Cerebral Cortex, 2016;1–10

doi: 10.1093/cercor/bhw228
Original Article

ORIGINAL ARTICLE

Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations

Lars Meyer[1], Molly J. Henry[2], Phoebe Gaston[3], Noura Schmuck[4], and Angela D. Friederici[1]

[1]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04303 Leipzig, Germany, [2]Brain and Mind Institute, University of Western Ontario, Ontario, Canada N6G 1H1, [3]Department of Linguistics, University of Maryland, College Park, MD 20742-7505, USA, and [4]Department of English and Linguistics, Johannes Gutenberg University, 55128 Mainz, Germany

Address correspondence to Lars Meyer, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, 04303 Leipzig, Germany. Email: lmeyer@cbs.mpg.de

Abstract

Language comprehension requires that single words be grouped into syntactic phrases, as words in sentences are too many to memorize individually. In speech, acoustic and syntactic grouping patterns mostly align. However, when ambiguous sentences allow for alternative grouping patterns, comprehenders may form phrases that contradict speech prosody. While delta-band oscillations are known to track prosody, we hypothesized that linguistic grouping bias can modulate the interpretational impact of speech prosody in ambiguous situations, which should surface in delta-band oscillations when grouping patterns chosen by comprehenders differ from those indicated by prosody. In our auditory electroencephalography study, the interpretation of ambiguous sentences depended on whether an identical word was either followed by a prosodic boundary or not, thereby signaling the ending or continuation of the current phrase. Delta-band oscillatory phase at the critical word should reflect whether participants terminate a phrase despite a lack of acoustic boundary cues. Crossing speech prosody with participants' grouping choice, we observed a main effect of grouping choice—independent of prosody. An internal linguistic bias for grouping words into phrases can thus modulate the interpretational impact of speech prosody via delta-band oscillatory phase.

Key words: delta band, electroencephalography, language, neural oscillations, speech comprehension

Introduction

For successful speech comprehension, single words are grouped into syntactic phrases (Lewis et al. 2006; Caplan and Waters 2013; Bonhage et al. 2014), because sentences' word count often exceeds the human storage capacity (Miller 1962; Levelt 1970; Wingfield and Butterworth 1984). Speech prosody acoustically marks the boundaries of syntactic phrases through salient acoustic cues, such as pitch changes, pauses, and syllabic duration increases (Frazier et al. 2006), allowing for the bottom-up identification of phrase boundaries and facilitating phrase formation and memorization (Bower and Springston 1970; Wingfield et al. 1984; Stine and Wingfield 1987; Farrell and Lelievre 2012). Generally, salient acoustic boundaries aid the perceptual grouping of tones, digits, and words (Sturges and Martin 1974; Reeves et al. 2000).

In spite of the role of speech prosody in identifying phrase boundaries, the ability to perceive those acoustic cues that delimit syntactic phrases in speech prosody does not entail the ability to group words into syntactic phrases: electroencephalography (EEG) indicates that prelinguistic infants can well perceive prosodic boundaries based on salient acoustic cues (Männel and Friederici 2009), but utilize prosodic boundaries for phrase formation only around 3 years of age (Männel and Friederici 2011), with an adult-like phrase formation emerging only with linguistic experience around 6 years of age (Hahne et al. 2004; Oberecker et al. 2005; Oberecker and Friederici 2006; Isobe 2007; Snedeker and Yuan 2008; Männel and Friederici 2011; Männel et al. 2013; Wiedmann and Winkler 2015). These findings suggest that syntactic knowledge can exercise a linguistic influence on phrase perception during speech comprehension. Indeed, the influence of syntactic knowledge is strong enough to draw the perceived occurrence of a salient acoustic event that is experimentally displaced in time from a syntactic-phrase boundary toward the time point of the nearest syntactic-phrase boundary (Fodor and Bever 1965; Garrett et al. 1966; Buxó-Lugo and Watson 2016).

On the one hand, magnetoencephalography (MEG) work has shown that bottom-up prosody perception relies on delta-band oscillations (i.e., 04 Hz) in the human auditory cortex, which align with speech prosody through entrainment (Bourguignon et al. 2013). This is in line with the general proposal that speech perception involves phase locking of neural oscillations with linguistically distinctive acoustic speech cues on hierarchical time scales (Luo and Poeppel 2007; Ghitza and Greenberg 2009; Ghitza 2011; Ding and Simon 2012; Giraud and Poeppel 2012; Ding et al. 2015). On the other hand, there is evidence for a modulatory role of delta-band oscillations during auditory sequence processing, as these were observed to internally increase auditory-cortical sensitivity during tone (Besle et al. 2011) and speech perception (Fontolan et al. 2014; Park et al. 2015)—in the absence of additional external acoustic cues that could have led to increased sensitivity. Likewise, delta-band oscillatory phase was observed to modulate responses from macaque auditory cortex during tone perception (Lakatos et al. 2005).

Based on prior evidence for an internal modulatory role of delta-band oscillations beyond their bottom-up entrainment to acoustic boundary cues during speech perception, we asked her whether delta-band oscillations might specifically serve the preferential grouping of sentences' single words into syntactic phrases. Our auditory EEG experiment employed ambiguous sentences, which always allowed for 2 alternative patterns of grouping their single words into syntactic phrases (Fig. 1*A*,*B*). Because the 2 alternative grouping patterns are dissociated by different acoustic boundary cues, comprehenders mostly determine phrase boundaries in a bottom-up manner to arrive at an interpretation (Lehiste et al. 1976; Price et al. 1991; Wightman et al. 1992; Pynte 1996; Schafer 1997; Carlson et al. 2001; Clifton et al. 2002; Snedeker and Casserly 2010). However, depending on the language under investigation (for review, see Sedivy and Spivey-Knowlton 1994; Webman-Shafran and Fodor 2015), comprehenders also have a consistent default grouping bias for 1 of the 2 grouping alternatives (e.g., 2-phrases grouping in German; Fig. 1*B*) and may thus establish phrase boundaries internally, even if the resulting syntactic phrase contradicts the acoustic cues in the speech input (Hemforth et al. 1998; Wiedmann and Winkler 2015). We hypothesized that participants' grouping bias can modulate the impact of acoustic cues on language interpretation during the formation of syntactic phrases in an ambiguous situation through delta-band oscillatory phase.

Materials and Methods

Participants

Forty-eight right-handed (Oldfield 1971; mean lateralization quotient = 92.57, standard deviation (SD) = 10.75) native speakers of German were tested (24 females; mean age = 24.63 years, SD = 4.12 years). The high number of participants was chosen to ensure both statistical power and non-confounded stimulation (see Materials and Methods). No participant reported neurological or hearing deficits. All participants had normal or corrected-to-normal vision and were naïve as to the purpose of the study. Participation was reimbursed with €17.50.

Materials

We used ambiguous German sentences to dissociate the relationship between delta-band oscillatory phase and grouping choice from the relationship between delta-band oscillatory phase and speech prosody. From a fixed set of 40 items, which were presented in 2 conditions and 2 counterbalanced versions (i.e., 160 stimuli altogether; see below), 40 stimuli were randomly assigned to the 2 experimental conditions. In the first experimental condition (1-phrase prosody), acoustic cues at the

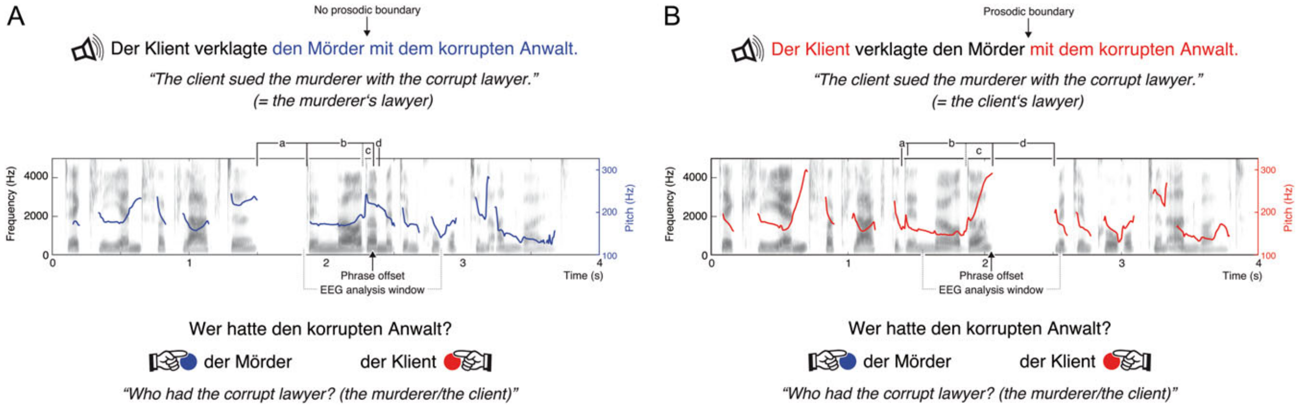

Figure 1. Overview of stimulus materials; (*A*) in 1-phrase sentences, the prepositional phrase "mit dem korrupten Anwalt (with the corrupt lawyer)" and the preceding object phrase "den Mörder (the murderer)" form the joint phrase "den Mörder mit dem korrupten Anwalt" "(the murderer with the corrupt lawyer)"; the spectrogram shows that "den Mörder" (b) is preceded by a long pause (a) and not accompanied by a lengthened final syllable (c) or a pause (d); flat pitch (c) indicates that the phrase will continue after word offset; (*B*) in 2-phrases sentences, the prepositional phrase "mit dem korrupten Anwalt (with the corrupt lawyer)" forms a separate phrase, interpreted as linking to the subject phrase "Der Klient (the client)"; the spectrogram shows that "den Mörder" (b) is not preceded by a pause (a), accompanied by a pitch rise (b) and a lengthened final syllable (c), and followed by a long pause (d), all indicating that the current phrase terminates after "Mörder"; behaviorally, participants indicated their grouping choice via button press (bottom panel).

offset of an object noun phrase (“den Mörder/the murderer”; Fig. 1A) indicated that the subsequent prepositional phrase (“mit dem korrupten Anwalt/with the corrupt lawyer”; Fig. 1A) formed a single joint syntactic phrase with the object noun phrase. In the second experimental condition (2-phrases prosody), acoustic cues at the offset of the object noun phrase indicated that the subsequent prepositional phrase formed a syntactic phrase distinct from the object noun phrase, referring to a subject noun phrase (“Der Klient/the client”; Fig. 1B). The second factor, grouping choice, resulted from participants answering a visual 2-alternative forced-choice comprehension question (“Wer hatte den korrupten Anwalt?/Who had the corrupt lawyer?”; Fig. 1, both panels) via button press for each stimulus, indicating either a 1-phrase grouping (“der Mörder/the murderer”; Fig. 1 both panels) or a 2-phrases grouping (“der Klient/the client”; Fig. 1, both panels).

We made 60 initial stimulus items in 2 counterbalanced versions each (i.e., 120 initial stimuli altogether; see below) to be piloted in a pre-study. To avoid confounding the experimental manipulation with syntactic-phrase length or across-items differences in lexical frequency (Kutas and Federmeier 2011), we matched word length and full-word and lemma frequency within sentential positions across experimental items (Biemann et al. 2004). We controlled for semantic grouping bias by matching full-word and lemma frequency of subject and object nouns across items (Hindle and Rooth 1993; MacDonald et al. 1994; Spivey-Knowlton and Sedivy 1995; Schafer 1997) and by counterbalancing subject and object nouns within item (Taraban and McClelland 1988). We then chose only those 40 items with minimal difference between the counterbalanced versions in the proportions of 1-phrase choices and 2-phrases choices (online questionnaire study; 155 participants; 90 females; mean age = 25.33 years, SD = 6.25 years; mean bias for 2-phrases choice = 10.31%, SD = 6.59%; Wales and Toner 1979; Price et al. 1991). Of each stimulus, a trained female speaker then recorded versions with 1-phrase and 2-phrases prosody in a soundproof cabin. Recordings were cut and normalized to 65 dB sound-pressure level. Onset and offset ramps of 5 ms length ruled out acoustic artifacts; average stimulus length was 3.89 s (SD = 0.25 s). Recordings with 1-phrase prosody differed from recordings with 2-phrases prosody across acoustic boundary cues (paired-samples *t*-tests; all $P < 0.001$; Table 1)—absolute pause length before and after the object phrase (Price et al. 1991; Pynte 1996), relative pause length (Snedeker and Casserly 2010), pitch slope during the phrase (Price et al. 1991; Carlson et al. 2001), and length of the phrase-final syllable (Lehiste et al. 1976; Wightman et al. 1992; Schafer 1997; Clifton et al. 2002).

The stimuli were distributed in a Latin square among 24 pseudo-randomized lists using MATLAB (The MathWorks, Inc.). Each list would contain 136 stimuli in total, including 96 additional filler items (e.g., “Nach einer Saison in der Bundesliga hat der Trainer den Stürmer gewürdigt./After a season in der German soccer league, the coach honored the striker”.; Meyer et al. 2012a, 2012b, 2013, 2014, 2015) to prevent strategy buildup. Importantly, each participant would only receive a single stimulus out of each item set to avoid a transfer of the interpretation of the relationship of the subject, object, and verb between stimuli, effectively avoiding induction of a within-items discourse-level grouping bias. For each stimulus, a comprehension question and 2 answer options were prepared. Display side for the answer options was counterbalanced within items and participants.

Procedure

Participants were seated in a dimly lit, electrically shielded, and soundproof cabin. Stimuli were presented using the Presentation software package (Neurobehavioral Systems, Inc.). Auditory stimuli were played through a pair of JBL XL 300 speakers (JBL Professional, Inc.), approximately 100 cm from each participant’s left and right front. Visual comprehension questions were shown in a proportional, sans-serif font, white letters against a gray background, on a Sony Trinitron Multiscan 300 GS monitor (Sony Corporation), approximately 70 cm from each participant’s front. A trial started with a green fixation cross of 1500 ms duration, which transitioned to red for subsequent stimulus playback. To reduce the number of blink artifacts, participants were instructed to blink during green crosses only. After playback and transition to green, the fixation cross would last for 1500 ms again, followed by a visual comprehension question to be answered within 5000 ms. The interpretation-choice options were shown underneath the comprehension question. Trials were presented in four blocks to avoid fatigue. The experiment would last for approximately 30 min (1.5 h including preparation).

Data Acquisition

Participants’ interpretation choices were recorded from a 2-button box. The EEG was recorded with a pair of BrainAmp DC amplifiers (Brain Products GmbH) from 64 Ag/AgCl scalp electrodes, mounted in an elastic cap (Electro Cap International, Inc.) according to the extended international 10–20 system. The vertical and horizontal electro-oculogram (EOG) was recorded from electrodes above and below the left eye and at the outer canthi of both eyes, respectively. The setup was online referenced to the left mastoid and grounded to the sternum. Electrode impedances were kept below 3 kΩ. The EEG and EOG were continuously recorded with a bandwidth from DC to 250 Hz at a sampling rate of 500 Hz. Individual electrode positions were determined with a Polhemus FASTRAK (Polhemus).

Table 1 Prosodic cues for 1-phrase prosody and 2-phrases prosody recordings and paired-samples *t*-statistics; object-phrase pitch rise was calculated as pitch slope between the initial and final samples of the object phrase.

Boundary cue	1-Phrase		2-Phrases		Statistic		
	Mean	SD	Mean	SD	t	df	P
First-pause duration (ms)	163	64	18	13	19.83	79	<0.001
Second-pause duration (ms)	10	14	429	62	−59.86	79	<0.001
Pause difference (ms)	−154	66	411	61	−54.69	79	<0.001
Final-syllable duration (ms)	163	63	313	70	−46.85	79	<0.001
Critical-phrase pitch rise (slope)	0.04	0.05	0.14	0.05	−14.36	79	<0.001

Data Analysis

All analyses were carried out in MATLAB (The MathWorks, Inc.). For behavioral analysis, we used methods from signal-detection theory: first, we assessed discriminability of the experimental conditions as an indicator of bottom-up processing by calculating participants' detection accuracy (i.e., d-prime). Second, we assessed grouping bias as an indicator of top-down processing by calculating participants' response bias (i.e., the criterion, c). More specifically, to verify that 1-phrase speech prosody reliably elicited 1-phrase groupings and 2-phrases speech prosody reliably elicited 2-phrases grouping, we calculated participants' detection accuracy (i.e., d-prime); we defined 1-phrase-prosody sentences as noise and 2-phrases-prosody sentences as signal (Macmillan and Creelman 2005). To verify that participants followed an internal 2-phrases grouping bias in spite of their ability to discriminate the 2-speech-prosody conditions, we calculated participants' response bias (i.e., the criterion, c). We hypothesized detection accuracy to be different from chance; we hypothesized a significant negative response bias. Due to non-normal distributions (Kolmogorov–Smirnov tests; both $d > 0.35$, both $P < 0.001$), both were assessed with a Wilcoxon sign-rank test.

For sensor-level analysis, we employed the Fieldtrip toolbox for EEG/MEG analysis (Oostenveld et al. 2011). Trials of 10 s length, time-locked to the offset of the critical object noun phrase, were extracted from the EEG, 5 s of data preceding phrase offset, 5 s of data following phrase offset, to avoid edge artifacts during filtering. Data were re-referenced offline to the average of all EEG electrodes excluding EOG and mastoid channels. To reduce slow drifts while avoiding phase shifts, we applied a sixth-order 2-pass Butterworth infinite impulse response (IIR) 0.1 Hz high-pass filter. For detection of muscle artifacts, we employed a semi-automatic approach, automatically identifying artifacts at $z = 3.5$ with a distribution-based identification approach, then rejecting artifacts based on visual inspection. On average, 16.46% (SD = 14.28%) of trials were rejected. Rejection rates did not differ between conditions (2-way ANOVA; speech prosody × grouping choice; all $F < 1.4$, all $P > 0.22$). Blinks and eye-movements were corrected using independent-components analysis (Makeig et al. 1996); to-be-corrected components were determined visually based on component topography and waveform. To derive delta-band oscillatory phase, the data were low-pass filtered with a sixth-order 2-pass Butterworth IIR 25 Hz low-pass filter, down-sampled to 100 Hz, and band-pass filtered with an optimal (Parks and McClellan 1972) 2148th-order linear-phase finite-impulse-response 0–4 Hz band-pass filter. Phase shift was corrected by an according time shift, and analytic phase and amplitude were derived through the Hilbert transform from 500 ms preceding phrase offset to 500 ms following phrase offset.

Phase data were analyzed in a factorial design, averaging across trials within participants, crossing the factors speech prosody (main effect: 1-phrase prosody vs. 2-phrases prosody, regardless of grouping choice) and grouping choice (main effect: 1-phrase grouping vs. 2-phrases grouping, regardless of speech prosody). Two participants were excluded from analysis at this step due to empty design cells resulting from artifact rejection. For statistical analysis, we implemented a cluster-permutation version of the Harrison–Kanji test (Harrison and Kanji 1988), which is appropriate for the analysis of circular data in 2-way designs (Berens et al. 2009); because of low concentration parameters in the data (i.e., $\kappa < 2$), a nonparametric version of the test was applied (i.e., an analog of the Chi-Square test for circular data; Zar 1999; Maris and Oostenveld 2007; Berens 2009). The algorithm was set to permute condition labels 5000 times to identify significant time–electrode clusters while controlling for false positives ($P < 0.05$; $\alpha = 0.05$). As we did not have a hypothesis concerning size (i.e., involving many sensors) or strength (i.e., involving sensors with strong effects) of to-be-expected clusters, we weighted both to equal extents, using the weighted cluster mass as a joint indicator (Hayasaka et al. 2004). A minimum of 3 neighboring electrodes was considered a cluster (Mellem et al. 2013; Meyer et al. 2013; Meyer et al. 2015). Because of the significant acoustic differences between the 1-phrase-prosody and 2-phrases-prosody conditions (see Materials and Methods), we hypothesized a main effect of speech prosody throughout our analysis time window. Additionally, we hypothesized participants' grouping bias (see Results) to surface as a main effect of grouping choice prior to the intonational phrase boundary.

Results

Behavioral Data

On average, there was a higher proportion of 2-phrase groupings of 2-phrases-prosody sentences (mean = 68.04%, SD = 17.19%) as compared with 1-phrase-prosody sentences (mean 54.42%, SD = 21.96%; Fig. 2A). Detection accuracy (i.e., d-prime; mean = 0.42, SD = 0.81) was significantly different from 0 ($Z = 3.43$, $P < 0.001$). This indicates acoustic discriminability of the 1-phrase-prosody and 2-phrases-prosody sentences (Fig. 2B). Response bias (i.e., the criterion:

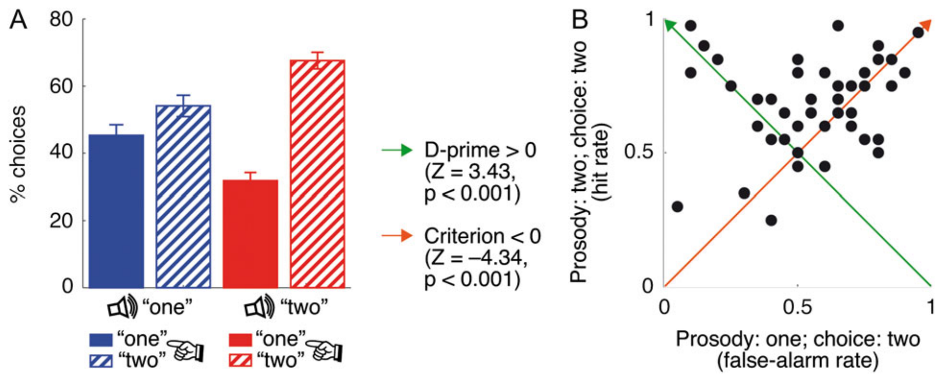

Figure 2. Behavioral results; (A) participants made significantly more 2-phrases choices to 2-phrases-prosody sentences as compared with 1-phrase sentences; (B) participants were clearly able to dissociate 1-phrase-prosody from 2-phrases-prosody sentences (i.e., most data points are above the orange line) and showed a clear segmentation bias independent of prosody toward a 2-phrases interpretation (i.e., most data points are above the green line).

$c = -\,z(\text{hit rate}) + z(\text{false alarm rate})/2$; Macmillan and Creelman 2005; mean = –0.21, SD = 0.40; Fig. 2*B*) was alsosignificantly different from 0 ($Z = -4.34$, $P < 0.001$), indicating a grouping bias toward a 2-phrases interpretation, independent of acoustic discriminability.

While significantly different from chance, the comparably low detection accuracy points to reduced discriminability of our experimental conditions. This may have resulted from our experimental control of secondary sources of interpretational bias (see Materials and Methods), such as semantic relatedness between either the subject or object and the prepositional phrase (e.g., Hindle and Rooth 1993; Schafer 1997; Snedeker and Yuan, 2008), verb-inherent bias (e.g., Grillo et al. 2015), and length differences between the subject and object (e.g., Webman-Shafran and Fodor 2015), which increase stimulus discriminability in real-life situations—yet do not reflect syntactic chunking bias. For the same reason, we take the significant difference of the response bias from 0 as an indication of syntactic chunking bias, independent of secondary disambiguating cues.

Sensor-Level Analyses

Sensor-level EEG analysis showed 2 broadly distributed significant clusters for the main effect of speech prosody (Fig. 3*B*), 1 from –0.50 to 0.33 s (weighted cluster mass = 4035.04, cluster-level $P < 0.001$, corrected; peak at electrode C3, peak-level $\chi^2(2) = 33.67$, peak-level $P < 0.001$), the other from 0.41 to 0.50 s (weighted cluster mass = 196.56, cluster-level $P < 0.05$, corrected; peak at electrode T7, peak-level $\chi^2(2) = 24.19$, peak-level $P < 0.05$). For the main effect of grouping choice, 2 focally distributed significant clusters were observed (Fig. 3*C*), the first from –0.30 s to –0.11 s over left temporo-parietal sensors (weighted cluster mass = 257.79, cluster-level $P < 0.05$, corrected; peak at electrode P3, peak-level $\chi^2(1) = 22.67$, peak-level $P < 0.05$), the second from –0.28 to –0.13 s over left-anterior sensors (weighted cluster mass = 150.42, cluster-level $P < 0.05$, corrected; peak at electrode AF7, peak-level $\chi^2(1) = 8.40$, peak-level $P < 0.05$). No interactions between speech prosody and interpretation choice were observed (Fig. 3*D*).

Sensor-Level Control Analyses

We aimed to further strengthen our interpretation that the main effect of grouping choice represents internal modulatory processing (see Discussion). To this end, we decided post hoc to test whether prosody entrainment during the critical time window for phrase continuation or termination was weaker when participants chose a 2-phrases interpretation, that is, when interpretation was in line with grouping bias. Reduced prosody entrainment would indicate a reduced reliance on acoustic cues in inferring the boundary of a syntactic phrase, in turn indicating an increased reliance on internal grouping bias. We calculated coherence between speech prosody (i.e., pitch tracks) and the single-trial band-pass-filtered EEG signal (Bourguignon et al. 2013) within conditions, within the frequency band, and across the time window of interest (i.e., 0–4 Hz, –0.5 to 0.5 s) at the peak electrodes of the sensor-level main-effect clusters

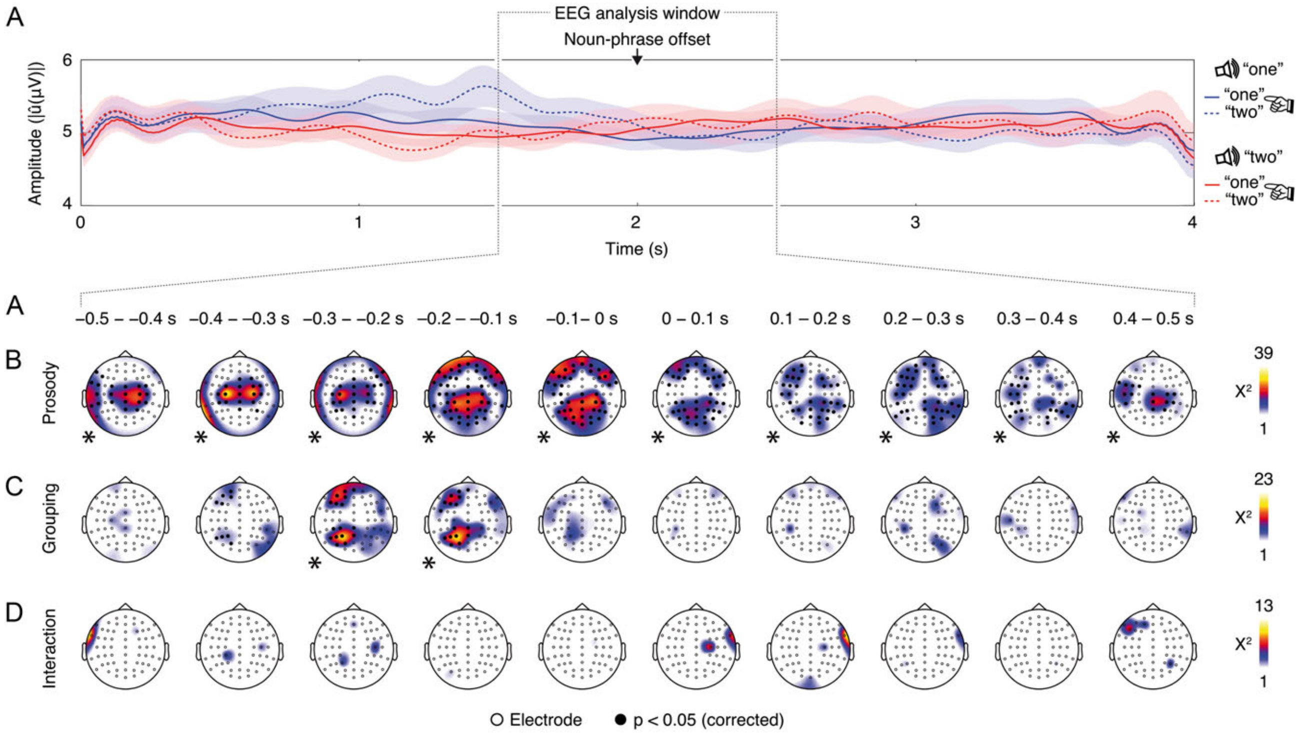

Figure 3. Sensor-level results; (*A*) delta-band amplitude envelope (i.e., absolute of band-pass-filtered and Hilbert-transformed EEG), averaged across all electrodes; blue lines mark 1-phrase-prosody sentences, red lines mark 2-phrases-prosody sentences; solid lines mark 1-phrase choices, dashed lines mark 2-phrases choices; error bands mark standard error; (*B*) statistical maps (χ^2) of significant differences between 1-phrase-prosody and 2-phrases-prosody sentences (i.e., main effect of speech prosody, irrespective of grouping choice) for the critical time window (–0.50 to 0.33 s, peak at electrode C3; 0.41–0.50 s, peak at electrode T7; $P < 0.05$, corrected); (*C*) statistical maps (χ^2) showing significant differences between 1-phrase and 2-phrases grouping choice (i.e., main effect of grouping choice, irrespective of speech prosody) for the critical time window (–0.30 to –0.11 s, peak at electrode P3; –0.28 to –0.13 s peak at electrode AF7; $P < 0.05$, corrected); (*D*) statistical maps (χ^2) of insignificant interaction between speech prosody and grouping choice for critical time window (*$P < 0.05$, corrected; 100-ms time windows chosen for illustration only).

(i.e., P3 and AF7, respectively; Fig. 4). We chose to restrict this analysis to the time window of interest, because the relevant acoustic boundary cues for inferring a phrase boundary from stimulus acoustics (i.e., bottom-up) were present during this time interval (see Table 1), and because any internal modulatory process leading to the termination of a syntactic phrase should coincide with the termination of the final word of the syntactic phrase. Single-participant coherence values were extracted for the frequency band within the delta band where coherence was maximal. To account for the unequal trial numbers across the four conditions, we corrected coherence values for distributional bias (Bokil et al. 2007; Bastos et al. 2016). We then used a Friedman test (Friedman 1937) to assess the effect of the grouping-choice factor on the coherence between pitch and delta-band signal, adjusting for the effect of the speech-prosody factor, because coherence values were non-normally distributed (Kolmogorov–Smirnov test, both $d > 0.48$, both $P < 0.001$). The effect was significant for both cluster peaks (P3: $\chi^2(1) = 4.69$, $P < 0.05$; AF7: $\chi^2(1) = 4.17$, $P < 0.05$), indicating a reduction in prosody entrainment, that is, decreased reliance on acoustic boundary cues in inferring the syntactic-phrase boundary. This may point to increased internal modulatory processing when participants chose a 2-phrases grouping, converging on their grouping bias (Fig. 4). This interpretation was corroborated by correlations between the behavioral grouping bias (i.e., the criterion, c) and the entrainment reduction (i.e., entrainment difference between the 2-phrases and 1-phrase groupings; P3: $r = 0.42$, $P < 0.005$; AF7: $r = 0.45$, $P < 0.005$; Fig. 4C); individuals with overall weaker prosody–delta coherence were more biased to choose 2-phrases groupings.

Because prior work has suggested that phase data for slow oscillatory frequencies can be confounded by the timing of sensory event-related brain potentials (ERPs) in the vicinity of the time window of interest, we controlled whether the main effect of grouping choice could have been partially driven by differences in the amplitudes of the sensory ERPs to the acoustic onset of the critical phrase (Table 1; Kayser et al. 2015). Sensory ERPs were calculated from the broadband EEG data, focusing on the main effect at the peak electrodes (i.e., P3 and AF7) during the canonical N100 and P200 time windows (N100: 50–150 ms; P200: 150–300 ms, respectively; 500 ms adjacent baseline interval). No significant main effects were found (N100 at P3: $F(1,45) = 2.22$, $P = 0.14$; N100 at AF7: $F(1,45) = 0.61$, $P = 0.44$; P200 at P3: $F(1,45) = 0.13$, $P = 0.72$; P200 at AF7 = $F(1,45) = 0.00$, $P = 0.99$). We validated this further through a circular–linear-correlation analysis (Berens 2009) between the non-significant ERP main effect and the significant phase main effect (i.e., difference scores). Correlations were non-significant (N100 at P3: $r = 0.28$, $P = 0.16$; N100 at AF7: $r = 0.27$, $P = 0.18$; P200 at P3: $r = 0.16$, $P = 0.57$; P200 at AF7: $r = 0.20$, $P = 0.40$), supporting the oscillatory nature of the observed phase main effect.

Discussion

The current results indicate that delta-band oscillatory phase during sentence processing predicts whether an internal bias

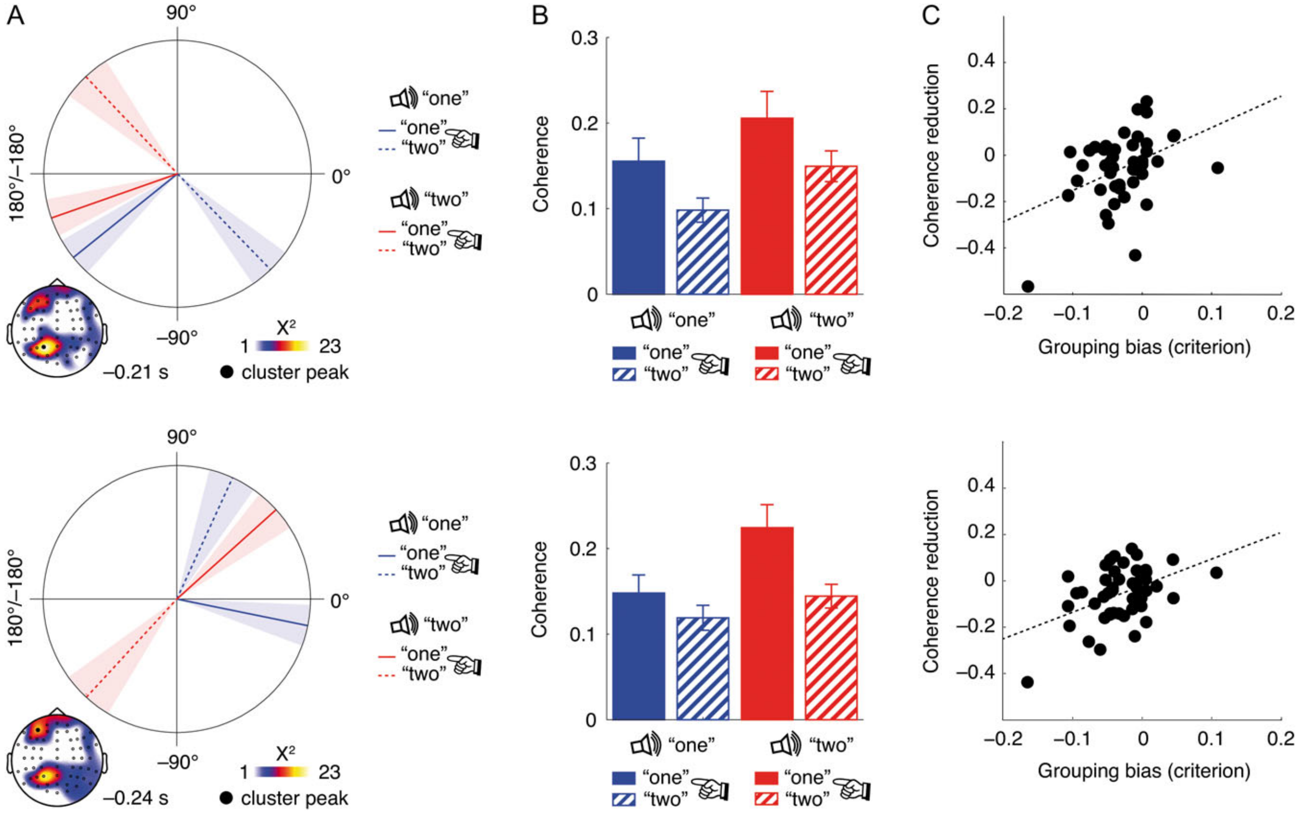

Figure 4. (*A*) Delta-band oscillatory phase at peak electrodes of sensor-level interaction clusters (top: P3; bottom: AF7); blue lines mark 1-phrase-prosody sentences, red lines mark 2-phrases-prosody sentences; solid lines mark 1-phrase choices, dashed lines mark 2-phrases groupings; error bands mark standard error; (*B*) participants showed significantly reduced prosody entrainment (i.e., coherence) when words were grouped according to syntactic bias, overriding speech prosody; (*C*) the coherence reduction was correlated with participants grouping bias, indicating that the reduced coherence during 2-phrases groupings reflects the reliance on internal modulatory mechanisms.

for grouping single words into phrases leads to a grouping pattern different from the pattern indicated bottom-up by the acoustic boundary cues of speech prosody. The behavioral data replicate the well-known grouping bias: whereas the presence of acoustic boundary cues in 2-phrases-prosody sentences results mostly in a 2-phrases grouping, the absence of acoustic boundary cues in 1-phrase-prosody sentences does not necessarily result in a 1-phrase grouping. Our EEG results provide the neural basis for this behavioral effect, strongly suggesting that the application of this internal grouping bias is driven by the state of delta-band oscillations during the point of ambiguity in the sentence: delta-band oscillatory phase predicts whether participants will follow their internal grouping bias and terminate a syntactic phrase, even in spite of a lack of acoustic cues that would warrant phrase termination.

We propose here that delta-band oscillations provide an internal linguistic searchlight for the formation of syntactic phrases during auditory language comprehension. This proposal extends recent elegant MEG work by Ding et al. (2015), who showed that delta-band oscillations are entrained at the pace of syntactic phrases—even if the boundaries of syntactic phrases are not marked by acoustic cues in speech prosody, but only by morphosyntactic cues. Critically, comprehenders showed an entrainment only when stimuli were presented in their mother tongue, but not in a foreign language—strongly suggesting that entrainment can depend on internal linguistic knowledge. While Ding et al. (2015) thus demonstrate bottom-up entrainment of delta-band oscillations to exogenous, non-acoustic temporal regularities at the phrasal rate (Ding and He 2016), we go beyond their findings in demonstrating that delta-band oscillatory phase can even induce preferred regularities in an ambiguous situation. In a hierarchical fashion, an internally formed syntactic phrase can thus modulate the interpretational impact of a phrase indicated by acoustic cues in the speech stream.

The neurocognitive dissociation of the internal induction of non-acoustic phrasal regularities from the bottom-up perception of acoustic regularities is supported by the developmental trajectory of the interpretation of ambiguous sentences (Snedeker and Yuan 2008). While EEG indicates that prelinguistic infants can well perceive prosodic boundaries (Männel and Friederici 2009), children aged 3–6 years cannot yet employ speech prosody to assign appropriate boundaries to ambiguous syntactic phrases (Choi and Mazuka 2003; Isobe 2007). From 6 years onwards, children will infer the pattern of syntactic phrases from speech prosody only, choosing 1-phrase interpretations for 1-phrase-prosody sentences, and choosing 2-phrases interpretations for 2-phrases-prosody sentences (Isobe 2007; Wiedmann and Winkler 2015). Adult speakers of most languages then have a grouping bias toward a 2-phrases interpretation, which can contradict the pattern of prosodic boundaries present in the speech input (Hemforth et al. 1998).

The proposed modulatory influence of delta-band oscillatory phase could be a functional mechanism that evaluates incoming, variable acoustic information against the preferred phrasal grain size required for optimal language comprehension. This idea is supported by corpus-linguistic and psycholinguistic work that suggests an optimal phrase length of 2 to 3 s for comprehension—within the bandwidth of delta-band oscillations. First, phrase duration in speech has a sweet spot around 2 to 3 s (Vollrath et al. 1992). When phrases exceed this duration, comprehenders are biased to terminate the current phrase and start a new one (Fodor 1998; Webman-Shafran and Fodor 2015). Second, when participants read sentences at a chosen pace, the closure positive shift (CPS), an ERP associated with the termination of syntactic phrases (Steinhauer et al. 1999; Hwang and Steinhauer 2011), appears periodically every 2 to 3 s (Roll et al. 2012; Schremm et al. 2015), and may thus be an ERP mirror image of the progression of the delta-band's oscillatory cycles. While acoustic cues can trigger a CPS (Gilbert, Boucher, & Jemel, 2015), the CPS appears also in the absence of acoustic cues (Roll et al. 2012; Schremm et al. 2015), and thus likely indexes either the internal structuring of the speech stream (Steinhauer et al. 1999; Frazier et al. 2006), or a combination of bottom-up acoustic and internal modulatory processing aspects. In either case, a relationship between the delta band, the phrase duration optimal for language comprehension, and the CPS would link classical psycholinguistic results to leading models of speech perception (Giraud and Poeppel 2012) and fundamental mechanisms of cortical information processing (Buzsaki 2006; Singer 2013; Friederici and Singer 2015). This would be an important step toward a psychophysically adequate view of language comprehension.

As an alternative interpretation, it might be possible that the observed effect of grouping choice on delta-band oscillatory phase and cortical pitch entrainment reflects reduced efficacy in auditory processing (i.e., deficient bottom-up processing) rather than reduced impact of auditory cues (i.e., modulation by internal grouping bias). This could entail that delta-band oscillations subserve 2 distinct functions during language comprehension: auditory delta would subserve the parsing of phrase-sized units from the speech stream, whereas syntactic delta would serve a modulatory purpose in applying preferential chunking patterns during internal readout (Ghitza, personal communication). Under this interpretation, grouping choice would be biased (i.e., 2-phrases grouping) when speech processing does not deliver sufficient information to determine the actual grouping pattern. It has been shown that cortical speech entrainment reduces with speech intelligibility (Peelle et al. 2013; Park et al. 2015), and that low-level auditory-processing deficits are accompanied by reduced cortical speech entrainment (Lehongre et al. 2011). With respect to the current result, however, reduced auditory efficacy would predict random grouping choices (i.e., 1-phrase grouping and 2-phrases grouping in half of trials each); yet, under reduced prosody entrainment, participants predominantly chose 2-phrases groupings instead, indicating the application of an internal, non-auditory bias.

In conclusion, we here presented evidence that delta-band oscillatory phase may provide an internal linguistic searchlight for the formation of optimal syntactic phrases during language comprehension. This extends prior evidence for an entrainment of delta-oscillations to both speech prosody and temporal linguistic regularities, arguing that internal linguistic biases for the grouping of words into syntactic phrases can modulate the impact of acoustic cues in speech prosody through delta-band oscillatory phase.

Funding

The Max Planck Society funded this research.

Notes

We are thankful to Maren Grigutsch for invaluable support with data analysis; Thomas Gunter and Björn Herrmann were very helpful in discussing the experimental paradigm. We are obliged to Ina Koch for EEG-data acquisition. *Conflict of Interest*: None declared.

References

Bastos AM, Schoffelen J-M. 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Frontiers in systems. Neuroscience. 9:175.

Berens P. 2009. CircStat: a MATLAB toolbox for circular statistics. J Stat Softw. 31:1–21.

Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE. 2011. Tuning of the human neocortex to the temporal dynamics of attended events. J Neurosci. 31:3176–3185.

Biemann C, Bordag S, Quasthoff U, Wolff C. 2004. Web services for language resources and language technology applications. Proceedings of the Fourth International Conference on Language Resources and Evaluation.

Bokil H, Purpura K, Schoffelen JM, Thomson D, Mitra P. 2007. Comparing spectra and coherences for groups of unequal size. J Neurosci Methods. 159:337–345.

Bonhage CE, Fiebach CJ, Bahlmann J, Mueller JL. 2014. Brain signature of working memory for sentence structure: enriched encoding and facilitated maintenance. J Cogn Neurosci. 26:1654–1671.

Bourguignon M, De Tiege X, de Beeck MO, Ligot N, Paquier P, Van Bogaert P, Goldman S, Hari R, Jousmaki V. 2013. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum Brain Mapp. 34:314–326.

Bower GH, Springston F. 1970. Pauses as recoding points in letter series. Exp Psychol. 83:421–430.

Buxó-Lugo A, Watson DG. 2016. Evidence for the influence of syntax on prosodic parsing. J Mem Lang. 90:1–13.

Buzsaki G. 2006.Rhythms of the brain. New York: Oxford University Press.

Caplan D, Waters G. 2013. Memory mechanisms supporting syntactic comprehension. Psychon Bull Rev. 20:243–268.

Carlson K, Clifton C, Frazier L. 2001. Prosodic boundaries in adjunct attachment. J Mem Lang. 45:58–81.

Choi Y, Mazuka R. 2003. Young children's use of prosody in sentence parsing. J Psycholinguist Res. 32:197–217.

Clifton C, Carlson K, Frazier L. 2002. Informative prosodic boundaries. Lang Speech. 45:87–114.

Ding N, He H. 2016. Rhythm of silence. Trends Cogn Sci. 20:82–84.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D. 2015. Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci. 19:158–164.

Ding N, Simon JZ. 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol. 107:78–89.

Farrell S, Lelievre A. 2012. The dynamics of access to groups in working memory. J Exp Psychol Learn Mem Cogn. 38:1659–1674.

Fodor JA, Bever TG. 1965. The psychological reality of linguistic segments. J Verbal Learning Verbal Behav. 4:414–420.

Fodor JD. 1998. Learning to parse? J Psycholinguist Res. 27:285–319.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL. 2014. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun. 5:4694.

Frazier L, Carlson K, Clifton C. 2006. Prosodic phrasing is central to language comprehension. Trends Cogn Sci. 10:244–249.

Friederici A. 2011. The brain basis of language processing: from structure to function. Physiol Rev. 91:1357–1392.

Friederici AD, Singer W. 2015. Grounding language processing on basic neurophysiological principles. Trends Cogn Sci. 6:329–338.

Friedman M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc. 32:675–701.

Garrett M, Bever T, Fodor J. 1966. The active use of grammar in speech perception. Percept Psychophys. 1:30–32.

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.

Ghitza O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychol. 2:130.

Giraud AL, Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 15:511–517.

Gilbert AC, Boucher VJ, Jemel B. 2015. The perceptual chunking of speech: A demonstration using ERPs. Brain Res. 1603:101–113.

Grillo N, Costa J, Fernandes B, Santi A. 2015. Highs and lows in english attachment. Cognition. 144:116–122.

Hahne A, Eckstein K, Friederici AD. 2004. Brain signatures of syntactic and semantic processes during children's language development. J Cogn Neurosci. 16:1302–1318.

Harrison D, Kanji G. 1988. The development of analysis of variance for circular data. J Appl Stat. 15:197–223.

Hayasaka S, Phan KL, Liberzon I, Worsley KJ, Nichols TE. 2004. Nonstationary cluster-size inference with random field and permutation methods. Neuroimage. 22:676–687.

Hemforth B, Konieczny L, Scheepers C, Strube G. 1998. Syntactic ambiguity resolution in German. In: Hillert D, editors. Sentence Processing: A Crosslinguistic Perspective. Bingley: Emerald, p. 293–312.

Hindle D, Rooth M. 1993. Structural ambiguity and lexical relations. Comput Ling. 19:103–120.

Hwang H, Steinhauer K. 2011. Phrase length matters: the interplay between implicit prosody and syntax in Korean "garden path" sentences. J Cogn Neurosci. 23:3555–3575.

Isobe M. 2007. The acquisition of nominal compounding in Japanese: A Parametric Approach. Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA).

Kayser SJ, Ince RA, Gross J, Kayser C. 2015. Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. J Neurosci. 35:14691–14701.

Kutas M, Federmeier KD. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol. 62:621–647.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol. 94:1904–1911.

Lehiste I, Olive JP, Streeter LA. 1976. Role of duration in disambiguating syntactically ambiguous sentences. J Acoust Soc Am. 60:1199–1202.

Lehongre K, Ramus F, Villiermet N, Schwartz D, Giraud AL. 2011. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. Neuron. 72:1080–1090.

Levelt W. 1970. Hierarchial chunking in sentence processing. Percept Psychophys. 8:99–103.

Lewis R, Vasishth S, Van Dyke J. 2006. Computational principles of working memory in sentence comprehension. Trends Cogn Sci. 10:447–454.

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

MacDonald MC, Pearlmutter NJ, Seidenberg MS. 1994. The lexical nature of syntactic ambiguity resolution. Psychol Rev. 101:676.

Macmillan N, Creelman C. 2005.Detection theory: a user's guide. Mahwan: Lawrence Erlbaum Associates.

Makeig S, Bell AJ, Jung T-P, Sejnowski TJ. 1996. Independent component analysis of electroencephalographic data. In: Touretzky DS, Mozer MC, Hasselmo ME, editors. Advances in neural information processing systems 8. Cambridge: MIT Press, p. 145–151.

Männel C, Friederici AD. 2009. Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults. J Cogn Neurosci. 21:1988–2006.

Männel C, Friederici AD. 2011. Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. Dev Sci. 14:786–798.

Männel C, Schipke CS, Friederici AD. 2013. The role of pause as a prosodic boundary marker: language ERP studies in German 3-and 6-year-olds. Dev Cogn Neurosci. 5:86–94.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods. 164:177–190.

Mellem MS, Friedman RB, Medvedev AV. 2013. Gamma- and theta-band synchronization during semantic priming reflect local and long-range lexical-semantic networks. Brain Lang. 127:440–451.

Meyer L, Cunitz K, Obleser J, Friederici A. 2014. Sentence processing and verbal working memory in a white-matter-disconnection patient. Neuropsychologia. 61:190–196.

Meyer L, Grigutsch M, Schmuck N, Gaston P, Friederici A. 2015. Frontal–posterior theta oscillations reflect memory retrieval during sentence comprehension. Cortex. 71:205–218.

Meyer L, Obleser J, Anwander A, Friederici AD. 2012a. Linking ordering in Broca's area to storage in left temporo-parietal regions: the case of sentence processing. Neuroimage. 62:1987–1998.

Meyer L, Obleser J, Friederici A. 2013. Left parietal alpha enhancement during working memory-intensive sentence processing. Cortex. 49:711–721.

Meyer L, Obleser J, Kiebel S, Friederici A. 2012b. Spatiotemporal dynamics of argument retrieval and reordering: an fMRI and EEG study on sentence processing. Front Psychol. 3:523.

Miller G. 1962. Decision units in the perception of speech. IRE Trans Inform Theory. 8:81–83.

Oberecker R, Friederici AD. 2006. Syntactic event-related potential components in 24-month-olds' sentence comprehension. Neuroreport. 17:1017–1021.

Oberecker R, Friedrich M, Friederici A. 2005. Neural correlates of syntactic processing in two-year-olds. J Cogn Neurosci. 17:1667–1678.

Oldfield R. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.

Oostenveld R, Fries P, Maris E, Schoffelen J. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci . 2011:1.

Park H, Ince RA, Schyns PG, Thut G, Gross J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol. 25:1649–1653.

Parks T, McClellan J. 1972. Chebyshev approximation for nonrecursive digital filters with linear phase. IEEE Trans Circuit Theory. 19:189–194.

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

Price P, Ostendorf M, Shattuck-Hufnagel S, Fong S. 1991. The use of prosody in syntactic disambiguation. J Acoust Soc Am. 90:2956–2970.

Pynte J. 1996. Prosodic breaks and attachment decisions in sentence parsing. Lang Cogn Process. 11:165–192.

Reeves C, Schmauder AR, Morris RK. 2000. Stress grouping improves performance on an immediate serial list recall task. J Exp Psychol Learn Mem Cogn. 26:1638.

Roll M, Lindgren M, Alter K, Horne M. 2012. Time-driven effects on parsing during reading. Brain Lang. 121:267–272.

Schafer A. 1997.Prosodic parsing: the role of prosody in sentence comprehension. Amherst: University of Massachusetts.

Schremm A, Horne M, Roll M. 2015. Brain responses to syntax constrained by time-driven implicit prosodic phrases. J Neurolinguistics. 35:68–84.

Sedivy J, Spivey-Knowlton M. 1994. The use of structural, lexical, and pragmatic information in parsing attachment ambiguities. In: Clifton C, Frazier L, Rayner K, editor. Perspectives on sentence processing. Hove: Psychology Press.

Singer W. 2013. Cortical dynamics revisited. Trends Cogn Sci. 17:616–626.

Snedeker J, Casserly E. 2010. Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. Lang Cogn Process. 25:1234–1264.

Snedeker J, Yuan S. 2008. Effects of prosodic and lexical constraints on parsing in young children (and adults). J Mem Lang. 58:574–608.

Spivey-Knowlton M, Sedivy JC. 1995. Resolving attachment ambiguities with multiple constraints. Cognition. 55:227–267.

Steinhauer K, Alter K, Friederici AD. 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci. 2:191–196.

Stine EL, Wingfield A. 1987. Process and strategy in memory for speech among younger and older adults. Psychol Aging. 2:272.

Sturges PT, Martin JG. 1974. Rhythmic structure in auditory temporal pattern perception and immediate memory. J Exp Psychol. 102:377–383.

Taraban R, McClelland JL. 1988. Constituent attachment and thematic role assignment in sentence processing: influences of content-based expectations. J Mem Lang. 27:597–632.

Vollrath M, Kazenwadel J, Krüger H-P. 1992. A universal constant in temporal segmentation of human speech. Naturwissenschaften. 79:479–480.

Wales R, Toner M. 1979. Intonation and ambiguity. In: Cooper WE, editor. Sentence processing: sentence processing: psycholinguistic studies presented to Merrill Garrett. Hillsdale: Erlbaum.

Webman-Shafran R, Fodor JD. 2015. Phrase length and prosody in on-line ambiguity resolution. J Psycholinguist Res. 45:447–474.

Wiedmann N, Winkler S. 2015. The influence of prosody on children's processing of ambiguous sentences. In: Winkler

S, editor. Ambiguity: Language and communication. Berlin: De Gruyter, p. 185–197.

Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. J Acoust Soc Am. 91:1707–1717.

Wingfield A, Butterworth B. 1984. Running memory for sentences and parts of sentences: syntactic parsing as a control function in working memory. In: Bouma H, Bouwhuis D, editor. Attention and performance X: control of language processes. London: Lawrence Erlbaum Associates. p. 351–364.

Wingfield A, Lombardi L, Sokol S. 1984. Prosodic features and the intelligibility of accelerated speech: syntactic versus periodic segmentation. J Speech Lang Hear Res. 27:128–134.

Zar JH. 1999. Biostatistical analysis, 4 ed. Upper Saddle River: Prentice Hall.