong evidence for a main effect of Word Position on SCF, $F(3, 60) = 5.23, p = .003, \eta_p^2 = .21, BF_{10} = 22.17$. These results show that sentential context gradually modulated the cortical word-processing in the sentences. Interestingly, the observed effect is positive (i.e., larger responses at later word positions), which we further address in the Discussion section.

Statistical analysis of Sen's slope (a nonparametric estimate of slope; see Methods section) revealed a significant positive difference from zero for behavioral SCF and neural SCF (RT: $\beta = 39.5 \pm 5.7$, mean ± *SEM*; paired $t$ test: $t(20) = 6.89, p < .001$; Bayesian $t$ test: $BF_{+0} = 3.40 \times 10^4$; ASSR: $\beta = 0.009 \pm 0.0031$, mean ± *SEM*; paired $t$ test:

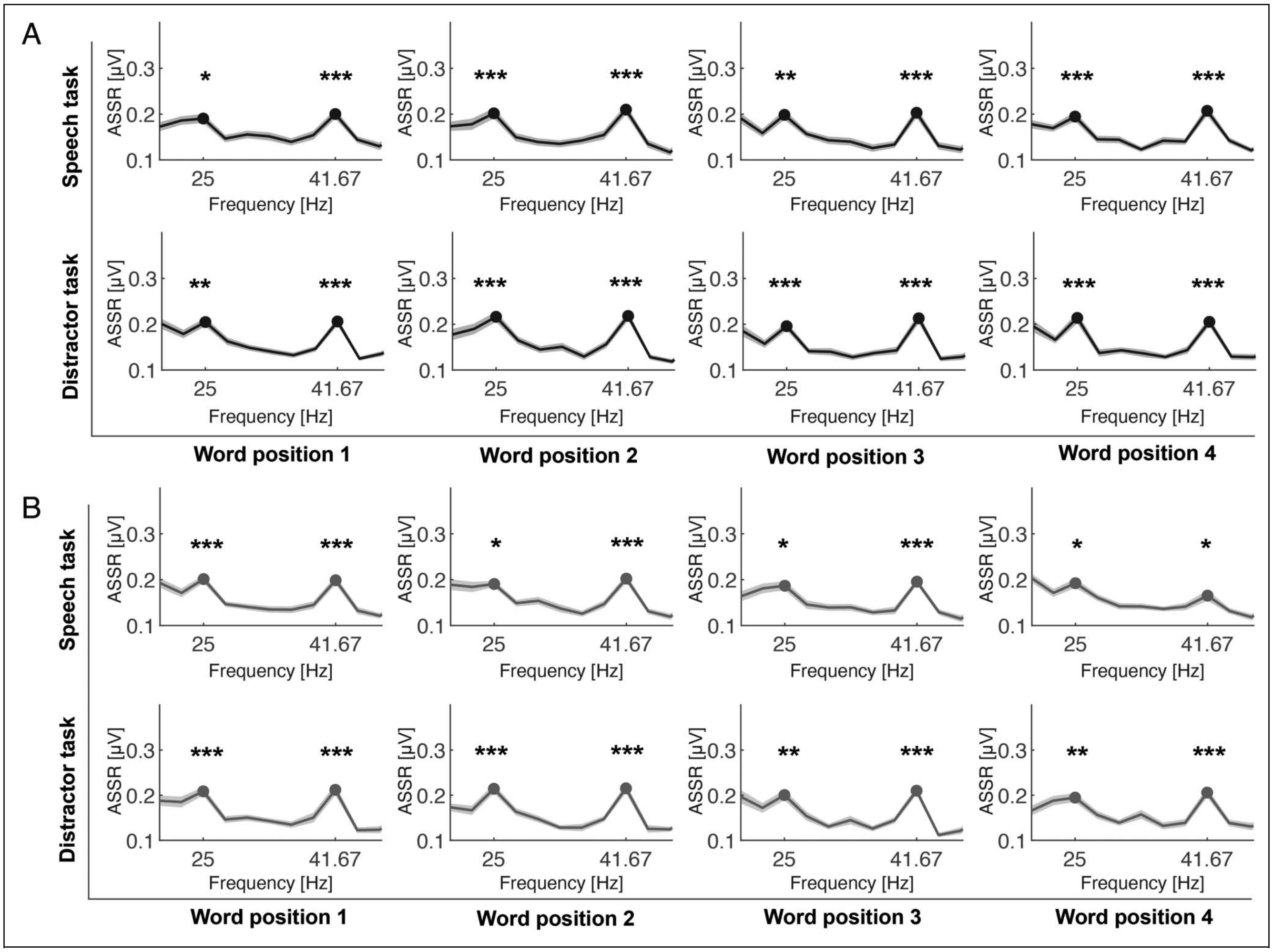

**Figure 3.** Neural responses to speech and distractor for each condition. Each plot shows ASSR in the range of the tagging frequencies (25 Hz and 42 Hz) averaged across channels in the central–frontal area. Asterisks indicate significant responses at the tagging frequency relative to the average responses at the six neighboring frequency bins. (A) Rows represent the two main experimental conditions involving structured speech stimuli (first: speech task; second: distractor task). Columns represent the four word positions. (B) Same as (A), but for the acoustic control condition (shuffled, unstructured speech) after unshuffling responses. Black lines and shaded areas represent mean ± *SEM* across participants. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

$t_{20} = 3.00, p = .005$; Bayesian *t* test: $BF_{+0} = 13.54$), indicating significant upward trends across word positions. These results support our conclusion that SCF increased monotonously across word positions.

To further test the effects of SCF on each word position, we compared SCF at each position to zero. We found that both behavioral SCF and neural SCF were significantly larger than zero at the last word position (RT: $t_{20} = 8.17$, $p < .001$, FDR-corrected, $BF_{+0} = 3.48 \times 10^5$; ASSR: $t_{20} = 4.15, p < .001$, FDR-corrected; $BF_{+0} = 135.00$), revealing that context facilitates both behavioral and neural responses to words especially at the end of the sentence.

## Potential Effect of Attention on the Buildup of SCF

To explore whether the observed buildup of SCF during sentence tracking requires listeners to pay selective attention to the speech, we tested whether the neural word-position effect persisted even when participants withdrew their attention from the speech stimuli. Contrary to the results above, we found no evidence for the main effect of Word Position on SCF of the 42-Hz ASSR when participants performed the distractor task (one-way repeated-measures ANOVA: $F(3, 60) = 0.31, p = .82, \eta_p^2 = .015$; Bayesian ANOVA: $BF_{10} = 0.094$; Figure 6). A two-way ANOVA with factors Word Position and Attention revealed a Word Position × Attention interaction effect on SCF; however, the evidence for the interaction effect is weak, $F(3, 60) = 3.07, p = .037, \eta_p^2 = .13, BF_{incl} = 0.89$ (comparing the model that contains the effect to an equivalent model without the effect). These results suggest that the buildup of SCF during sentence tracking depends on the listener's attention to the speech.

## No Transfer of SCF to Auditory Processing of Concurrent Sound Streams

We further explored whether the observed buildup of SCF during sentence tracking transfers to the processing of concurrent sound streams. To this end, we assessed

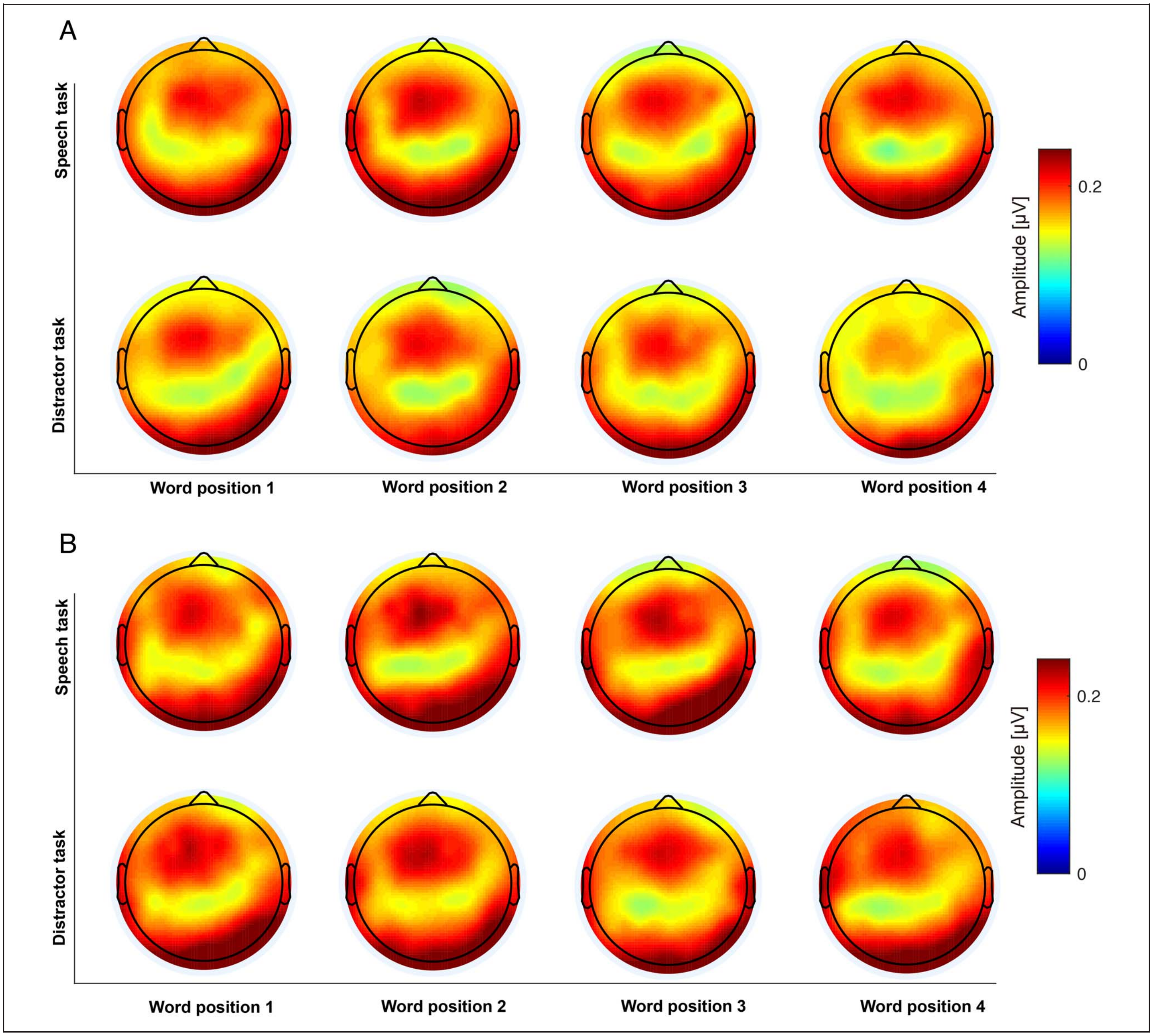

**Figure 4.** Topographic maps of neural responses to speech for each condition. Each plot shows the spatial distribution of the 42-Hz ASSR across the scalp. (A) Rows represent the two main experimental conditions involving structured speech stimuli (first: speech task; second: distractor task). Columns represent the four word positions. (B) Same as (A), but for the acoustic control condition (shuffled, unstructured speech) after unshuffling the responses. The 42-Hz ASSR was most prominent in frontocentral regions.

whether the sentential context facilitated also participants' responses to the distractor tone. We found that SCF of RTs to the distractor tone was nearly zero at each word position and did not differ significantly across these positions (one-way repeated-measures ANOVA: $F(3, 60) = 1.78, p = .16, \eta_p^2 = .08$; Bayesian ANOVA: $BF_{10} = 0.42$; Figure 7A). Moreover, the latter nonsignificant across-word variations were significantly smaller than those observed in responses to the speech (interaction Word Position × Sound Stream: $F(3, 60) = 27.33, p < .001, \eta_p^2 = .577$, $BF_{incl} = 9.59 \times 10^7$).

Consistent with this behavioral result, we found no main effect of Word Position on neural responses to the distractor tone, as measured by SCF of the 25-Hz ASSR (one-way repeated-measures ANOVA: $F(3, 60) = 1.04, p = .38, \eta_p^2 = .05$; Bayesian ANOVA: $BF_{10} = 0.88$, Figure 7B). Similar to the behavioral results above, a two-way ANOVA revealed a significant Word Position × Sound Stream interaction, $F(3, 60) = 2.97, p = .039, \eta_p^2 = .129$, $BF_{incl} = 1.06$. In summary, these results indicate that the observed buildup of SCF on auditory word processing did not transfer to the processing of the concurrent distractor tone.

## Relation between SCF of Neural and Behavioral Responses to Speech

To explore whether the SCF of behavioral and neural responses across word positions were functionally

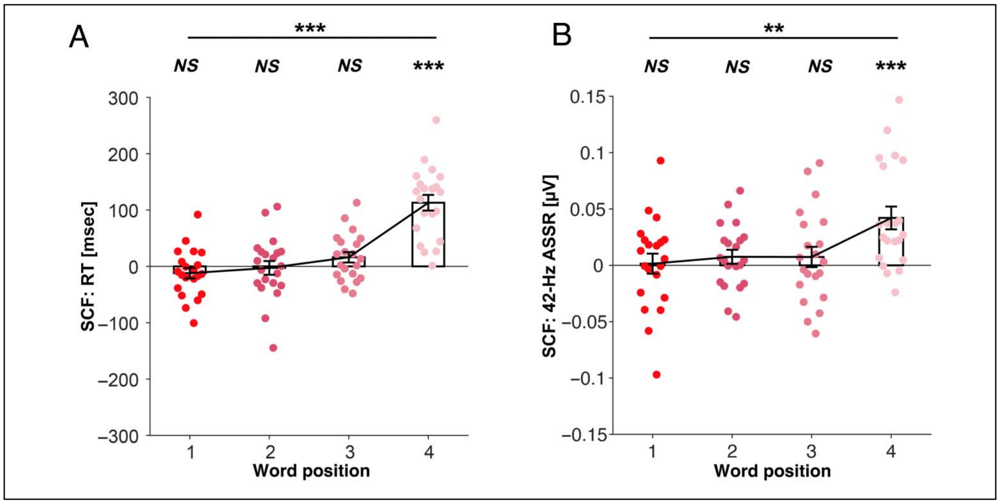

**Figure 5.** Across-word pattern of SCF: behavioral and neural responses to individual words. (A) The bar plots show SCF of participants' RT to individual words (structured speech stimuli minus control stimuli) as a function of word position within attended sentences. SCF of RTs increased monotonically. Note that positive SCF values represent accelerated RTs and negative values represent decelerated RTs. Therefore, sentential context gradually accelerated behavioral responses, revealing a word-position effect on linguistic word processing. (B) Same as (A), but for neural responses to words. Sentential context resulted in stronger ASSR amplitudes at later word positions. The significance marker above each bar indicates whether SCF was significantly larger than zero (corrected for multiple comparisons). Significant SCF was observed in the last word position for both behavioral and neural responses. Red dots represent individual participants. Bars and error bars represent mean ± *SEM* across participants. *NS* = nonsignificant. $^{**}p < .01$, $^{***}p < .001$.

coupled, we tested the correlation between SCF of RTs and SCF of 42-Hz ASSR (same data as in Figure 5). A repeated-measures correlation analysis revealed a significant positive correlation between SCF of RTs and SCF of 42-Hz ASSR ($r_{\mathrm{rm}} = .414$, $p < .001$, 95% CI [0.187, 0.599]; Figure 8). This neural-behavioral result indicates that the contextual facilitation of cortical processing of individual words was associated with a corresponding facilitation of the behavioral detection to these words, suggesting a link between acoustic and linguistic levels of auditory word processing.

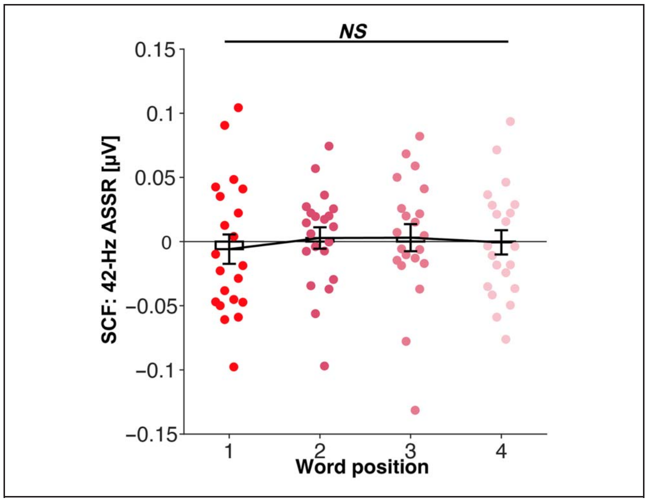

**Figure 6.** Across-word pattern of SCF: neural responses to unattended individual words. The bar plots show SCF of participants' neural responses to individual words (SCF: structured speech stimuli minus control stimuli) as a function of word position within unattended sentences. Contrary to the significant buildup observed during the speech task (Figure 5B), no significant word-position effect on SCF was observed when participants withdrew their attention from the speech to perform the distractor task. Red dots represent individual participants. Bars and error bars represent mean ± *SEM* across participants. *NS* = nonsignificant.

## Effect of Syntax on the Buildup of SCF

In an additional exploratory analysis, we investigated whether the buildup of SCF during sentence tracking depends on the syntax of the speech (i.e., syntax NVAN vs. AANV). A two-way ANOVA including Word Position (word interval: first, second, third, or fourth) and Speech Syntax (syntax NVAN and AANV) as factors revealed weak evidence for an interaction (Word Position × Syntax) effect on SCF of RTs, $F(3, 60) = 3.24$, $p = .028$, $\eta_{\mathrm{p}}^2 = .14$, $\mathrm{BF}_{\mathrm{incl}} = 1.61$ (Figure 9). Post hoc comparisons for each syntax confirmed a main effect of Word Position on SCF in syntax NVAN as well as syntax AANV (syntax NVAN: $F(3, 60) = 25.07$, $p < .001$, $\eta_{\mathrm{p}}^2 = .56$, $\mathrm{BF}_{10} = 3.67 \times 10^{8;}$ syntax AANV: $F(3, 60) = 9.23$, $p < .001$, $\eta_{\mathrm{p}}^2 = .32$, $\mathrm{BF}_{10} = 1.92 \times 10^{3}$). Statistical analysis of Sen's slope revealed an upward trend across word positions for each syntax (syntax NVAN: $\beta = 52.0 \pm 7.8$; syntax AANV: $\beta = 29.7 \pm 6.6$; mean ± *SEM*) that was significantly stronger in syntax NVAN than syntax AANV (paired *t* test, $t(20) = 2.40$, $p = .026$). Post hoc pairwise comparisons between the two syntaxes at each individual position showed no significant difference at any position ($p > .05$, FDR-corrected). These results suggest that the observed Word Position × Syntax interaction effect on SCF of RTs reflects differences in across-word trend, rather than a difference at a single specific word position. Applying the same two-way ANOVA as above to the neural data revealed no such interaction effect on SCF of 42-Hz ASSR, $F(3, 60) = 1.26$, $p = .30$, $\eta_{\mathrm{p}}^2 = .059$, $\mathrm{BF}_{\mathrm{incl}} = 0.082$. In summary, these exploratory results suggest that the speech syntax may influence the temporal shape of SCF of linguistic word processing.

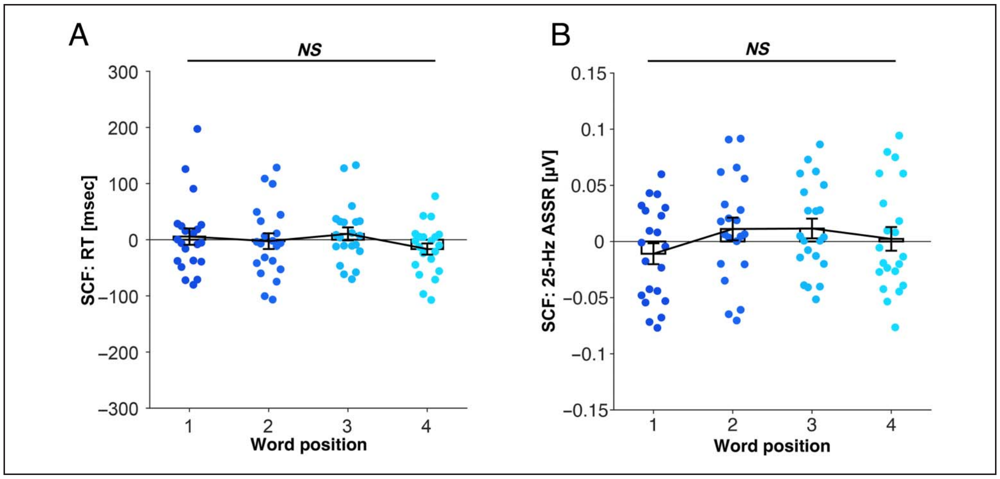

**Figure 7.** Across-word pattern of SCF: behavioral and neural responses to the distractor tone. (A) The bar plots show SCF of participants' RT to the distractor tone (SCF: structured speech stimuli minus control stimuli) as a function of word position within unattended sentences. No effect of word position on SCF was observed. (B) Same as (A), but for neural responses to the distractor tone (SCF of 25-Hz ASSR). Blue dots represent individual participants. Bars and error bars represent mean ± *SEM* across participants. *NS* = nonsignificant.

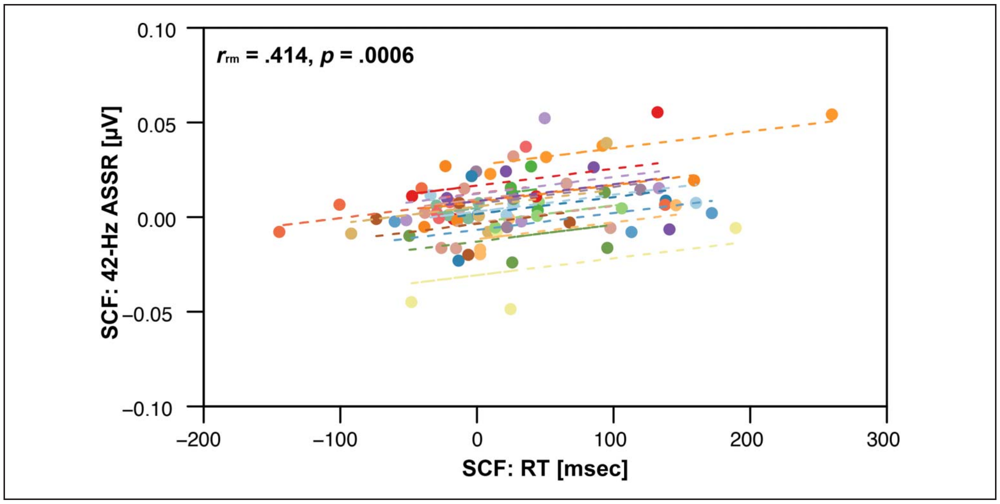

**Figure 8.** Relation between SCF of behavioral and neural responses to speech. The scatterplot shows results from a repeated-measures correlation analysis testing for a functional coupling between SCF of RTs and SCF of 42-Hz ASSR to words across word positions. Dots with the same color represent responses to words at different positions from the same participant. The corresponding lines show the repeated-measures correlation fit for each participant. Correlation coefficient $r_{\mathrm{rm}}$ and $p$ value describe, respectively, the strength and statistical significance of the common intra-individual coupling.

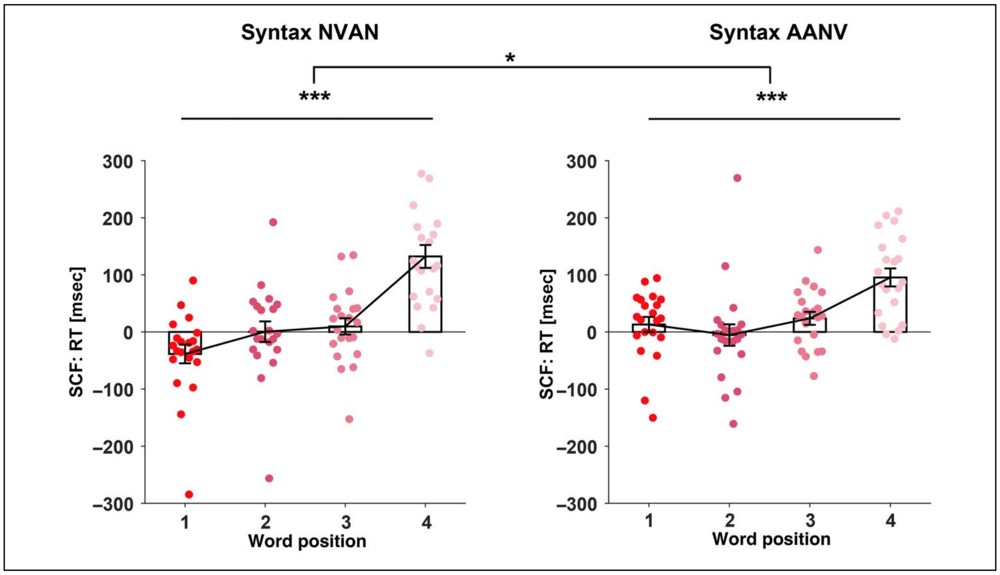

**Figure 9.** Across-word pattern of SCF: behavioral responses to individual words in sentences of syntax NVAN and syntax AANV. The bar plots show SCF of participants' RT to individual words (structured speech stimuli minus control stimuli) as a function of word position within attended sentences of syntax NVAN (left plot) and syntax AANV (right plot). Overall, sentential context gradually accelerated behavioral responses at later word positions, but the exact across-word pattern depended significantly on the syntax.

## DISCUSSION

In this study, we investigated the dynamics of auditory word processing during the tracking of continuous meaningful speech. We found that (i) sentential context alone incrementally facilitates behavioral and neural responses to individual words as the sentence unfolds. This SCF (ii) depends on listeners' endogenous selective attention to the speech; (iii) affects selectively the processing of the speech stream, not that of a concurrent sound stream; and (iv) may be shaped by the syntax of the speech. In addition, we found that the facilitation of behavioral and neural responses to individual words in our paradigm was coupled. Overall, these results provide evidence for the hypothesis that the dynamics in auditory processing during sentence tracking can be driven exclusively by buildup of top–down factors, specifically sentential predictions derived from the speech.

### Sentential Predictability Elicits Word-processing Dynamics

Our observation that sentential context incrementally facilitates perceptual word detection during sentence tracking aligns with previous work showing that word processing may be accelerated across word positions in sentences (Lam et al., 2016; Van Petten & Kutas, 1990, 1991; Marslen-Wilson & Tyler, 1975). Our finding of a word-position effect on word processing that originates exclusively from sentential context (i.e., in absence of variations in word phonetics, frequency, familiarity, and overt prosody) is novel. It provides strong evidence to the notion that the facilitation of the processing of later-occurring words can be attributed to a gradual increase in sentential predictability (Clark, 2013; Hagoort & Indefrey, 2013; Van Petten & Luka, 2012; DeLong, Urbach, & Kutas, 2005; Van Petten & Kutas, 1991). As a sentence unfolds, semantic context incrementally builds up, and the accruing context increases the predictability of later words in the sentence. Consequently, listeners in our study may have recognized the later words more quickly than the earlier words because their meaning was more predictable.

Although syntactic predictions can influence linguistic analysis of individual words during sentence tracking (see Introduction section), they unlikely caused the word-position effects observed in our study. We kept the syntax constant across all sentences within a given block, implying no variations in syntactic predictability across word positions. Alternatively, the observed facilitation at later words may have been mediated by covert prosody (prosodic boundaries internally projected onto the stimulus in the absence of overt prosodic phrase boundary markers) that listeners possibly generated from the sentential context (Glushko, Poeppel, & Steinhauer, 2022; Breen, 2014). Although our study cannot distinguish between these two accounts (semantics vs. covert prosody), it does unequivocally demonstrate a purely top–down influence on word processing that originates from sentential context. Similar position effects have also been reported at the level of syllables. EEG studies found that word-initial syllables elicit a larger N1 component than word-medial syllables and the processing of the syllables is modulated by predictability (Astheimer & Sanders, 2011; Sanders & Neville, 2003). Thus, studies at syllable and word levels jointly demonstrate the temporal dynamics of speech processing and support a role of predictability in these dynamics.

Consistent with our behavioral results, the analysis of neural responses to speech revealed a significant word-position effect. We found that sentential context incrementally facilitates cortical processing of auditory words (as measured with ASSR) as the sentence unfolds. As explained above, the primary origin of these dynamics in cortical word processing may be attributed to semantic predictability. This interpretation is in line with neural findings showing effects of semantic comprehension on much slower phase-locked cortical responses to natural speech envelope (Kaufeld et al., 2020; Pefkou, Arnal, Fontolan, & Giraud, 2017; Park, Ince, Schyns, Thut, & Gross, 2015; Peelle, Gross, & Davis, 2013; Ahissar et al., 2001). Our measure of cortical word processing (42-Hz ASSR) likely reflects phase-locked responses to the artificially induced rapid modulations of the word waveform (Picton, John, Purcell, & Plourde, 2003; Kuwada, Batra, & Maher, 1986; Rodriguez, Picton, Linden, Hamel, & Laframboise, 1986). Therefore, the neural results may indicate a primary effect of sentential predictability on cortical processing of lower-level acoustic speech-signal features.

The positive direction of the effect (i.e., higher semantic predictability leading to stronger auditory-evoked response) may at first sight seem to conflict with predictive coding accounts, which typically demonstrate reduced neural activity for highly predictable content (Grisoni et al., 2021; Kutas & Federmeier, 2011). However, we should emphasize that in our finding, predictability led to increased ASSR, representing the processing of the amplitude modulations of the acoustic signal, rather than linguistic processing. This suggests that increased predictability of linguistic content allowed listeners to shift processing resources from the linguistic analysis of the ongoing speech signal to its auditory analysis, which probably led to a more faithful auditory cortical representation of the speech signal and accelerated word recognition.

We further observed a significant correlation between the facilitation in behavioral and neural responses to individual words: Words that were detected more quickly tended to elicit stronger word processing. This suggests that the enhancement of neural responses to the acoustic speech signal improved the perceptual detectability of the target word.

### Potential Effect of Selective Attention on the Buildup of SCF

The SCF of neural word processing was observable only when listeners paid selective attention to the speech. This extends previous findings showing contributions of selective attention to the processing of individual words. For example, speech studies using continuous isochronous syllable sequences have shown that listeners' attention may be required for the grouping of consecutive syllables into words (Ding et al., 2018; Makov et al., 2017). Moreover, it has been shown that listeners detect target words in sentences more rapidly when they are semantically primed to focus their attention on these words (Cutler & Fodor, 1979). Our results extend these findings to the grouping of words into sentences and, more importantly, the incremental SCF of auditory word processing. As explained above, the observed word-processing dynamics likely originated from sentential predictability, not from phonetic or overt prosodic effects that may be more immune to selective attention (Bosker, Sjerps, & Reinisch, 2020). As such, it is likely that our listeners grouped the words into sentences and could thereby extract contextual information only when they paid attention to the speech. Whether selective attention is generally necessary for sentential predictions—and, by extension, word-processing dynamics—remains unclear from our study and would require testing a wider range of speech stimuli and tasks.

## SCF Does Not Transfer to Processing of Concurrent Sound Streams

We observed no SCF of the processing of the concurrent distractor tone (or significant temporal changes therein), suggesting that the dynamics in word processing did not transfer to the processing of the distractor. In contrast, some previous studies observed a word-position effect on acoustic event detection during speech processing; for example, shorter RTs were found for clicks occurring during later versus early words (Lobina et al., 2018; Holmes & Forster, 1970). The difference to these previous results may reflect methodological differences: In the previous studies, the concurrent sound was an occasional brief acoustic event and participants were required to pay attention simultaneously to a click and the meaning of the ongoing speech. In contrast, our study used a continuous distractor tone encouraging auditory stream segregation and two separate tasks requiring no division of attention. Thus, our null result showing no effect of sentential speech context on the processing of a simultaneous distractor tone may indicate that sentential predictions affect only the processing of the sound stream from which these predictions are derived, not input that is perceptually separate from that stream. Put differently, SCF of acoustic word processing (i.e., processing of the amplitude variations of the sound wave of words) probably occurs at a later stage than the perceptual segregation of the speech from concurrent streams, which may occur within ~200 msec after speech onset (Alain, Arsenault, Garami, Bidelman, & Snyder, 2017; Bidelman & Yellamsetty, 2017; Bidelman & Alain, 2015) in auditory structures as early as the cochlear nucleus (Pressnitzer, Sayles, Micheyl, & Winter, 2008). It should be noted that, in contrast to the speech stream, the ongoing tone contained no identifiable structure, implying that listeners could not extract valid predictions from it. A temporally structured tone might have made it more likely to observe a significant temporal (across-word) pattern also in tone processing.

## Speech Syntax Modulates the Buildup of SCF

Our exploratory analysis of syntax effects revealed that the buildup of SCF may be influenced by the speech syntax. Syntax modulated the temporal shape of the behavioral, but not the neural, word-position effect, which may be related to asymmetries between our behavioral and neural measures. The behavioral measure probably captured linguistic processes, whereas the neural measure captured more acoustics-related processes. Therefore, the syntax probably affected the facilitation of linguistic, not acoustic, word processing.

A potential explanation for this side observation may be differences in the hierarchical syntactic structure of NVAN and AANV. NVAN has a more specific verb phrase (consisting of a verb, adjective, and noun, e.g., "eat fresh fish"), whereas AANV has a more specific noun phrase (consisting of an adjective, adjective, and noun, e.g., "Two big dogs"). Listeners probably built an internal representation of the hierarchical syntactic structure during speech processing (Friederici, 1995, 2002), and the differences in these structures might have resulted in differences in the buildup of this representation or its utilization for the analysis of the incoming speech signal. However, this interpretation is tentative and requires thorough verification with a better-suited study design using a larger number of syntactic structures and combining sentences of various lengths.

## Conclusion

In summary, our study reveals that the acoustic and linguistic processing of auditory words during sentence tracking builds up dynamically and these dynamics can be driven exclusively by top–down factors, in particular sentential predictions derived from the processed speech. These factors may depend on the listener's selective attention to the speech and affect only the processing of speech, not that of perceptually separate sound streams. It appears that auditory and semantic processes during sentence tracking interact reciprocally: Auditory processing of individual words may inform the formation of syntactic and semantic structure, and predictions derived from these structures may modulate the auditory processing of the words, provided that the listener pays selective attention to the speech stream. These conclusions could be drawn because of our control of phonetics and prosody, at the expense of the naturalness of our auditory stimuli. Future work is encouraged to explore whether these findings generalize to natural speech.

# APPENDIX

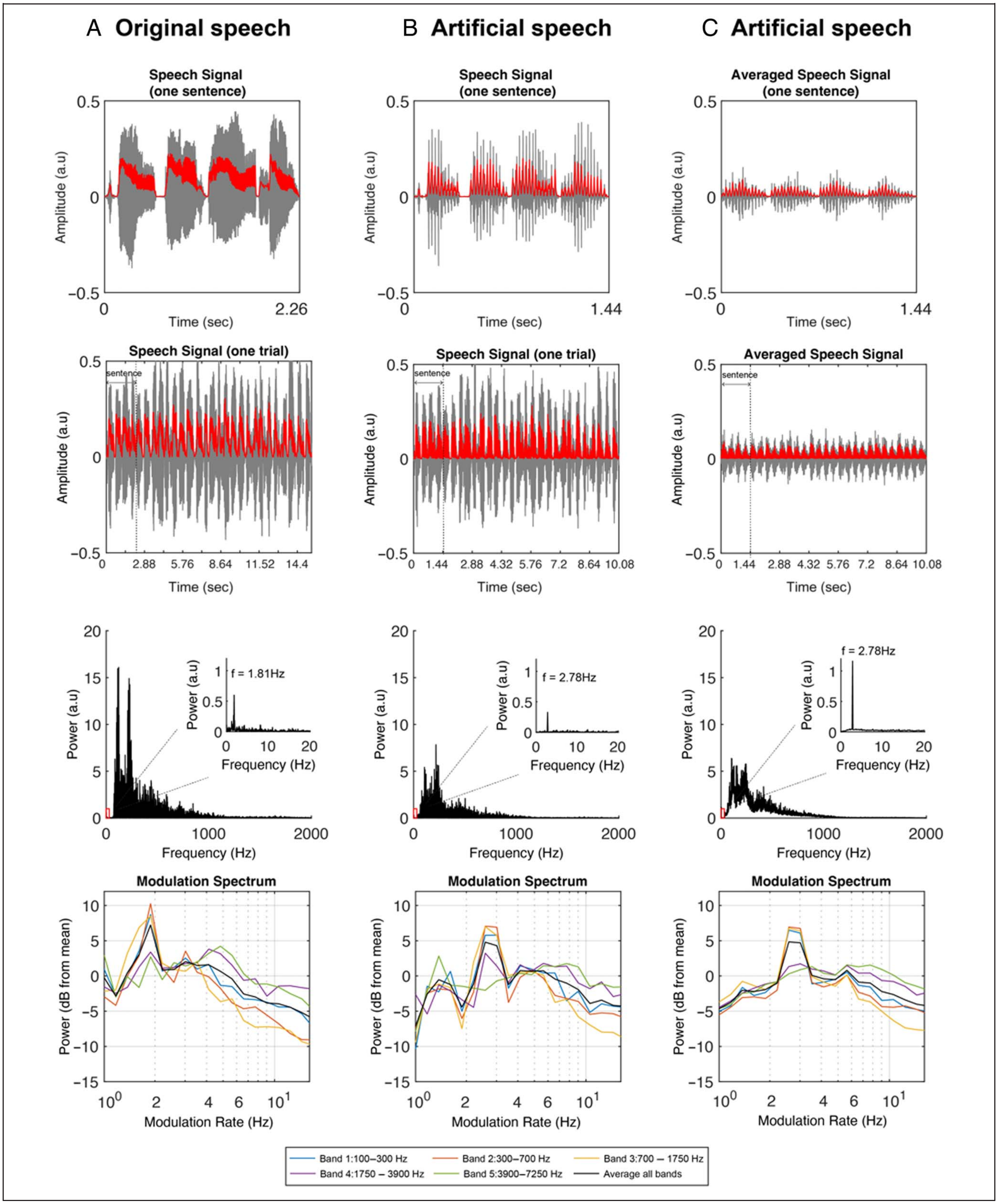

**Figure A1.** Speech acoustics. (A, B, and C) show, respectively, an exemplary original excerpt, a time-compressed version of the same excerpt, and the average of all time-compressed excerpts. The first to fourth row show, respectively, the sound waveform of a single sentence (in gray, amplitude envelope is shown in red), the sound waveform of a single trial (seven sentences), the average power spectrum of all sentences, and the average power spectrum of the amplitude envelopes of all sentences (modulation spectrum, shown for the broadband envelope as well as for non-overlapping narrow envelope bands).

Reprint requests should be sent to Min Wu, Maastricht University, Oxfordlaaan, 55, Maastricht, 6200 MD, the Netherlands, or via e-mail: m.wumin@maastrichtuniversity.nl.

### Data Availability Statement

Data and analysis scripts have been deposited in Open Science Framework (https://osf.io/qzvsn/).

### Author Contributions

Min Wu: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Writing—Original draft. Hans Rutger Bosker: Conceptualization; Methodology; Writing—Review & editing. Lars Riecke: Conceptualization; Data curation; Methodology; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

### Funding Information

This work was supported by Maastricht University and China Scholarship Council (CSC; https://dx.doi.org/10.13039/501100004543), grant number; 201906320078 to M. W.

### 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.

### Note

[1] We chose to frame the present study in terms of prediction without taking a particular stance in the debate between prediction and integration (Pickering & Gambi, 2018).

## 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

Alain, C., Arsenault, J. S., Garami, L., Bidelman, G. M., & Snyder, J. S. (2017). Neural correlates of speech segregation based on formant frequencies of adjacent vowels. *Scientific Reports*, *7*, 40790. https://doi.org/10.1038/srep40790, PubMed: 28102300

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, *49*, 3512–3516. https://doi.org/10.1016/j.neuropsychologia.2011.08.014, PubMed: 21875609

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456, PubMed: 28439244

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

Bidelman, G. M., & Alain, C. (2015). Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept. *Neuropsychologia*, *68*, 38–50. https://doi.org/10.1016/j.neuropsychologia.2014.12.020, PubMed: 25542675

Bidelman, G. M., & Yellamsetty, A. (2017). Noise and pitch interact during the cortical segregation of concurrent speech. *Hearing Research*, *351*, 34–44. https://doi.org/10.1016/j.heares.2017.05.008, PubMed: 28578876

Bosker, H. R., Sjerps, M. J., & Reinisch, E. (2020). Temporal contrast effects in human speech perception are immune to selective attention. *Scientific Reports*, *10*, 5607. https://doi.org/10.1038/s41598-020-62613-8, PubMed: 32221376

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, *8*, 37–50. https://doi.org/10.1111/lnc3.12061

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204. https://doi.org/10.1017/S0140525X12000477, PubMed: 23663408

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*, 49–59. https://doi.org/10.1016/0010-0277(79)90010-6, PubMed: 436402

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*, 3914–3932. https://doi.org/10.1162/jocn_a_00084, PubMed: 21745006

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. https://doi.org/10.1038/nn1504, PubMed: 16007080

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *134*, 9–12. https://doi.org/10.1016/j.jneumeth.2003.10.009, PubMed: 15102499

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*, 1025–1066. https://doi.org/10.1162/COLI_a_00160

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

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *Journal of Neuroscience*, *38*, 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017, PubMed: 29255005

Drijvers, L., Jensen, O., & Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping*, *42*, 1138–1152. https://doi.org/10.1002/hbm.25282, PubMed: 33206441

Fedorenko, E., Scott Terri, L., Brunner, P., Coon William, G., Pritchett, B., Schalk, G., et al. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, E6256–E6262. https://doi.org/10.1073/pnas.1612132113, PubMed: 27671642

Ferreira, F., & Qiu, Z. (2021). Predicting syntactic structure. *Brain Research*, *1770*, 147632. https://doi.org/10.1016/j.brainres.2021.147632, PubMed: 34453937

Foltz, A. (2021). Using prosody to predict upcoming referents in the L1 and the L2: The role of recent exposure. *Studies in Second Language Acquisition*, *43*, 753–780. https://doi.org/10.1017/S0272263120000509

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259–281. https://doi.org/10.1006/brln.1995.1048, PubMed: 7583190

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8, PubMed: 15866191

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*, 262–268. https://doi.org/10.1016/j.tics.2012.04.001, PubMed: 22516238

Giani, A. S., Ortiz, E., Belardinelli, P., Kleiner, M., Preissl, H., & Noppeney, U. (2012). Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *Neuroimage*, *60*, 1478–1489. https://doi.org/10.1016/j.neuroimage.2012.01.114, PubMed: 22305992

Glushko, A., Poeppel, D., & Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*, *12*, 14759. https://doi.org/10.1038/s41598-022-18162-3, PubMed: 36042220

Gransier, R., van Wieringen, A., & Wouters, J. (2017). Binaural interaction effects of 30–50 Hz auditory steady state responses. *Ear and Hearing*, *38*, e305–e315. https://doi.org/10.1097/AUD.0000000000000429, PubMed: 28358747

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2021). Correlated brain indexes of semantic prediction and prediction error: Brain localization and category specificity. *Cerebral Cortex*, *31*, 1553–1568. https://doi.org/10.1093/cercor/bhaa308, PubMed: 33108460

Guediche, S., Reilly, M., Santiago, C., Laurent, P., & Blumstein, S. E. (2016). An fMRI study investigating effects of conceptually related sentences on the perception of degraded speech. *Cortex*, *79*, 57–74. https://doi.org/10.1016/j.cortex.2016.03.014, PubMed: 27100909

Hagoort, P., & Indefrey, P. (2013). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347–362. https://doi.org/10.1146/annurev-neuro-071013-013847, PubMed: 24905595

Holmes, V. M., & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception & Psychophysics*, *7*, 297–301. https://doi.org/10.3758/BF03210171

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541–573. https://doi.org/10.1016/j.jml.2007.06.013, PubMed: 19190719

Jaeger, M., Bleichner, M., Bauer, A.-K., 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., Zou, J., Zhou, T., & Ding, N. (2018). Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nature Communications*, *9*, 5374. https://doi.org/10.1038/s41467-018-07773-y, PubMed: 30560906

Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, *40*, 9467–9475. https://doi.org/10.1523/JNEUROSCI.0302-20.2020, PubMed: 33097640

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*, 788–799. https://doi.org/10.1038/s41593-020-0660-4, PubMed: 32601411

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59. https://doi.org/10.1080/23273798.2015.1102299, PubMed: 27135040

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123, PubMed: 20809790

Kuwada, S., Batra, R., & Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hearing Research*, *21*, 179–192. https://doi.org/10.1016/0378-5955(86)90038-9, PubMed: 3700256

Lam, N. H. L., Schoffelen, J. M., Udden, J., Hulten, A., & Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *Neuroimage*, *142*, 43–54. https://doi.org/10.1016/j.neuroimage.2016.03.007, PubMed: 26970187

Lobina, D. J., Demestre, J., & Garcia-Albea, J. E. (2018). Disentangling perceptual and psycholinguistic factors in syntactic processing: Tone monitoring via ERPs. *Behavior Research Methods*, *50*, 1125–1140. https://doi.org/10.3758/s13428-017-0932-4, PubMed: 28707215

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., et al. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17. https://doi.org/10.18637/jss.v088.i02

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703. https://doi.org/10.1037/0033-295X.101.4.676, PubMed: 7984711

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Zion Golumbic, E. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*, 7772–7781. https://doi.org/10.1523/JNEUROSCI.0168-17.2017, PubMed: 28626013

Marslen-Wilson, W., & Tyler, L. (1975). Processing structure of sentence perception. *Nature*, *257*, 784–786. https://doi.org/10.1038/257784a0, PubMed: 1186856

Oliver, G., Gullberg, M., Hellwig, F., Mitterer, H., & Indefrey, P. (2012). Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise. *Bilingualism: Language and Cognition*, *15*, 841–857. https://doi.org/10.1017/S1366728912000089

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., 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

Pefkou, M., Arnal, L., Fontolan, L., & Giraud, A.-L. (2017). θ-Band and β-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, *37*, 2882–2816. https://doi.org/10.1523/JNEUROSCI.2882-16.2017, PubMed: 28729443

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*, 1002–1044. https://doi.org/10.1037/bul0000158, PubMed: 29952584

Picton, T. W., John, M. S., Purcell, D. W., & Plourde, G. (2003). Human auditory steady-state responses: The effects of recording technique and state of arousal. *Anesthesia and Analgesia*, *97*, 1396–1402. https://doi.org/10.1213/01.ane.0000082994.22466.dd, PubMed: 14570657

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

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026, PubMed: 31103785

Pressnitzer, D., Sayles, M., Micheyl, C., & Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology*, *18*, 1124–1128. https://doi.org/10.1016/j.cub.2008.06.053, PubMed: 18656355

Rodriguez, R., Picton, T., Linden, D., Hamel, G., & Laframboise, G. (1986). Human auditory steady state responses: Effects of intensity and frequency. *Ear and Hearing*, *7*, 300–313. https://doi.org/10.1097/00003446-198610000-00003, PubMed: 3770325

Ross, B., Borgmann, C., Draganova, R., Roberts, L., & Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *Journal of the Acoustical Society of America*, *108*, 679–691. https://doi.org/10.1121/1.429600, PubMed: 10955634

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, *15*, 228–240. https://doi.org/10.1016/s0926-6410(02)00195-7, PubMed: 12527097

Schoonhoven, R., Boden, C., Verbunt, J., & De Munck, J. (2003). A whole head MEG study of the amplitude-modulation-following response: Phase coherence, group delay and dipole source analysis. *Clinical Neurophysiology*, *114*, 2096–2106. https://doi.org/10.1016/S1388-2457(03)00200-1, PubMed: 14580607

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*, 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*, 311–327. https://doi.org/10.1111/lnc3.12151

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language*, *26*, 638–657. https://doi.org/10.1016/0749-596X(87)90107-0

Uddin, S., Heald, S. L. M., Van Hedger, S. C., Klos, S., & Nusbaum, H. C. (2018). Understanding environmental sounds in sentence context. *Cognition*, *172*, 134–143. https://doi.org/10.1016/j.cognition.2017.12.009, PubMed: 29272740

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393. https://doi.org/10.3758/BF03197127, PubMed: 2381317

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition*, *19*, 95–112. https://doi.org/10.3758/BF03198500, PubMed: 2017035

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*, 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015, PubMed: 22019481

Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. (2007). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*, *16*, 1790–1796. https://doi.org/10.1093/cercor/bhj115, PubMed: 16421329

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