UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Event-related responses reflect chunk boundaries in natural speech

Anurova, Irina; Vetchinnikova, Svetlana; Dobrego, Aleksandra; Williams, Nitin; Mikusova, Nina ...

2022-07-15
Academic Press Inc.
http://hdl.handle.net/10138/344509
Anurova, I, Vetchinnikova, S, Dobrego, A, Williams, N, Mikusova, N, Suni, A, Mauranen, A & Palva, S 2022, 'Event-related responses reflect chunk boundaries in natural speech', NeuroImage, vol. 255, 119203. https://doi.org/10.1016/j.neuroimage.2022.119203

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

NeuroImage 255 (2022) 119203



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Event-related responses reflect chunk boundaries in natural speech



Irina Anurova[a,b,*], Svetlana Vetchinnikova[c], Aleksandra Dobrego[c], Nitin Williams[a,c], Nina Mikusova[c], Antti Suni[c], Anna Mauranen[c,1], Satu Palva[a,d,1,*]

[a] Helsinki Institute of Life Sciences, Neuroscience Center, University of Helsinki, Finland
[b] BioMag Laboratory, HUS Medical Imaging Center, Helsinki, Finland
[c] Department of Languages, University of Helsinki, Finland
[d] Centre for Cognitive Neuroscience, Institute of Neuroscience and Psychology, University of Glasgow, United Kingdom

ARTICLE INFO

Keywords:
Natural speech
Chunking
Interruptions
MEG
EEG
Emitted potential
Closure positive shift

ABSTRACT

Chunking language has been proposed to be vital for comprehension enabling the extraction of meaning from a continuous stream of speech. However, neurocognitive mechanisms of chunking are poorly understood. The present study investigated neural correlates of chunk boundaries intuitively identified by listeners in natural speech drawn from linguistic corpora using magneto- and electroencephalography (MEEG). In a behavioral experiment, subjects marked chunk boundaries in the excerpts intuitively, which revealed highly consistent chunk boundary markings across the subjects. We next recorded brain activity to investigate whether chunk boundaries with high and medium agreement rates elicit distinct evoked responses compared to non-boundaries. Pauses placed at chunk boundaries elicited a closure positive shift with the sources over bilateral auditory cortices. In contrast, pauses placed within a chunk were perceived as interruptions and elicited a biphasic emitted potential with sources located in the bilateral primary and non-primary auditory areas with right-hemispheric dominance, and in the right inferior frontal cortex. Furthermore, pauses placed at stronger boundaries elicited earlier and more prominent activation over the left hemisphere suggesting that brain responses to chunk boundaries of natural speech can be modulated by the relative strength of different linguistic cues, such as syntactic structure and prosody.

1. Introduction

Speech unfolds as a continuous flow of acoustic information where extraction of meaning has been proposed in theoretical models to be achieved by chunking speech into smaller units (Christiansen and Chater, 2016; Sinclair and Mauranen, 2006). Chunking language is understood to be an automatized, integrated, and multi-level process (e.g., Bonhage et al., 2017). A considerable body of research involving short, controlled stimuli suggests that segmentation of oral speech is associated with neural correlates at multiple timescales simultaneously (Doelling et al., 2014; Ghitza and Greenberg, 2009; Giraud and Poeppel, 2012; Gross et al., 2013; Henke and Meyer, 2021; Peelle et al., 2013). For example, segmentation at the syllabic rate has been demonstrated with cortical entrainment in the theta band (Peelle et al., 2013) while segmentation at the phrasal level has been associated with delta (1–4 Hz) band rhytmicity (Ding et al., 2016). However, little is known about segmentation of continuous spontaneous speech. Interest in natural speech is nevertheless growing in neuroscience research, extending to multi-level units, narratives, and conversations (Blanco-Elorrieta and Pylkkänen, 2017; Blank and Fedorenko, 2017; Brennan et al., 2012; Chai et al., 2016; Gross et al., 2013; Kauppi et al., 2017; Keitel et al., 2018; Lerner et al., 2011; Nguyen et al., 2019; Saalasti et al., 2019; Silbert et al., 2014; Simony et al., 2016; Willems et al., 2016). Artificially constructed stimuli are beset with the problem of ecological validity and it is therefore not clear to what extent the findings are generalizable to naturally occurring speech. The syntax of natural, spontaneous speech differs from traditional sentence grammars, which are based on written text, to the extent that normal, ordinary speech is believed to require its own grammar (Brazil et al., 1995; Kaltenböck et al., 2011; Leech, 2000; Ono and Thompson, 1995). For example, over 35% of the units of conversational speech can be considered as consisting of non-clausal material, that is, stretches of speech which do not fit the definition of a clause (Biber et al., 1999). Furthermore, syntactic structures in natural spoken language do not seem to have clear periodicity. Therefore, to overcome these problems, Ono and Thompson (1995) suggested that spoken language should be studied in terms of prosodic phrases. Prosodic

* Corresponding authors at: University of Helsinki, P.O. Box 21, FI-00014, Finland.
E-mail addresses: anurova.irina@gmail.com (I. Anurova), satu.palva@helsinki.fi (S. Palva).
[1] These authors contributed equally.

https://doi.org/10.1016/j.neuroimage.2022.119203.
Received 1 October 2021; Received in revised form 22 March 2022; Accepted 8 April 2022
Available online 9 April 2022.
1053-8119/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

units in natural speech were shown to occur approximately once per second in different languages (Inbar et al., 2020) and form a consistent low-frequency rhythm. A more detailed analysis employing Tones and Break Indices (ToBI) annotation system, which allows distinguishing between different types of hierarchically organized prosodic units, revealed that sequences of 2 to 5 intermediate intonation phrases are periodic at 0.8 to 1.6 Hz within a longer and more distinct superordinate intonation phrase (Stehwien and Meyer, 2021). Such amplitude modulations of the speech envelope were suggested to be sufficiently regular for neural tracking (Stehwien and Meyer, 2021). Indeed, brain oscillatory activity was shown to be synchronized to *prosodic chunks* of non-grammatical digit strings providing evidence for acoustically-driven speech segmentation in the delta range (Ghitza, 2020; Rimmele et al., 2021). In compliance with this, the benefit of acoustic prosodic segmentation was demonstrated in a behavioral study employing an auditory working memory task (Ghitza, 2017). Furthermore, prosody, available at an early point in a sentence, has been shown to allow listeners to predict eventual syntactic structure during online sentence processing (Beach, 1991; Schafer et al., 2000). These studies support the hypothesis that prosodic phrasing is central to spoken language comprehension and may guide the cognitive formation of syntactic and semantic units (Cutler et al., 1997; Frazier et al., 2006; Stehwien and Meyer, 2021). Moreover, prosodic chunking has been proposed to allow the effective decoding of a single coherent piece of information per prosodic unit without exceeding working memory capacity (Inbar et al., 2020; Stehwien and Meyer, 2021).

At the neuronal level, it has been found that the closure of a prosodic phrase elicits a distinct event-related potential (ERP) component – the closure positive shift (CPS) starting around or even before the onset of a pause separating consecutive prosodic phrases (Bögels et al., 2011; Steinhauer, 2003). Importantly, when a pause is removed while other indicators of a prosodic boundary, such as prefinal lengthening and boundary tone, are kept intact, the CPS is still present (Steinhauer et al., 1999). This component has also been observed in responses to breaks separating musical phrases (Knösche et al., 2005), and to intonational phrases in pseudoword and hummed sentences (Pannekamp et al., 2005), which suggests an important role for low level acoustic-phonetic cues in prosodic chunking. On the other hand, the CPS has been observed in responses to commas in reading (Steinhauer, 2003; Steinhauer and Friederici, 2001), and even in the absence of a comma after long syntactic constituents (Hwang and Steinhauer, 2011). Furthermore, the CPS can reflect the interaction between prosody and other cues such as context and syntax. For example, the CPS elicited by acoustically identical prosodic boundaries has been found to be modulated by contextual predictability of a boundary (Kerkhofs et al., 2007). The CPS has also been observed in the absence of an overt prosodic boundary if the boundary is predictable on syntactic grounds (Itzhak et al., 2010). In all, these studies suggest that the CPS is not driven exclusively by bottom-up acoustic information, but rather reflects more abstract phrasing based on the integration of several linguistic cues (Bögels et al., 2011).

The importance of abstract linguistic information such as syntactic structure and semantics in segmenting continuous speech has also been demonstrated in several behavioral studies and in studies of oscillatory activity. In a boundary detection task, where the strength of prosodic cues was thoroughly controlled, the probability of boundary marking was higher for syntactically licensed compared to non-licensed locations in spoken sentences (Buxó-Lugo and Watson, 2016). At the neurophysiological level, the tracking of perceptually relevant speech is associated with the entrainment of neuronal oscillations in different frequency bands, whereby different oscillatory frequencies would track different hierarchical linguistic units (Ding et al., 2017; Doelling et al., 2014; Kaufeld et al., 2020; Keitel et al., 2018; Teng et al., 2018). Importantly, neural tracking of abstract linguistic structures may be observed when prosody is not present in the stimuli. For example, synchronization of delta activity has been found to different types of *non-prosodic chunks*, such as phrases and sentences (Ding et al., 2016) or word pairs constructed on the basis of abstract rules (Jin et al., 2020). These findings suggest that speech segmentation could be achieved by neural networks including auditory, motor and association cortex through neuronal oscillations which allow parsing the sound input at separate timescales and form Linguistic Trees (Morillon et al., 2019; Poeppel and Assaneo, 2020). However, enhanced neural tracking of the speech envelope at the phrasal timescale for naturally spoken sentences compared to jabberwocky controls with morphemes and sentential prosody suggests the impact of both acoustically-driven bottom-up and contextually-invoked top-down processing on spoken language segmentation (Kaufeld et al., 2020).

The present study was designed to answer two questions: how continuous natural speech is intuitively chunked up by listeners, and to what extent, if any, it is possible to find neurocognitive correlates to intuitively perceived chunks. We set up two experiments, one at a behavioral level and another using magneto- and electroencephalography (MEEG). In contrast to previous studies, we did not construct the stimuli according to any pre-defined notion of a chunk, such as prosodically or syntactically driven unit. Instead, following Sinclair and Mauranen (2006), we hypothesized that listeners who are fluent in the language spontaneously identify chunk boundaries in real-time speech; therefore, a 'chunk' can be defined as an intuitive unit reliably identified by naïve listeners. In this sense it can be regarded as a pre-theoretical notion. We further hypothesized that such a unit is unlikely to be driven by just one type of linguistic cue, since language processing is holistic and based on the simultaneous integration of all available information: prosodic, semantic, syntactic and even sociolinguistic (Hanulíková et al., 2012; Van Berkum et al., 2008).

Naturalistic speech stimuli comprised of short excerpts of speech events were extracted from linguistic corpora consisting of both a soundtrack and its transcript. In the behavioral experiment, we sought to establish the degree of convergence among naïve listeners with regard to chunk boundaries by evaluating the consistency of intuitive chunk boundary marking across participants. We examined the effect of pause length, prosody, and syntactic structure on chunk boundary perception. We then used MEEG to study neurocognitive mechanisms in relation to intuitive chunking based on the behavioral experiment by recording evoked responses to pauses inserted (1) at chunk boundaries with high and medium agreement rates across participants and (2) at non-boundaries, i.e., at locations where participants did not mark a boundary. In previous research, omission or delay of final words or word fragments had been found to elicit either a biphasic negative-positive emitted potential (EP) (Bendixen et al., 2014; Besson et al., 1997; Mattys et al., 2005) or a monophasic positive response (Nakano et al., 2014), which were suggested to reflect a conceptual surprise and a reanalysis of the syntactic anomaly.

We hypothesized that silent pauses inserted at intuitive chunk boundary locations and at non-boundaries would elicit different event-related (ER) activity. More specifically, we expected that silent pauses inserted at intuitive chunk boundary locations would be associated with boundary-specific responses, while silent pauses inserted at non-boundaries would be perceived as interruptions. However, since information from top-down and bottom-up sources had been shown to be integrated in a highly interactive way at different levels of language organization, we also expected to detect a modulatory effect of syntactic structure on boundary-related brain activity.

2. Materials and methods

2.1. Overview of the approach

This study was comprised of two parts. First, a behavioural experiment was conducted. In the experiment, the participants simultaneously listened to speech extracts and followed their transcripts on tablet computers, while intuitively marking chunk boundaries in excerpts of natural speech. Based on these data, we estimated chunk boundaries

agreement rates across participants. Next, we used the identified chunk boundaries to study their neuronal correlates using combined magneto- and electroencephalography (MEEG). In this experiment, we inserted silent pauses of 2 s at the locations of chunk boundaries as well as non-boundaries and estimated ERs to these silent pauses.

2.2. Participants

For the behavioral experiment, we obtained data from 104 neurologically healthy volunteers that were students of the University of Helsinki with no background in linguistics, aged 20–39 (71 females, 31 males, 2 other; 94 right-handed, 5 left-handed, 5 ambidextrous). All were fluent non-native speakers of English, and none reported dyslexia. The MEEG data and anatomical magnetic resonance images (MRIs) were acquired from 20 volunteers (mean age ± standard deviation (S.D.): 30 ± 6.7 years, 12 males, 18 right-handed and 2 left-handed). The participants in the behavioral and the MEEG studies did not overlap. All participants were healthy, with no history of neurological or psychiatric disorders including dyslexia, and proficient in English as measured with the Elicited Imitation Task (Culbertson et al., 2020; Yan et al., 2016).

In both studies, all participants were non-native, although proficient English speakers. In the behavioral study, we recruited subjects of 22 different native languages: Finnish speakers were a prevalent group of 62, followed by Spanish (8), Chinese (5), Arabic (3,) German (3) and others. In the MEEG study, the participants represented 7 different native languages: Russian (9), Finnish (5), Tamil (2), Danish, Turkish, German, and Ukrainian.

The study was performed according to Declaration of Helsinki and approved by an ethical committee of the Helsinki University Central Hospital. Written informed consent was obtained from all participants prior to the experiment.

2.3. Speech stimuli

The speech stimuli were created using three corpora of authentic speech recorded in university environments: the Michigan Corpus of Academic Spoken English (MICASE), the Corpus of English as a Lingua Franca in Academic Settings (ELFA) and the Vienna-Oxford International Corpus of English (VOICE). All three corpora comprise native and non-native speech from fluent English speakers of different backgrounds, accents, and cultures, recorded in a natural environment. We selected 195 10-45-second-long semantically coherent and grammatically well-formed extracts from speech events typical of academic communication, such as lectures, seminars, conference presentations and discussions. The extracts were meaningful on their own outside the larger context of a speech event, and did not contain unintelligible or unfinished words, laughter, long pauses, overlapping speech, speaker changes, frequent hesitations, or repetitions. To ensure comprehensibility, we also controlled for specialized and low-frequency vocabulary. The extracts were reproduced by a trained speaker, who mimicked the original intonation patterns with high precision. The speaker was bilingual, with English as one native language. Recordings took place in an acoustically shielded studio at the phonetics laboratory, University of Helsinki, Finland.

2.4. Linguistic annotation of speech stimuli

All boundaries between two consecutive words were annotated for pause length, prosodic boundary strength and clausal syntactic structure. Pause annotation was carried out using WebMAUS (Schiel, 1999), which automatically aligns audio recording to its transcript, and Praat (Boersma and Weenink, 2017). Prosodic boundary strength was estimated with the Wavelet Prosody Toolkit, an unsupervised system which performs continuous wavelet analysis (CWT) based on fundamental frequency, energy envelope and word duration (excluding pauses and breaths) and finds prosodic boundaries (Suni, 2017; Suni et al., 2017). Specifically, the method checks each word boundary for energy and fundamental frequency minima, as well as minima of differences between durations of adjacent words (pre-boundary lengthening). Prosodic boundary was defined by tracking minima across all those scales in the resulting scalograms.

Applying CWT allows for examining both local and wider context of the prosodic signals, which has shown to be beneficial, yielding good agreement between CWT method and expert annotations of ToBi break indices (Suni et al., 2017). The depth of the minima across wavelet scales is taken into account, resulting in a continuous boundary strength score.

In syntactic annotation we defined a clause as a constituent structured around a verb phrase and included both finite and non-finite clauses in the definition. Each clause has a clausal boundary at the beginning and at the end and no clausal boundaries within it. This annotation was mapped on a scale from 1 to 4 to reflect syntactic boundary strength from the perspective of constituent structure which is hierarchical in nature and posits part-whole relationships between smaller and larger constituents (see e.g., Carnie, 2010; Huddleston and Pullum, 2002). Thus, word boundaries where one clause ends and the next one begins were assumed to be syntactically the strongest and assigned a value of 4 (clausal/clausal or C/C), places where a clause ends but a new one does not start immediately a value of 3 (clausal/non-clausal or C/NC), places where a new clause starts but the clause in which it is embedded is not yet finished, a value of 2 (non-clausal/clausal or NC/C) and places without a single clausal boundary a value of 1 (non-clausal/non-clausal or NC/NC). Fig. 1 illustrates an example of syntactic annotation.

In addition, we evaluated the mean duration of all intuitively defined chunks separated by statistically significant boundaries by measuring intervals (in seconds) from the onset of a preceding chunk to the onset of a consecutive chunk, so that a natural gap separating the chunks was included into the earlier one.

2.5. Behavioral chunking experiment

To collect data on intuitively perceived locations of chunk boundaries, we used a custom web-based tablet application *ChunkitApp* (Vetchinnikova et al., 2017). The application displays transcripts of audio recordings at the same time as they play in participants' headphones (Fig. 2, A). The participants were asked to listen to the recordings and mark boundaries between chunks as they felt appropriate by tapping interactive tilde symbols (~) in real time. Concurrent listening and access to the transcripts was selected to avoid problems due to the fast rate of spontaneous speech that requires attention to the speech signal which would compromise the concurrent marking. The *ChunkitApp* (Vetchinnikova et al., 2017) approach enables pausing the soundtrack, but with disappearance of text simultaneously. It is important to note that the transcript did not convey a normal reading experience, since every word was interleaved with a tilde symbol and due to concurrent attention to speech signal. The trial comprised of 26–109 words (54 ± 15, mean ± S.D.). The experiment comprised of 98 trials in Set 1 and 97 trials in Set 2 where each stimulus was presented once.

2.6. Testing agreement rates for statistical significance

The participants were free to mark any word boundary as a chunk boundary. To determine the statistical significance of agreement rates for each word boundary, we used permutation statistics. We permuted marked and unmarked boundaries within each individual participant 1M times. Then, in each permutation we calculated the total number of boundary markings for each word boundary which gave us 1M permuted agreement rates for each word boundary. To estimate the probability that the observed agreement rate occurred by chance, we calculated the number of times when the observed or a higher agreement rate occurred across 1M permutations as well as the number of times when the observed or a lower agreement rate occurred across 1M permutations, divided by 1M. To avoid zero p-values in

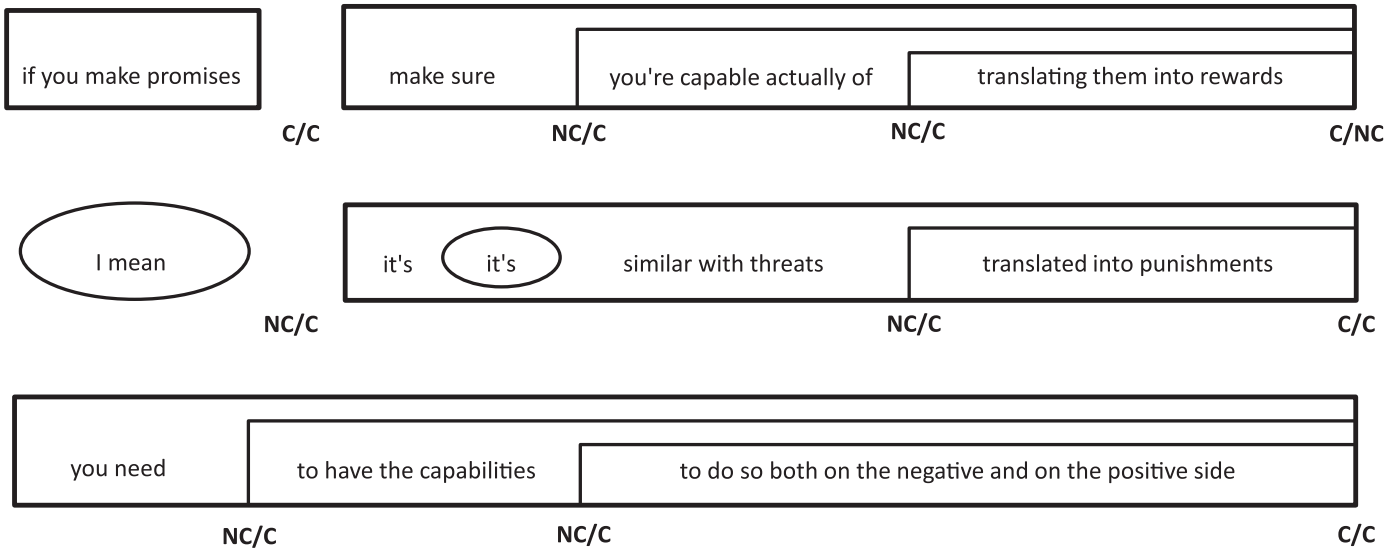


Fig. 1. An example of syntactic annotation. The extract starts with a dependent *if*-clause followed by the main clause which embeds a *that*-clause containing an embedded non-finite *ing*-clause. All clauses are located within rectangles. Non-clausal material, that is, material which is not syntactically integrated into either the preceding or the following clause, such as a repetition (*it's*) and a discourse marker (*I mean*) is enclosed in ovals. Boundaries annotated as C/C, NC/C, or C/NC are marked in the extract, all other boundaries between two consecutive words are NC/NC.

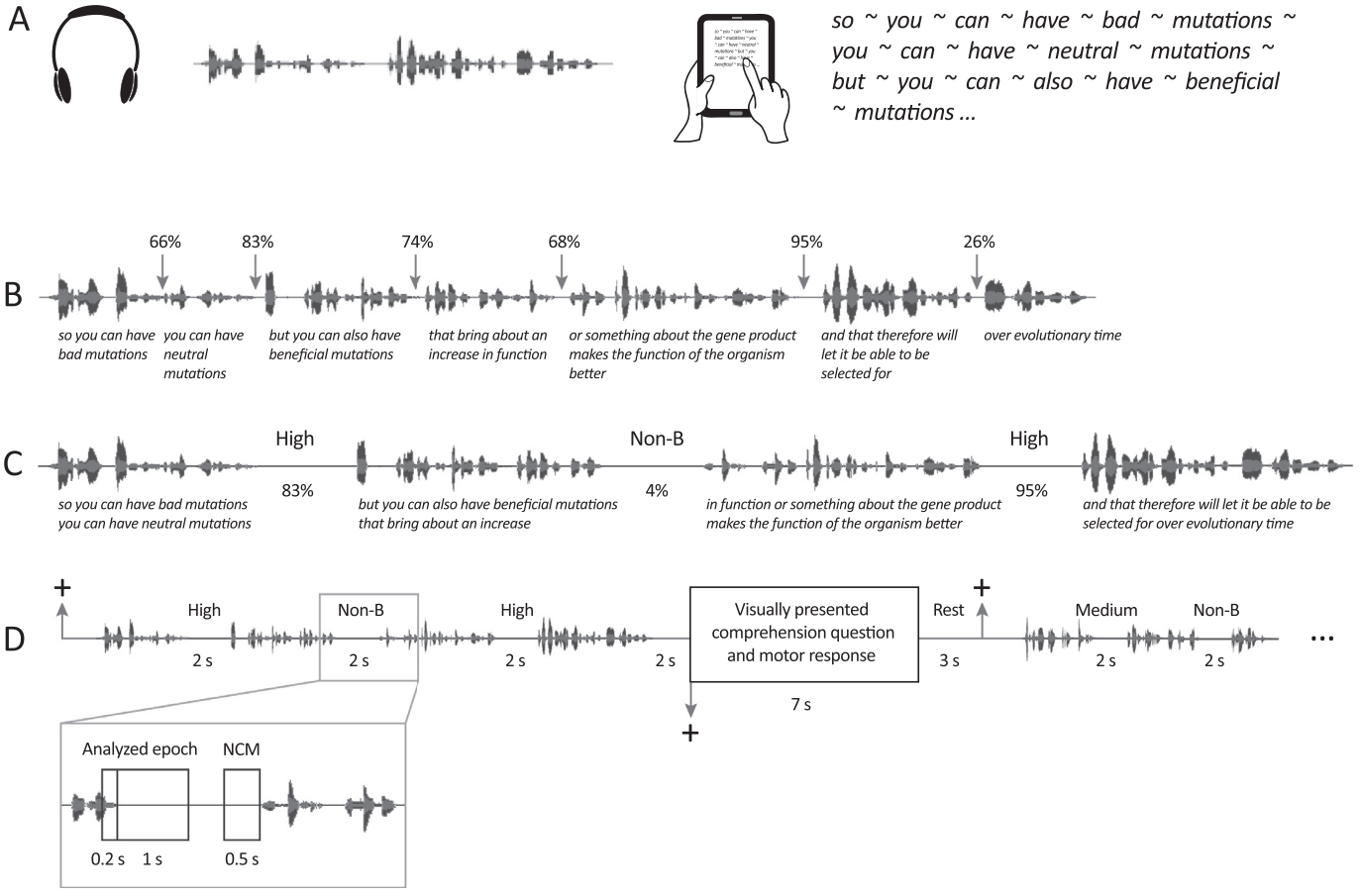


Fig. 2. Examples of stimuli and experimental paradigm in the behavioral (A-B) and MEEG experiments (C-D). **A**. The schematic diagram of simultaneously presented auditory stimulus and its transcript. **B**. An example of an original stimulus used in the behavioral experiment (for more examples see Appendix 1). Arrows point to the locations of all significant boundaries identified in the behavioral experiment. The values above the arrows show agreement rate on a chunk boundary (the percent of subjects who marked a boundary in a given space between two orthographic words). **C**. An example of a stimulus used in the electrophysiological experiment. The 2-second silent pauses were inserted at the locations selected on the basis of the behavioral results. "High" and "Non-B" (Non-boundary) indicate experimental conditions for the present excerpt, the values below – agreement rate of a chunk boundary. **D.** The schematic diagram of the experimental design. "+" indicates the fixation cross; vertical arrows indicate time points when the fixation cross is turned on (up) and off (down). "High", "Medium" and "Non-B" indicate experimental conditions, values below – time intervals in seconds.

cases where the observed agreement rate did not occur in the permutations, we defined p as the upper bound $p_u = (b + 1)/(m + 1)$ where b is the number of times when permuted agreement rate is equal or more extreme than the observed and m is the number of permutations (Phipson and Smyth, 2010; Puoliväli et al., 2020). We doubled the value to get a two-tailed p-value. To account for multiple comparisons, we applied the Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995) using MultiPy package for Python (Puoliväli et al., 2020) at $\alpha = 0.05$. Thus, if the agreement rate was higher than expected based on the null distribution, the boundary was considered a significant boundary. If the agreement rate was lower than expected based on the null distribution, the boundary was considered a significant non-boundary.

2.7. Analysis of the relationship between the linguistic cues and the agreement rate

To relate the agreement rate (dependent variable) to clausal syntactic structure, prosodic strength and pause length (independent variables), we fit a multiple linear regression model to Set 1 and Set 2 separately. Clause structure was represented as a continuous variable varying on a scale from 1 to 4 as described in Linguistic annotation of speech stimuli. Pause length was square root transformed and assigned to ten equally spaced bins in order to fulfil the assumption of linear relationship between independent and dependent variables. All independent variables were then normalized with a z-score. In determining statistical significance of regression coefficients, we used the bootstrapping approach to account for violations of model assumptions in non-normality of residuals, heteroscedasticity, and the presence of outliers. Bootstrap-based confidence intervals for each regression coefficient were determined by estimating distributions of 10 000 samples of regression coefficients from randomly resampled (with replacement) versions of the original dataset. In these resampled datasets, data for some boundaries was expressed more than once and for some, not at all. 99.9% confidence intervals for each regression coefficient were obtained by finding values at 0.05 and 99.95 percentile of the bootstrap distributions. P-values for each regression coefficient were the proportion of oppositely signed values in its bootstrap distribution. $P < 0.001$ was considered statistically significant.

3. MEEG experiment

3.1. Speech stimuli for the MEEG experiment

Based on the results of the behavioral chunking experiment (Fig. 2, B), we selected speech boundaries with high and medium agreement rates (see the Results section, Behavioral chunking experiment) and non-boundaries for the subsequent MEEG study. We inserted 2-second pauses at the locations of non-boundaries ("Non-boundary" condition) and at boundaries of medium and high agreement rate ("Medium" and "High" conditions). Non-boundaries were defined as boundaries which were marked by fewer than 5% of participants. All excerpts contained from two to seven 2-second silent pauses (Fig. 2, C). As a rule, each speech stimulus included at least one pause inserted within a chunk and one inserted at a chunk boundary. The mean distance between two consecutive pauses (±S.D.) was 4.1 ± 1.4 s. For the Non-boundary condition, 197 trials were used, for the Medium – 230, and for the High – 257. The stimuli were normalized in their intensity using the RMS (root mean square) function.

3.2. Experimental design and task

During the MEEG experiment, the speech stimuli were presented binaurally through plastic tubes and earpieces. The delivery of the stimuli was controlled by Presentation software (release 19.0, Neurobehavioral Systems, Inc., San Francisco, USA), which was also used for collecting the behavioral data (correct and incorrect responses, and reaction times). Presentation of each new excerpt was signaled by a fixation cross, which appeared in the middle of a screen (Fig. 2, D). During the experiment, the subjects were instructed to keep visual fixation on the cross throughout the presentation of a whole speech excerpt (including silent pauses). Two seconds after the offset of a speech stimulus, a comprehension question was presented on the screen for 7 s. In a forced-choice paradigm, the subjects had to answer this question by lifting either their right index or right middle finger from an optical sensor corresponding to "yes" or "no" response respectively. Efficiency of task performance was evaluated as the percentage of correct responses to the comprehension questions and mean reaction times. After a 3-s resting interval, when the fixation cross was turned off, the presentation of a new excerpt started. An experimental session included six blocks with the mean duration (±S.D.) of 21.6 ± 0.4 min and two breaks.

3.3. Acquisition of neuroimaging data

Concurrent 64-channel EEG and 306-channel (204 planar gradiometers and 102 magnetometers) MEG (Elekta-MEGIN) data were collected at the BioMag Laboratory, HUS Medical Imaging Center, in a magnetically shielded room (Euroshield Oy, Eura, Finland). The EEG was recorded with an Ag/AgCl-electrode cap; the reference electrode was placed on the nose. Vertical and horizontal electro-oculograms (EOG) were recorded in order to control eye movements. The recording high-pass filter was 0.03 Hz, and the sampling rate was 1000 Hz. Prior to data acquisition, the locations of five head position indicator (HPI) coils attached to the subject's head, and locations of all EEG electrodes were determined with respect to the three cardinal points (nasion and two preauricular points) by using a 3-D digitizer (Polhemus) in order to localize the position of the subject's head within the magnetometer helmet. The head position was localized at the beginning of each experimental block for further co-registration the MEG coordinate system with the subject's structural MRI. High-resolution 3D T1-weighted MPRAGE images ($1 \times 1 \times 1$ mm^3) were acquired for each subject (TR = 2530 ms, TE = 3.3 ms, TI = 1100 ms, flip angle = 7°) with a 3-T whole-body MRI scanner (Magnetom Skyra, Siemens) at the Advanced Magnetic Imaging Centre (Aalto University, Espoo, Finland). The volume consisted of 176 sagittal slices (FOV = 256 × 256 mm^2, 256 × 256 matrix, slice thickness = 1 mm).

3.4. MEEG data analysis

3.4.1. Preprocessing

Maxfilter with temporal signal space separation (tSSS) (Taulu and Simola, 2006) (Elekta Neuromag Ltd., Finland) was used for suppressing extra-cranial noise from MEG sensors and for the interpolation of bad channels. MEG channels exhibiting noise during the recordings were manually marked as bad prior to MaxFiltering procedure. The MEEG data were segmented into epochs starting 200 ms before and ending 1000 ms after the onset of the inserted silent pause. Epochs with signals exceeding peak-to-peak amplitude of 5 pT for magnetometers and 140 μV for EOG channels were automatically excluded from the analysis. The remaining epochs were visually inspected and those containing blinks or movement artifacts were rejected manually. Bad EEG channels were interpolated using the spherical spline method (Perrin et al., 1989). For the analysis of the event-related potentials (ERPs), the original reference system was used, while for the source reconstruction, the evoked responses were re-referenced to the average reference.

3.4.2. Analysis of event-related activity

Event-related sensor-level EEG time-series data were filtered with a passband of 0.5–20 Hz with FIR (finite impulse response) filter. Data were then averaged across trials and baseline-corrected using a 200-ms-time interval preceding a stimulus onset (i.e., the onset of a silent pause).

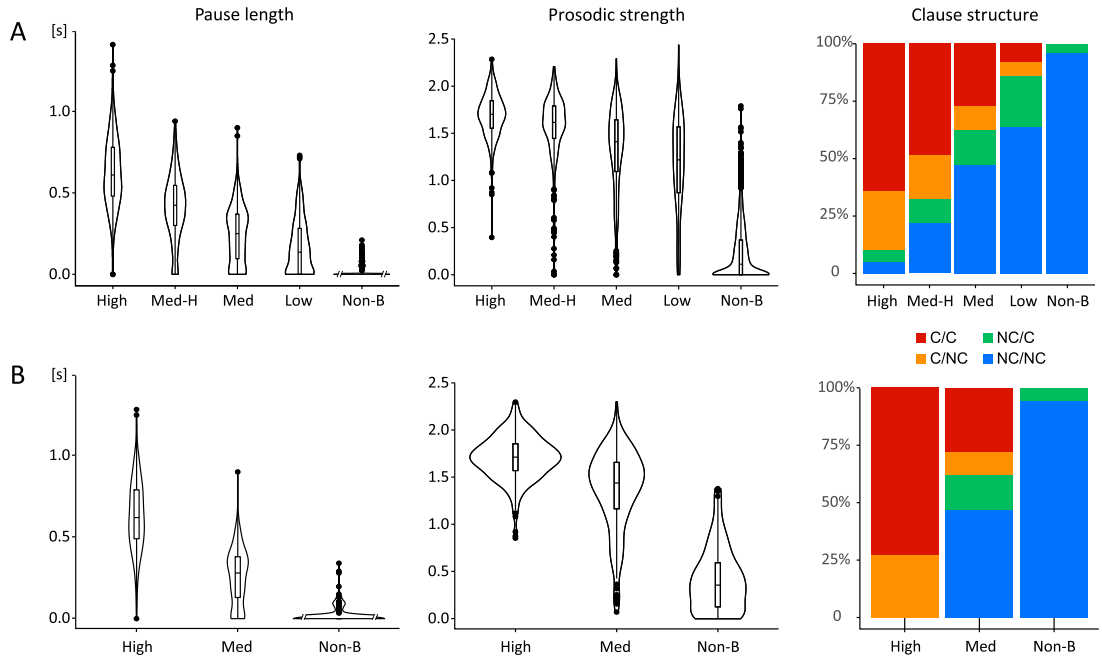


Fig. 3. Distribution of linguistic features with the agreement rate. A. The distribution of linguistic features (pause length, prosodic strength, and clause structure) across the boundary categories (High, Medium-High (Med-H), Medium (Med), Low, and Non-Boundary (Non-B)) in the behavioral experiment. B. The distribution of linguistic features across the experimental conditions (High, Medium, and Non-Boundary) in the MEEG experiment. C/C = clausal/clausal boundary; C/NC = clausal/non-clausal boundary; NC/C = non-clausal/clausal boundary; NC/NC = non-clausal/non-clausal boundary.

The ERPs were then analyzed within a set of 10 electrodes (Fig. 4, A), where the ERP components were most prominent (Fig 3, C). The ERP amplitudes were then averaged separately over three equidistant 120-ms time windows: 110–230 ms, 250–370 ms and 390–510 ms and then over the 10 electrodes for each participant. The first and the last time windows corresponded to the negative and positive deflections of ERPs observed in the Non-boundary condition respectively, and the second – to the positive deflections in the Medium and High conditions (Fig. 4, B).

3.4.3. Surface parcellation and Source reconstruction of MEEG data

Anatomical preprocessing included an automatic volumetric segmentation of the individual MRIs, surface reconstruction and surface parcellations, using FreeSurfer image analysis suite (Fischl, 2012) (http://surfer.nmr.mgh.harvard.edu/). Cortical parcellation and labeling was performed in accordance with the Destrieux atlas (Destrieux et al., 2010). MNE software (http://martinos.org/mne/stable/index.html) (Gramfort et al., 2014, 2013) was used for source reconstruction. Source configurations underlying the evoked responses were modeled using three-layer boundary element conductivity models created for each subject. The boundaries for the skin, skull and brain surfaces were determined using the watershed algorithm. The MEEG data were co-localized with individual anatomical images, and forward solutions were then calculated based on the volume conductor and the transformation information. The surface-based source spaces contained 8196 vertex locations for two hemispheres, with inter-vertex separation of 6 mm. For the preparation of the inverse operator, the Noise Covariance Matrices (NCMs) were computed using the last 500 ms of inserted pauses (i.e., 1500-2000 ms from a pause onset) which included no task-related activity. The inverse operator was computed using the depth weighting factor of 0.8, and loose source orientations with the weighting factor of 0.2 for the source variances of the dipole components that are tangential to the cortical surface. The dSPM method (Dale et al., 2000) was used for source localization. Source reconstructions were performed separately for each subject and experimental condition. The individual source solutions were then morphed to a common fsaverage template (Fischl et al., 1999) provided by the FreeSurfer software. Morphing an individual subject's source estimates to a common reference space was based on a spherical representation of the cortex computed using the spherical registration of FreeSurfer.

3.5. Statistical analysis of the MEEG data

The main hypothesis in this study was specifically to test whether inserting pauses at boundaries with different agreement rates (high and medium), and at non-boundaries induces different evoked responses. The null hypothesis here is that pauses inserted at different boundary locations elicit similar ERs so that there are no neural or ER correlates with the intuitive boundary markings. That is, if the intuitive boundaries would not be reflected in the brain activity.

3.5.1. Statistical analysis of the evoked potentials from sensor level EEG data

First, we calculated the statistical significance of evoked responses (ERs) in different time-windows and conditions in respect to the prestimulus baseline using the *t*-test. To control for multiple comparisons, p-values were corrected with the FDR correction.

Next, the statistical analysis of the ERP amplitudes was conducted using the 2-way ANOVA with the factors of Condition (Non-boundary, Medium and High) and Time window (3). The Greenhouse-Geisser correction was used for factors with more than two levels. Post-hoc analyses were performed using the Newman–Keuls test.

3.5.2. Statistical analysis of source activity

Time courses of the source activity were extracted from the predefined regions of interest (ROIs) within bilateral temporal and right frontal lobes. The temporal ROIs included the primary (transverse temporal (or Heschl's) gyrus and sulcus) and non-primary auditory cortex: posterior (planum temporale and posterior part of the lateral fissure), anterior (inferior insula) and lateral (STG and STS) parts. The inferior frontal ROIs included inferior frontal gyrus (pars opercularis, pars triangularis and pars orbitalis), inferior frontal sulcus, and vertical and horizontal rami of the anterior lateral fissure (Fig. 5, B). First, the time

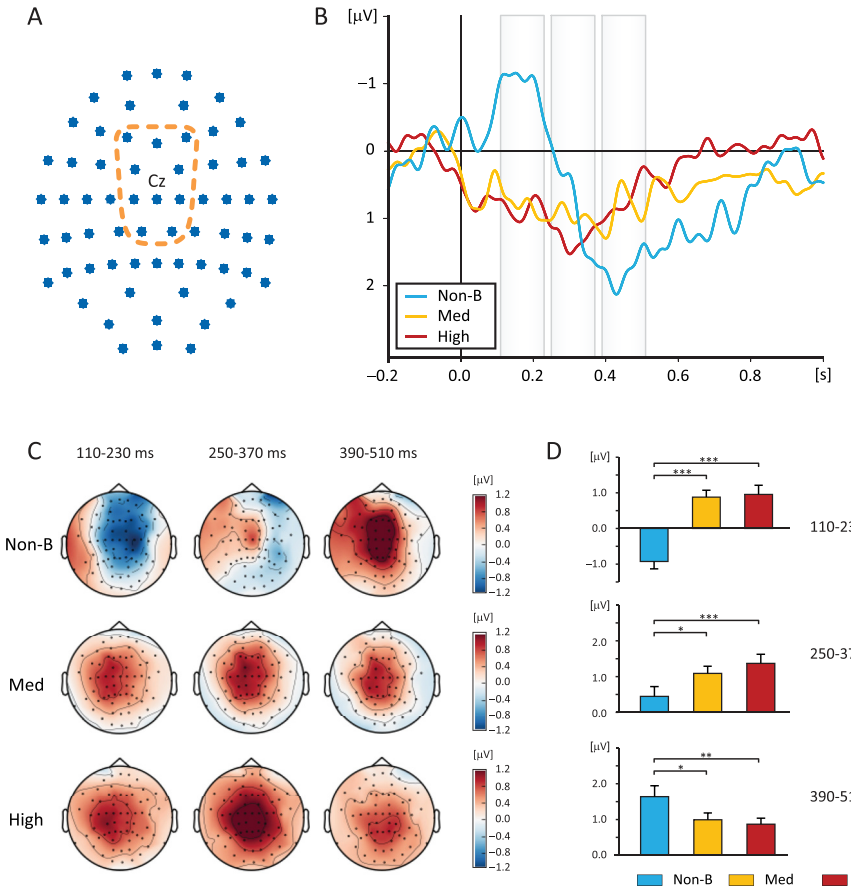


Fig. 4. Group-averaged ERPs for silent pauses inserted at chunk boundaries and within chunks. A. The 10 channels that were used for the analysis of the ERP amplitudes. B. The group-averaged ERPs recorded in the Non-boundary, Medium and High conditions at the Cz site. Shaded areas between the vertical lines – 120-ms time windows for the analysis of the ERP amplitudes: 110-230, 250-370 and 390-510 ms. C. The group-averaged topomaps for the mean amplitudes (μV) within the analyzed time windows in the Non-boundary, Medium and High conditions. D. The effect of condition on the amplitude of the ERPs for each analyzed time window. The vertical lines indicate SEM. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

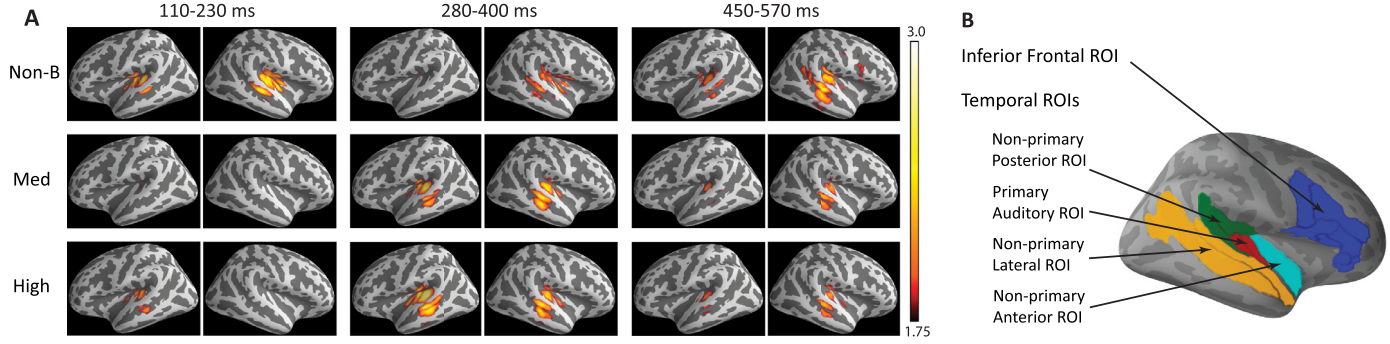


Fig. 5. Cortical localization of the boundary effects. A. The group averaged cortical localization for each condition and analyzed time windows. The maps are superimposed onto a standard brain template from the Freesurfer. B. Regions of interest (ROIs) for the analysis of the source amplitudes based on Destrieux parcellation).

courses were extracted for each ROI as averages across vertices at each time point separately for each subject, and then, for illustrative purpose, averaged across the group of subjects (Fig. 6, A).

For statistical evaluation of the ROI activity, three equidistant 120-ms time windows were used: 110–230 ms, 280–400 ms and 450–570 ms. The mean dSPM values were analyzed using 3-way ANOVA with the factors of Condition (Non-boundary, Medium and High), ROI (Primary, Posterior, Anterior and Lateral), and Hemisphere (Left and Right) separately for each time window. For the temporal sources, statistical evaluation of the source coordinates was performed in the same time windows as those used for the analysis of the amplitudes. However, the centers of mass were computed for the whole temporal ROIs which included all four subdivisions (Primary, Posterior, Anterior and Lateral) (Fig. 5, B). After converting the array of source vertices to MNI coordinates, the coordinate values were evaluated statistically using 2-way ANOVA with the factors of Condition and Hemisphere. In all statistical tests, the Greenhouse-Geisser correction was used for factors with more than two levels. Post-hoc analyses were performed using the Newman–Keuls test. For the inferior frontal ROI, coordinate values of the source activity observed during the third time window in the Non-boundary condition were evaluated in order to prove its consistency among the participants.

4. Results

4.1. Behavioral chunking experiment

We first identified chunk boundaries based on the agreement rate across the participants. The results showed that boundaries marked by more than 22% of participants were statistically significant chunk boundaries and boundaries with the agreement rate of less than 5% were statistically significant non-boundaries. Statistically significant chunk

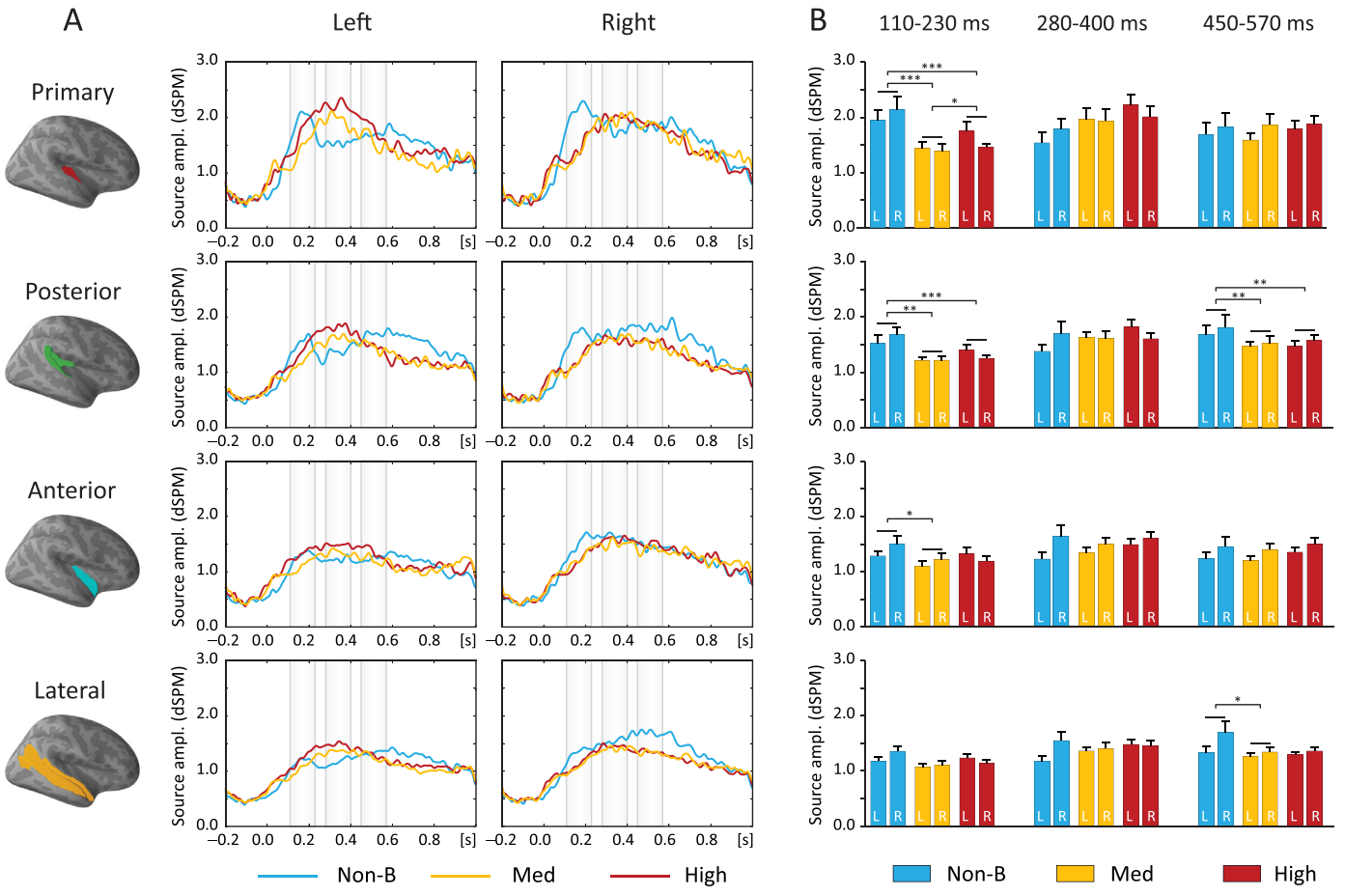


Fig. 6. Temporal evolution of source activity A. Time courses for the separate Temporal ROIs estimated for the left and right hemisphere. B. The effect of condition on the source amplitude for each analyzed time window. L = left hemisphere, R = right hemisphere. All other designations as in Fig. 4.

boundaries were then divided into four equal-sized bins (quartiles) of low (22-33%), medium (35-57%), medium-high (58-79%) and high (80-100%) agreement rates.

Fig. 3 shows the distribution of linguistic features across the boundary categories: pause length, prosodic strength, and clause structure. All three linguistic features clearly decrease in strength from boundaries with higher agreement rates to boundaries with lower agreement rates and non-boundaries. Furthermore, while pause duration and the proportion of completed clauses decrease gradually with decreasing agreement rate, prosodic strength drops abruptly from chunk boundaries to non-boundaries.

The mean duration of intuitively defined chunks (±S.D.) was 2.4 ± 1.1 s for the two sets of speech stimuli.

To further test the effect of different linguistic features on agreement rate, we fitted a multiple linear regression model to the continuous version of the agreement rate. This analysis revealed that each of the features studied – pause length, clausal syntactic structure and prosodic strength – has a statistically significant relationship with the agreement rate ($p < 0.001$). The regression coefficients of the features with their bootstrapped 99.9% confidence intervals (CI) are provided in Table 1. Since all variables were standardized, the regression coefficients can be interpreted in terms of variable importance, and show that pause length has the strongest relationship with the agreement rate, followed by syntactic structure and prosodic strength. Each pair of independent variables was correlated. The correlation between clause structure and prosodic strength was 0.58 for Set 1 and 0.55 for Set 2, between clause structure and pause length 0.61 for Set 1 and 0.58 for Set 2, and between prosodic strength and pause length 0.89 for Set 1 and 0.9 for Set 2.

Table 1
Results of multiple linear regression relating agreement rate to linguistic cues.

Linguistic cues	regression coefficients with 99.9% CI		VIF	
	Set 1	Set 2	Set 1	Set 2
Pause length	5.64 [4.65, 6.52]	5.54 [4.7, 6.41]	5.2	5.45
Clause structure	4.3 [3.76, 4.88]	3.66 [3.16, 4.18]	1.6	1.52
Prosodic strength	1.69 [0.84, 2.59]	1.24 [0.47, 2.03]	4.93	5.15

CI – confidence intervals, VIF – Variance Inflation Factor

These correlations suggest multicollinearity between the independent variables, which might influence the estimated regression coefficients and their standard deviations. We used the Variance Inflation Factor (VIF) to gauge the influence of the multicollinearity on regression coefficients. The VIF of each of the independent variables was less than 10 (Table 1), indicating that their estimated regression coefficients can be interpreted (Altman and Krzywinski, 2016; Kulesa et al., 2015). Moreover, as mentioned above, we estimated confidence intervals of the regression coefficients using a bootstrapping approach (Kulesa et al., 2015), which typically yields wider confidence intervals since it accounts for observed collinearity between independent variables. Hence, the results of our analyses are robust to the observed collinearity between independent variables, and the model reliably estimates the unique contribution of each cue to chunk boundary perception.

Since there are many boundaries marked by fewer than 5% of participants, we tested whether the outcome of the analyses would be different if these boundaries were excluded from the analysis. The results did not change, i.e., each of the independent variables still had a statistically significant relationship with the agreement rate (see Table A.1 in Appendices for the regression coefficients).

Overall, the results of the regression analysis indicate that in real-time speech chunking listeners use all three cues – pause length, clause structure and prosodic strength.

5. MEEG experiment

5.1. Linguistic features of the stimuli

Distributions of the linguistic characteristics of the boundaries selected for the MEEG experiment are shown in Fig. 3. In the High agreement-rate condition, the boundaries always occurred at the end of a clause and were characterized by longer original pauses (633 ± 21 ms, mean ± S.D.) and higher prosodic strength (1.45 ± 0.33). The boundaries in the Medium agreement-rate condition were more syntactically variable: they occurred within a clause (47.4%), at the end of a clause (37.4%) and at the start of an embedded clause (15.2%). They were also prosodically weaker (1.18 ± 0.45) and contained a shorter original pause (260 ± 172 ms). The non-boundaries never occurred at the end of a clause, had only a short pause or no pause at all (16 ± 49 ms), and a very low prosodic strength (0.25 ± 0.30).

5.2. Behavioral data

The analysis of responses to comprehension questions indicated that all the subjects were able to understand the speech excerpts. The mean accuracy of task performance (± standard error of mean (S.E.M)) was 79 ± 1.6%, and the mean reaction time was 4.2 ± 0.2 s.

5.3. The effect of boundary agreement rate on ERPs

We first estimated ERPs from EEG sensor-level data separately for each condition (High, Medium and Non-boundary) relative to the 200-ms baseline preceding the silent pause. The ERPs differed considerably in their waveforms and scalp distributions between the non-boundary and boundary conditions (Fig. 4, B and C).

Pauses inserted at non-boundaries elicited a biphasic response with its negative phase peaking around 150-200 ms from the pause onset, and a positive phase peaking at around 450 ms. Responses to pauses inserted at the medium and high agreement-rate chunk boundaries were monophasic and peaked around 300 ms in the High and 400 ms in the Medium condition.

In the Non-boundary condition, the first peak was significantly more negative compared to the baseline in the earliest time window ($t(19) = -4.5$, $p_{(FDR)} < 0.001$), and the second peak was more positive in the latest time window ($t(19) = 5.4$, $p_{(FDR)} < 0.001$). In the second time window the ERP amplitude did not differ significantly from the baseline. In the High and Medium conditions, the ERP amplitudes were significantly more positive compared to the baseline in all three time windows ($t(19) > 3.7$, $p_{(FDR)} < 0.01$).

Further statistical analysis using ANOVA revealed main effects of Condition ($F(2,38) = 5.9$, $p < 0.01$, $\eta_p^2 = 0.24$, Power = 0.85) and Time window ($F(2,38) = 25.2$, $p < 0.001$, $\eta_p^2 = 0.57$, Power = 1.00), and Condition and Time window interaction ($F(2,38) = 26.8$, $p < 0.001$, $\eta_p^2 = 0.59$, Power = 1.00). Overall, the ERP amplitude was more negative in the first compared to the second ($p < 0.001$) and third ($p < 0.001$) time windows, and in the Non-boundary condition the amplitude was more negative than in the High ($p < 0.01$) and Medium ($p < 0.01$) ones. Furthermore, the ERP amplitudes differed significantly among conditions in all analyzed time windows. The post-hoc analysis showed that during the first two windows, the amplitude of the ERP recorded in the Non-boundary condition was significantly more negative compared to the Medium ($p < 0.001$ and $p < 0.05$ for the fist and the second window respectively) and High ($p < 0.001$ and $p < 0.001$) conditions. During the latest time window, by contrast, the ERP amplitude in the Non-boundary condition was significantly more positive than in the Medium ($p < 0.05$) and High ($p < 0.01$) conditions (Fig. 4, D). No significant differences were observed between the High and Medium conditions.

5.4. The effect of boundary agreement rate on evoked responses in source space

We next estimated event-related activity from source reconstructed MEEG data. Evoked responses were localized to both primary and non-primary auditory cortices and inferior frontal areas (Fig. 5, A).

Temporal cortices were consistently activated during all experimental conditions, while activation of the right inferior frontal cortex was observed in the Non-boundary condition during the third time window from 450 to 570 ms. Similarly to the ERPs at the EEG sensor level, evoked responses in the source space also differed among the agreement-rate conditions. During the first and the last time windows, activity was stronger in the Non-boundary condition, while during the second time window it was stronger in the High and Medium conditions (Fig. 5, A and Fig. 6, A and B).

To investigate these differences more closely, we performed statistical evaluation of time courses for event-related data averaged over four temporal ROIs (Fig. 4, B). This analysis revealed that during the first time window, the amplitude of activity was affected by the Condition ($F(2,38) = 5.7$, $p < 0.05$, $\eta_p^2 = 0.23$, Power = 0.84), being higher in the Non-boundary compared to the High ($p < 0.05$) and Medium ($p < 0.01$) conditions. The main effect of ROI was also significant ($F(3,57) = 45.0$, $p < 0.001$, $\eta_p^2 = 0.70$, Power = 1.00). The amplitude of activity gradually decreased from the Primary to the Posterior, then to the Anterior and finally to the Lateral ROIs (Primary > other ROIs, $p < 0.001$; Posterior > Anterior and Lateral, $p < 0.05$ and 0.001 respectively; Anterior > Lateral, $p < 0.05$). Significant ROI and Condition interaction ($F(6,114) = 4.0$, $p < 0.05$, $\eta_p^2 = 0.17$, Power = 0.97) indicated that the effect of the Condition varied among the ROIs (Fig. 6). In the Primary ROI, the activity in the Non-boundary condition was stronger compared to the High ($p < 0.001$) and Medium ($p < 0.001$) conditions, and in the High condition it was stronger than in the Medium condition ($p < 0.05$). In the Posterior ROI, the activity was stronger in the Non-boundary compared to the High ($p < 0.01$) and Medium ($p < 0.001$) conditions. In the Anterior ROI, significant difference was observed between the Non-boundary and the Medium conditions only ($p < 0.05$). No differences were found in the Lateral ROI. Finally, there was a Condition and Hemisphere interaction, indicating interhemispheric differences in the effect of the Condition. In the left hemisphere, activity in the Medium condition was lower than in the Non-boundary ($p < 0.01$) and High ($p < 0.05$) conditions. Furthermore, in the Non-boundary condition activity was stronger in the right compared to the left hemisphere ($p < 0.05$), while in the High condition, it was stronger in the left than in the right hemisphere ($p < 0.05$). In the second time window, the source activity also differed among the ROIs ($F(3,57) = 21.7$, $p < 0.001$, $\eta_p^2 = 0.53$, Power = 1.00), being stronger in the Primary than in all other ROIs ($p < 0.001$), and in the Posterior stronger than in the Anterior ($p < 0.05$) and Lateral ($p < 0.05$) ROIs. Hemisphere and Condition interaction ($F(2,38) = 6.0$, $p < 0.05$, $\eta_p^2 = 0.24$, Power = 0.86) indicated that in the left hemisphere, activity during the pauses inserted at chunk boundaries was stronger compared to the activity in the Non-boundary condition (High > Non-boundary, $p < 0.001$, Medium > Non-boundary, $p < 0.01$), and that this drop in activity was confined to the left hemisphere (Non-boundary left < Non-boundary right, $p < 0.01$). In the third time window, the effect of ROI was similar to the one in the second time window ($F(3,57) = 29.5$, $p < 0.001$, $\eta_p^2 = 0.61$, Power = 1.00). The source activity was stronger in the Primary than in all other ROIs ($p < 0.001$), and in the Posterior stronger than in the Anterior ($p < 0.001$) and Lat-

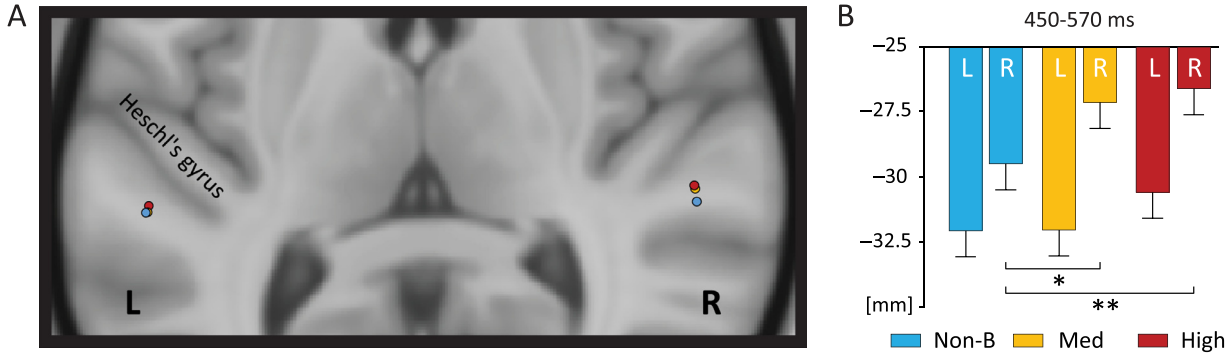


Fig. 7. A. The mean locations of the combined Temporal sources estimated for the 450-570 ms time window overlaid on the standard MNI template. B. The effect of condition on the Temporal source coordinates along the y-axes for the 450-570 ms time window. All other designations as in Fig. 4.

eral ($p < 0.001$) ROIs. The interaction between the ROI and Condition ($F(6,114) = 2.8$, $p < 0.05$, $\eta_p^2 = 0.13$, Power = 0.87) indicated, that significant differences among the conditions were observed in non-primary ROIs, Posterior (Non-boundary > High and Medium, $p < 0.01$) and Lateral (Non-boundary > Medium, $p < 0.05$), but not in the Primary one.

5.5. The effect of boundary agreement rate on source locations

The MNI coordinates of the centers of mass of the combined temporal ROIs differed significantly between the left and right hemisphere along *y*-axes in all analyzed time windows ($F(1,19) = 18.4$, $p < 0.001$, $\eta_p^2 = 0.49$, Power = 0.98 for the first time window, $F(1,19) = 19.2$, $p < 0.001$, $\eta_p^2 = 0.50$, Power = 0.99 for the second, and ($F(1,19) = 15.9$, $p < 0.001$, $\eta_p^2 = 0.46$, Power = 0.97 for the third). The right-hemispheric sources were located 4 mm more anteriorly in respect to their left-hemispheric homologs (Fig. 7). Furthermore, *y*-coordinates of the sources of activity during the third time window differed significantly among the conditions ($F(2,38) = 4.6$, $p < 0.05$, $\eta_p^2 = 0.19$, Power = 0.74). The center of mass in the Non-boundary condition was located more posteriorly with respect to source locations in other conditions. Moreover, sub-areal segregation of the source locations was observed exclusively in the right hemisphere. Planned comparison using Student's T-test followed with a Bonferroni correction confirmed significant differences along *y*-axes between the source locations in the Non-boundary and two other conditions (Non-boundary vs. High, 3 mm difference, $p < 0.01$, and Non-boundary vs. Medium, 2 mm difference, $p < 0.05$) in the right but not in the left hemisphere. No significant effects were observed along *x*- or *z*-axes. The right-hemispheric inferior frontal activity was consistently observed in all participants during the latest time window in the Non-boundary condition. The mean coordinates of the right inferior frontal source corresponded to the opercular part of the inferior frontal gyrus.

6. Discussion

This study investigated neurocognitive mechanisms underlying the intuitive chunking of natural speech using a combined behavioral and electrophysiological study. We first identified intuitive chunk boundaries from a behavioural experiment in which the participants simultaneously listened to speech extracts and followed their transcripts on tablet computers, while intuitively marking chunk boundaries in excerpts of continuous, spontaneous speech. Listeners converged on placing chunk boundaries, showing that boundaries with high agreement rates correlated with the presence of a variety of linguistic cues, such as prosodic boundary strength, clause structure, and pause length. These data thus confirmed the presence of intuitive chunk boundaries in naïve listeners (Sinclair and Mauranen, 2006), so that boundaries were based on holistic perception not driven by only one type of linguistic cue (Hanulíková et al., 2012; Van Berkum et al., 2008).

In the next experiment, we used MEEG to investigate possible neuronal markers of intuitive chunking and more specifically, whether pauses inserted at non-boundaries, which were experienced as interruptions of speech, would be associated with different evoked responses from those elicited by pauses inserted at chunk boundaries with high and medium agreement rates. As hypothesized, we found that pauses inserted at chunk boundaries with high and medium agreement rates across participants elicited a closure positive shift with the sources over bilateral auditory cortices. By contrast, pauses placed at non-boundaries were perceived as interruptions and elicited a biphasic emitted potential with sources located in the bilateral auditory areas with right-hemispheric dominance, and in the right inferior frontal cortex. The data establishes that intuitive boundary marking and the resulting chunk boundaries are correlated with the neuronal activity as measured with MEEG.

6.1. The influence of using a second language

The data from which the stimuli were drawn was based on two corpora of academic speech, which represented both native speakers and fluent non-native speakers of English, thus reflecting the reality of contemporary academic conferences and international universities. Neither of databases was compiled to represent any single (or more) first-language group of the speakers. The listeners participating in the present study likewise represented international speakers of English as a group, with a variety of first languages, without seeking a representative sample of any first language group. Therefore, the influence of any specific first language would not be possible. In support of this choice, another recent study (Dobrego et al., under revision) compared intuitive chunking behavior in listeners representing first and second-language speakers of English and found no differences in either inter-subject agreement on chunking, or in the mean chunk length between the groups. Furthermore, in a study by Frost and colleagues (Frost et al., 2017), no effect of language background on the use of a syllable duration cue for speech segmentation was found. This also supports Sinclair and Mauranen (2006) original hypothesis, which posits that 'fluent speakers' of the language will chunk speech in essentially the same way.

6.2. Monophasic ERs to intuitive chunk boundaries

Evoked responses to chunk boundaries of high and medium agreement rate were monophasic, with the sources over bilateral primary and non-primary auditory areas and peaked at 300-400 ms from the pause onset. The left-hemispheric activity for the boundaries with high agreement rate manifested slightly earlier than that for the medium one, the maximal difference between the two conditions thus being during the first time window (Figs. 5 and 6). More prominent left-hemispheric activity in the High than Medium agreement-rate condition may reflect

modulation of the responses by for example syntactic structure, since in contrast to Medium condition, all the stimuli in the High condition were associated with the completion of a clause. The evoked responses to high and medium agreement-rate chunk boundaries may correspond to the CPS, which is characterized by bilateral positivity originally found in the temporal vicinity of intonation boundaries (Friederici and Alter, 2004; Steinhauer et al., 1999). Furthermore, positive shifts were found in responses to verbal or non-verbal temporal groups characterized by a specific prosodic feature such as the lengthening of a final syllable, and were linked to the perceptual chunking of speech (Gilbert et al., 2015). Importantly, Gilbert et al.'s (2015) study found similar positivity irrespective of the presence of meaningful linguistic forms. Our results expand on these findings by showing that in natural speech, brain responses to chunk boundaries are influenced both by prosody and syntactic clause structure so that the strongest responses are observed when these two features converge. The contribution of the clause structure factor is evident from enhanced left-hemispheric responses in the High compared to the Medium condition characterized by a smaller proportion of completed clauses. This finding confirms our hypothesis of modulatory effect of syntactic structure on boundary-related brain activity. At the level of scalp-recorded ERPs, which, due to large volume conduction in EEG sensor level data reflect integrated activity from both hemispheres, no significant difference between the High and Medium conditions was observed. However, source modeling of MEEG data allowed disentangling neuronal activity originated from anatomically precise areas located in the left and right hemispheres, and thus revealed condition-related amplitude differences specific to the left hemisphere. This result is in line with previous findings concerning the rightward shift of the CPS amplitude maximum as a function of reduced linguistic content (Pannekamp et al., 2005) and further with the notion that speech comprehension involves parsing of sequential auditory information according to abstract grammatical rules into a set of word groups forming nested linguistic trees (Dehaene et al., 2015). However, it is very likely that other factors also play a role in the emergence CPS in the chunking of natural speech. This calls for further research.

6.3. Biphasic ERs to speech interruption

Responses to pauses at non-boundaries leading to interruptions of speech were different from those found for chunk boundaries with high agreement rates. These results are in line with the findings showing that integrating speech in time depends on temporal expectancies and attention (Scharinger et al., 2017).

Compared to the monophasic responses to high and medium agreement-rate boundaries, non-boundaries were associated with biphasic evoked responses, with a negative wave peaking around 150-200 ms and a following positive wave with the latency around 450 ms. A similar pattern of activity has earlier been shown in response to interruptions of continuous auditory stimuli such as speech and music (Besson et al., 1997; Besson and Faïta, 1995) and was referred to as an emitted potential, i.e., a fully endogenous response elicited in the absence of sensory events (Weinberg et al., 1974, 1970). The first phase of the emitted potential may represent the omission mismatch negativity elicited by expected but missing auditory input (Bendixen et al., 2014; Horvath et al., 2010; Pihko et al., 1997; Raij et al., 1997; Tervaniemi et al., 1994; Yabe et al., 1998, 1997). However, it may also include an obligatory off-response to the abrupt cessation of the speech sound in the Non-boundary condition.

Source reconstruction of the first peak specific to the Non-boundary condition revealed prominent activity in the temporal lobes with significant right-hemispheric dominance (Fig. 5). Further exploration of the distribution of this activity across temporal subregions allowed us to determine that the first phase of the response to chunk interruption originated mainly in the primary and posterior auditory cortex (Fig. 6). In both ROIs, the activity was stronger in the Non-boundary than in the High and Medium conditions.

The second, positive phase of the emitted potential may correspond to the P3a, which often follows the MMN in conditions requiring reorienting attention to facilitate the processing of novelty (Polich, 2007). Neural generators of the later phase of the emitted potential were localized predominantly in the right-hemispheric temporal and inferior frontal cortices (Fig. 5). The involvement of the inferior frontal cortex in the generation of the P3a is in line with earlier results from electrophysiological (Zora et al., 2020) and neuroimaging studies (Asplund et al., 2010; Corbetta and Shulman, 2002; Ford et al., 1994), as well as from clinical studies in patients with frontal lesions (Knight, 1984; Knight et al., 1995; Szczepanski and Knight, 2014). In contrast to the first phase of the emitted potential, when the maximal difference between the Non-boundary and two other conditions was observed over the primary auditory cortex, a significant increase of the later activity was confined to the non-primary auditory cortex. Chunk interruptions compared to chunk boundaries elicited more prominent and sustained activity within the bilateral posterior, the right lateral temporal, and the right prefrontal areas (Fig. 6). This might be caused by the maintenance of precise prosodic information in working memory compensating for the interruption. Hence, more prominent responses to chunk interruptions compared to chunk boundaries may result from enhanced activity within the prosody-related right-hemispheric network. Furthermore, the increase of activity in association areas may also be related to generating top-down predictions regarding a forthcoming continuation of an interrupted chunk and consequently building up a holistic image of this chunk during the delay period (Blumenthal-Dramé et al., 2017; Bornkessel-Schlesewsky et al., 2016). Such predictions may be based on preceding information from different linguistic domains, such as the semantic context, syntax, and prosody.

6.4. Cortical sources of evoked responses to chunk boundaries and chunk interruptions

In line with previous studies, the source localization of ER activity elicited by speech stimuli revealed neuronal sources in the superior temporal cortices which have been shown to be sensitive to spectro-temporal features of speech corresponding to different phonemes (Chang et al., 2010; Mesgarani et al., 2014), and to the dynamics of speech sounds compared to other sounds with similar acoustic content (Nora et al., 2020; Overath et al., 2015). The analysis of source coordinates revealed that the center of mass in the evoked activity was located in the planum temporale in all experimental conditions (Fig. 7). As expected, the right-hemispheric sources were located 4 mm more anteriorly in respect of their left-hemispheric homologs, reflecting the anatomical hemispheric asymmetry of the Heschl's gyrus and the planum temporale (Westbury et al., 1999). This interhemispheric difference was consistent during all conditions and analyzed windows. In addition, a significant difference in the source locations among all three conditions was found during the latest time window: the right-hemispheric generator of the response elicited in the Non-boundary condition was located significantly more posteriorly in respect of the sources of activity in the Medium and High conditions. However, this finding should be considered with some caution, since the mean difference in source coordinates along the y-axes was small, although the relative positions of the sources of responses to chunk boundaries and chunk interruptions were consistent among the participants.

There is a growing body of evidence that speech processing takes place in two anatomically distinct and functionally specialized pathways (Hickok and Poeppel, 2007; Rauschecker, 2011; Rauschecker and Scott, 2009; Sammler et al., 2018). The left-hemispheric ventral pathway is involved in processing semantic content (DeWitt and Rauschecker, 2012; Pylkkänen, 2019; Rauschecker and Scott, 2009) and local syntactic structure building

(Friederici et al., 2006), while the dorsal route is linked to speech production (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009) and the building of complex syntactic structures (Wilson et al., 2011). Right-hemispheric ventral and dorsal pathways are both involved in processing speech prosody (Fruhholz et al., 2015; Sammler et al., 2015). Furthermore, the integrity of the right-hemispheric dorsal pathway has been found to be crucial for building linguistic prosodic structure (Sammler et al., 2018).

In view of the more dorsal source locations and more prominent responses in the right posterior auditory cortex during chunk interruptions, it is possible to suggest that breaking the intonation contour is a crucial factor modulating brain activity during chunk interruptions. Thus, the integrity of the intonation contour may be considered as an essential property of the chunk. However, we did not find any evidence for a direct correspondence between intuitive chunks and prosodic units. The mean chunk duration in our study was 2.4 s, suggesting that an intuitive chunk may include more than one intermediate intonation phrase (Stehwien and Meyer, 2021) or prosodic phrase (Inbar et al., 2020). However, it is noteworthy that the speech material used in the present study might have biased the duration of prosodic phrases to some extent, since we extracted excerpts exclusively from monologic events such as lectures, presentations, etc., recorded in an academic environment. Thus, it would be important to continue investigating chunk properties and language processing more generally as a holistic, integrated process rather than continue to separate and isolate the roles of individual factors in accounting for segmentation. Such an approach would hold more promise for ecological validity in future research.

7. Conclusions

In conclusion, we have demonstrated that chunk boundaries intuitively marked by naïve participants with high and medium agreement rates in a behavioral experiment elicit distinct evoked activity in the brain compared to non-boundaries. As hypothesized, listeners spontaneously identified chunk boundaries driven by the presence of multiple linguistic cues including prosody and syntactic cues therefore supporting the idea that language processing is holistic. At the neural level, boundaries with high-and medium agreement rates induced monophasic ERs similar to CPS that is typically observed at the end of prosodic boundaries. Furthermore, lest-hemispheric increase in source activity in response to high compared to medium agreement rate boundaries suggests that these monophasic responses were modulated by syntactic structure. In contrast, pauses inserted at non-boundaries induced a distinct evoked response pattern, an emitted potential. Future studies are needed to investigate whether chucking of natural speech is reflected in the brain rhythmicity known to contribute to speech segmentation (Henke and Meyer, 2021; Keitel et al., 2018; Morillon et al., 2019; Rimmele et al., 2021).

Data and code availability statement

The data underlying the results in figures are available from the corresponding authors upon reasonable request. The data was collected under provision of informed consent of the participants. Access to the data will be granted in line with that consent, subject to approval by the project ethics board and under a formal Data Sharing Agreement. The software used to process and analyze the data (FreeSurfer, MNE Python, WebMAUS, Praat, Wavelet Prosody Toolkit, and MultiPy) are open source and freely available.

Declaration of Competing Interest

The authors declare no conflict of interest.

Credit authorship contribution statement

Irina Anurova: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Svetlana Vetchinnikova:** Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – review & editing. **Aleksandra Dobrego:** Formal analysis, Investigation, Writing – review & editing. **Nitin Williams:** Formal analysis, Methodology, Visualization. **Nina Mikusova:** Data curation, Formal analysis. **Antti Suni:** Data curation, Software. **Anna Mauranen:** Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Writing – original draft, Writing – review & editing. **Satu Palva:** Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Acknowledgments

This study was supported by grants from the Finnish Cultural Foundation (Grant # 00160622) to Prof. A. Mauranen, and from Sigrid Jusélius Foundation to Prof. S. Palva. We wish to thank Tuomas Puoliväli for his help with statistical analysis of the behavioral data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119203.

References

Altman, N., Krzywinski, M., 2016. Regression diagnostics. Nat. Methods 13, 385–386. doi:10.1038/nmeth.3854.

Asplund, C.L., Todd, J.J., Snyder, A.P., Marois, R., 2010. A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. Nat. Neurosci. 13, 507–512. doi:10.1038/nn.2509.

Beach, C.M., 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations. J. Mem. Lang. 30, 644–663. doi:10.1016/0749-596X(91)90030-N.

Bendixen, A., Scharinger, M., Strauss, A., Obleser, J., 2014. Prediction in the service of comprehension: modulated early brain responses to omitted speech segments. Cortex 53, 9–26. doi:10.1016/j.cortex.2014.01.001.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc.. Ser. B (Methodol.) 57, 289–300.

Besson, M., Faïta, F., 1995. An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. J. Exp. Psychol. Hum. Percept. Perform. 21, 1278–1296. doi:10.1037/0096-1523.21.6.1278.

Besson, M., Faïta, F., Czternasty, C., Kutas, M., 1997. What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. Biol. Psychol. 46, 3–23. doi:10.1016/s0301-0511(96)05215-5.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., 1999. Longman Grammar of Spoken and Written English. Longman, Harlow.

Blanco-Elorrieta, E., Pylkkänen, L., 2017. Bilingual language switching in the lab vs. in the wild: the spatio-temporal dynamics of adaptive language control. J. Neurosci. doi:10.1523/JNEUROSCI.0553-17.2017.

Blank, I.A., Fedorenko, E., 2017. Domain-general brain regions do not track linguistic input as closely as language-selective regions. J. Neurosci. 37, 9999–10011. doi:10.1523/JNEUROSCI.3642-16.2017.

Blumenthal-Dramé, A., Glauche, V., Bormann, T., Weiller, C., Musso, M., Kortmann, B., 2017. Frequency and chunking in derived words: a parametric fMRI study. J. Cogn. Neurosci. 29, 1162–1177. doi:10.1162/jocn_a_01120.

Boersma, P., Weenink, D., 2017. Praat: Doing Phonetics by Computer. University of Amsterdam.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D.J., 2011. Prosodic breaks in sentence processing investigated by event-related potentials. Lang. Linguist. Compass 5, 424–440. doi:10.1111/j.1749-818X.2011.00291.x.

Bonhage, C.E., Meyer, L., Gruber, T., Friederici, A.D., Mueller, J.L., 2017. Oscillatory EEG dynamics underlying automatic chunking during sentence processing. Neuroimage 152, 647–657. doi:10.1016/j.neuroimage.2017.03.018.

Bornkessel-Schlesewsky, I., Staub, A., Schlesewsky, M., 2016. The timecourse of sentence processing in the brain. In: Hickok, G., Small, S.L. (Eds.), Neurobiology of Language. Academic Press, San Diego, pp. 607–620. doi:10.1016/B978-0-12-407794-2.00049-3.

Brazil, D., Sinclair, J., Carter, R., 1995. A Grammar of Speech. Oxford University Press, Oxford.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D.J., Pylkkänen, L., 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. Brain Lang. 120, 163–173. doi:10.1016/j.bandl.2010.04.002.

Buxó-Lugo, A., Watson, D.G., 2016. Evidence for the influence of syntax on prosodic parsing. J. Mem. Lang. 90, 1–13. doi:10.1016/j.jml.2016.03.001.

Carnie, A., 2010. Constituent structure. Oxford Surveys in Syntax and Morphology, 2nd ed. Oxford University Press, Oxford; New York.
Chai, L.R., Mattar, M.G., Blank, I.A., Fedorenko, E., Bassett, D.S., 2016. Functional network dynamics of the language system. Cereb Cortex 26, 4148–4159. doi:10.1093/cercor/bhw238.
Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432. doi:10.1038/nn.2641.
Christiansen, M.H., Chater, N., 2016. The now-or-never bottleneck: a fundamental constraint on language. Behav. Brain Sci. 39, e62. doi:10.1017/S0140525X1500031X.
Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215. doi:10.1038/nrn755.
Culbertson, G., Andersen, E., Christiansen, M.H., 2020. Using utterance recall to assess second language proficiency. Lang. Learn. 70, 104–132. doi:10.1111/lang.12399.
Cutler, A., Dahan, D., van Donselaar, W., 1997. Prosody in the comprehension of spoken language: a litERATURE REVIEw. Lang. Speech 40, 141–201. doi:10.1177/002383099704000203.
Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26, 55–67. doi:10.1016/s0896-6273(00)81138-1.
Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., Pallier, C., 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. Neuron 88, 2–19. doi:10.1016/j.neuron.2015.09.019.
Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15. doi:10.1016/j.neuroimage.2010.06.010.
DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. 109, E505–E514. doi:10.1073/pnas.1113427109.
Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). Front. Hum. Neurosci. 11, 481. doi:10.3389/fnhum.2017.00481.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164. doi:10.1038/nn.4186.
Dobrego, A., Konina, A., Mauranen, A., n.d. Continuous speech segmentation by L1 and L2 speakers of English: the role of syntactic and prosodic cues. 2022
Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85 (Pt 2), 761–768. doi:10.1016/j.neuroimage.2013.06.035.
Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781. doi:10.1016/j.neuroimage.2012.01.021.
Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284. doi:10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4.
Ford, J.M., Sullivan, E.V., Marsh, L., White, P.M., Lim, K.O., Pfefferbaum, A., 1994. The relationship between P300 amplitude and regional gray matter volumes depends upon the attentional system engaged. Electroencephalogr. Clin. Neurophysiol. 90, 214–228. doi:10.1016/0013-4694(94)90093-0.
Frazier, L., Carlson, K., Clifton, C., 2006. Prosodic phrasing is central to language comprehension. Trends Cogn. Sci. 10, 244–249. doi:10.1016/j.tics.2006.04.002.
Friederici, A.D., Alter, K., 2004. Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang. 89, 267–276. doi:10.1016/S0093-934X(03)00351-1.
Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwander, A., 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. Proc. Natl. Acad. Sci. U. S. A. 103, 2458–2463. doi:10.1073/pnas.0509389103.
Frost, R.L.A., Monaghan, P., Tatsumi, T., 2017. Domain-general mechanisms for speech segmentation: the role of duration information in language learning. J. Exp. Psychol. Hum. Percept. Perform. 43, 466–476. doi:10.1037/xhp0000325.
Fruhholz, S., Gschwind, M., Grandjean, D., 2015. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. Neuroimage 109, 27–34. doi:10.1016/j.neuroimage.2015.01.016.
Ghitza, O., 2020. "Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019). Lang. Cognit. Neurosci. 35, 1100–1105. doi:10.1080/23273798.2020.1737720.
Ghitza, O., 2017. Acoustic-driven delta rhythms as prosodic markers. Lang. Cognit. Neurosci. 32, 545–561. doi:10.1080/23273798.2016.1232419.
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. doi:10.1159/000208934.
Gilbert, A.C., Boucher, V.J., Jemel, B., 2015. The perceptual chunking of speech: a demonstration using ERPs. Brain Res. 1603, 101–113. doi:10.1016/j.brainres.2015.01.032.
Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. doi:10.1038/nn.3063.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hamalainen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 267. doi:10.3389/fnins.2013.00267.
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hamalainen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. doi:10.1016/j.neuroimage.2013.10.027.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11, e1001752. doi:10.1371/journal.pbio.1001752.
Hanulíková, A., van Alphen, P.M., van Goch, M.M., Weber, A., 2012. When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. J. Cogn. Neurosci. 24, 878–887. doi:10.1162/jocn_a_00103.
Henke, L., Meyer, L., 2021. Endogenous oscillations time-constrain linguistic segmentation: cycling the garden path. Cereb Cortex 31, 4289–4299. doi:10.1093/cercor/bhab086.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. doi:10.1038/nrn2113.
Horvath, J., Muller, D., Weise, A., Schroger, E., 2010. Omission mismatch negativity builds up late. Neuroreport 21, 537–541. doi:10.1097/WNR.0b013e3283398094.
Huddleston, R., Pullum, G.K., 2002. The Cambridge Grammar of the English Language. Cambridge University Press, Cambridge doi:10.1017/9781316423530.
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. doi:10.1162/jocn_a_00001.
Inbar, M., Grossman, E., Landau, A.N., 2020. Sequences of intonation units form a ~ 1 Hz rhythm. Sci. Rep. 10, 15846. doi:10.1038/s41598-020-72739-4.
Itzhak, I., Pauker, E., Drury, J.E., Baum, S.R., Steinhauer, K., 2010. Event-related potentials show online influence of lexical biases on prosodic processing. Neuroreport 21, 8–13. doi:10.1097/WNR.0b013e328330251d.
Jin, P., Lu, Y., Ding, N., 2020. Low-frequency neural activity reflects rule-based chunking during speech listening. eLife 9, e55613. doi:10.7554/eLife.55613.
Kaltenböck, G., Heine, B., Kuteva, T., 2011. On thetical grammar. Stud. Lang. 35, 852–897. doi:10.1075/sl.35.4.03kal.
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. J. Neurosci. 40, 9467–9475. doi:10.1523/JNEUROSCI.0302-20.2020.
Kauppi, J.P., Pajula, J., Niemi, J., Hari, R., Tohka, J., 2017. Functional brain segmentation using inter-subject correlation in fMRI. Hum. Brain Mapp. 38, 2643–2665. doi:10.1002/hbm.23549.
Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. 16, e2004473. doi:10.1371/journal.pbio.2004473.
Kerkhofs, R., Vonk, W., Schriefers, H., Chwilla, D.J., 2007. Discourse, syntax, and prosody: the brain reveals an immediate interaction. J. Cogn. Neurosci. 19, 1421–1434. doi:10.1162/jocn.2007.19.9.1421.
Knight, R.T., 1984. Decreased response to novel stimuli after prefrontal lesions in man. Electroencephalogr. Clin. Neurophysiol. 59, 9–20. doi:10.1016/0168-5597(84)90016-9.
Knight, R.T., Grabowecky, M.F., Scabini, D., 1995. Role of human prefrontal cortex in attention control. Adv. Neurol. 66, 21–34 discussion 34-36.
Knösche, T.R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O.W., Friederici, A.D., 2005. Perception of phrase structure in music. Hum. Brain Mapp. 24, 259–273. doi:10.1002/hbm.20088.
Kulesa, A., Krzywinski, M., Blainey, P., Altman, N., 2015. Sampling distributions and the bootstrap. Nat. Methods 12, 477–478. doi:10.1038/nmeth.3414.
Leech, G., 2000. Grammars of spoken English: new outcomes of corpus-oriented research. Lang. Learn. 50, 675–724. doi:10.1111/0023-8333.00143.
Lerner, Y., Honey, C.J., Silbert, L.J., Hasson, U., 2011. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. 31, 2906–2915. doi:10.1523/JNEUROSCI.3684-10.2011.
Mattys, S.L., Pleydell-Pearce, C.W., Melhorn, J.F., Whitecross, S.E., 2005. Detecting silent pauses in speech: a new tool for measuring on-line lexical and semantic processing. Psychol. Sci. 16, 958–964. doi:10.1111/j.1467-9280.2005.01644.x.
Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal Gyrus. Science 343, 1006–1010. doi:10.1126/science.1245994.
Morillon, B., Arnal, L.H., Schroeder, C.E., Keitel, A., 2019. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci. Biobehav. Rev. 107, 136–142. doi:10.1016/j.neubiorev.2019.09.012.
Nakano, H., Rosario, M.A., Oshima-Takane, Y., Pierce, L., Tate, S.G., 2014. Electrophysiological response to omitted stimulus in sentence processing. Neuroreport 25, 1169–1174. doi:10.1097/WNR.0000000000000250.
Nguyen, M., Vanderwal, T., Hasson, U., 2019. Shared understanding of narratives is correlated with shared neural responses. Neuroimage 184, 161–170. doi:10.1016/j.neuroimage.2018.09.010.
Nora, A., Faisal, A., Seol, J., Renvall, H., Formisano, E., Salmelin, R., 2020. Dynamic time-locking mechanism in the cortical representation of spoken words. eNeuro 7. doi:10.1523/ENEURO.0475-19.2020, ENEURO.0475-19.2020.
Ono, T., Thompson, S.A., 1995. What can conversation tell us about syntax? Current Issues in Linguistic Theory. John Benjamins, Amsterdam.
Overath, T., McDermott, J.H., Zarate, J.M., Poeppel, D., 2015. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nat. Neurosci. 18, 903–911. doi:10.1038/NN.4021.
Pannekamp, A., Toepel, U., Alter, K., Hahne, A., Friederici, A.D., 2005. Prosody-driven sentence processing: an event-related brain potential study. J. Cogn. Neurosci. 17, 407–421. doi:10.1162/0898929053279450.
Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23, 1378–1387. doi:10.1093/cercor/bhs118.

Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. doi:10.1016/0013-4694(89)90180-6.

Phipson, B., Smyth, G.K., 2010. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol. 9. doi:10.2202/1544-6115.1585, Article 39.

Pihko, E., Leppasaari, T., Leppanen, P., Richardson, U., Lyytinen, H., 1997. Auditory event-related potentials (ERP) reflect temporal changes in speech stimuli. Neuroreport 8, 911–914. doi:10.1097/00001756-199703030-00019.

Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. Nat. Rev. Neurosci. 21, 322–334. doi:10.1038/s41583-020-0304-4.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. doi:10.1016/j.clinph.2007.04.019.

Puoliväli, T., Palva, S., Palva, J.M., 2020. Influence of multiple hypothesis testing on reproducibility in neuroimaging research: a simulation study and Python-based software. J. Neurosci. Methods 337, 108654. doi:10.1016/j.jneumeth.2020.108654.

Pylkkänen, L., 2019. The neural basis of combinatory syntax and semantics. Science 366, 62–66. doi:10.1126/science.aax0050.

Raij, T., McEvoy, L., Makela, J.P., Hari, R., 1997. Human auditory cortex is activated by omissions of auditory stimuli. Brain Res. 745, 134–143. doi:10.1016/s0006-8993(96)01140-7.

Rauschecker, J.P., 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear. Res. 271, 16–25. doi:10.1016/j.heares.2010.09.001.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. doi:10.1038/nn.2331.

Rimmele, J.M., Poeppel, D., Ghitza, O., 2021. Acoustically driven cortical δ oscillations underpin prosodic chunking. eNeuro 8. doi:10.1523/ENEURO.0562-20.2021, ENEURO.0562-20.2021.

Saalasti, S., Alho, J., Bar, M., Glerean, E., Honkela, T., Kauppila, M., Sams, M., Jaaskelainen, I.P., 2019. Inferior parietal lobule and early visual areas support elicitation of individualized meanings during narrative listening. Brain Behav. 9, e01288. doi:10.1002/brb3.1288.

Sammler, D., Cunitz, K., Gierhan, S.M.E., Anwander, A., Adermann, J., Meixensberger, J., Friederici, A.D., 2018. White matter pathways for prosodic structure building:a case study. Brain Lang. 183, 1–10. doi:10.1016/j.bandl.2018.05.001.

Sammler, D., Grosbras, M.H., Anwander, A., Bestelmeyer, P.E., Belin, P., 2015. Dorsal and ventral pathways for prosody. Curr. Biol. 25, 3079–3085. doi:10.1016/j.cub.2015.10.009.

Schafer, A.J., Speer, S.R., Warren, P., White, S.D., 2000. Intonational disambiguation in sentence production and comprehension. J. Psycholinguist. Res. 29, 169–182. doi:10.1023/a:1005192911512.

Scharinger, M., Steinberg, J., Tavano, A., 2017. Integrating speech in time depends on temporal expectancies and attention. Cortex 93, 28–40. doi:10.1016/j.cortex.2017.05.001.

Schiel, F., 1999. Automatic phonetic transcription of non-prompted speech. In: Ohala, J.J. (Ed.), Presented at the 14th International Congress of Phonetic Sciences, San Francisco, pp. 607–610. doi:10.5282/ubm/epub.13682.

Silbert, L.J., Honey, C.J., Simony, E., Poeppel, D., Hasson, U., 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. Proc. Natl. Acad. Sci. U. S. A. 111, E4687–E4696. doi:10.1073/pnas.1323812111.

Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., Hasson, U., 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. Nat. Commun. 7, 12141. doi:10.1038/ncomms12141.

Sinclair, J.M., Mauranen, A., 2006. Linear unit grammar: integrating speech and writing. Studies in Corpus Linguistics. John Benjamins, Amsterdam.

Stehwien, S., Meyer, L., 2021. Rhythm comes, rhythm goes: short-term periodicity of prosodic phrasing. PsyArXiv doi:10.31234/OSF.IO/C9SGB.

Steinhauer, K., 2003. Electrophysiological correlates of prosody and punctuation. Brain Lang. 86, 142–164. doi:10.1016/s0093-934x(02)00542-4.

Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. 2, 191–196. doi:10.1038/5757.

Steinhauer, K., Friederici, A.D., 2001. Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. J. Psycholinguist. Res. 30, 267–295. doi:10.1023/a:1010443001646.

Suni, A., 2017. Wavelet Prosody Toolkit.

Suni, A., Simko, J., Aalto, D., Vainio, M., 2017. Hierarchical representation and estimation of prosody using continuous wavelet transform. J. Comput. Speech Lang. 45, 123–136.

Szczepanski, S.M., Knight, R.T., 2014. Insights into human behavior from lesions to the prefrontal cortex. Neuron 83, 1002–1018. doi:10.1016/j.neuron.2014.08.011.

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759–1768. doi:10.1088/0031-9155/51/7/008.

Teng, X., Tian, X., Doelling, K., Poeppel, D., 2018. Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. Eur. J. Neurosci. 48, 2770–2782. doi:10.1111/ejn.13742.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., Naatanen, R., 1994. Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biol. Psychol. 38, 157–167. doi:10.1016/0301-0511(94)90036-1.

Van Berkum, J.J.A., van den Brink, D., Tesink, C.M.J.Y., Kos, M., Hagoort, P., 2008. The neural integration of speaker and message. J. Cogn. Neurosci. 20, 580–591. doi:10.1162/jocn.2008.20054.

Vetchinnikova, S., Mauranen, A., Mikusova, N., 2017. ChunkitApp: Investigating the relevant units of online speech processing. In: Presented at the Interspeech 2017: 18th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, pp. 811–812.

Weinberg, H., Walter, W.G., Cooper, R., Aldridge, V.J., 1974. Emitted cerebral events. Electroencephalogr. Clin. Neurophysiol. 36, 449–456. doi:10.1016/0013-4694(74)90201-6.

Weinberg, H., Walter, W.G., Crow, H.J., 1970. Intracerebral events in humans related to real and imaginary stimuli. Electroencephalogr. Clin. Neurophysiol. 29, 1–9. doi:10.1016/0013-4694(70)90074-x.

Westbury, C.F., Zatorre, R.J., Evans, A.C., 1999. Quantifying variability in the planum temporale: a probability map. Cereb Cortex 9, 392–405. doi:10.1093/cercor/9.4.392.

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., van den Bosch, A., 2016. Prediction during natural language comprehension. Cereb Cortex 26, 2506–2516. doi:10.1093/cercor/bhv075.

Wilson, S.M., Galantucci, S., Tartaglia, M.C., Rising, K., Patterson, D.K., Henry, M.L., Ogar, J.M., DeLeon, J., Miller, B.L., Gorno-Tempini, M.L., 2011. Syntactic processing depends on dorsal language tracts. Neuron 72, 397–403. doi:10.1016/j.neuron.2011.09.014.

Yabe, H., Tervaniemi, M., Reinikainen, K., Naatanen, R., 1997. Temporal window of integration revealed by MMN to sound omission. Neuroreport 8, 1971–1974. doi:10.1097/00001756-199705260-00035.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R.J., Naatanen, R., 1998. Temporal window of integration of auditory information in the human brain. Psychophysiology 35, 615–619. doi:10.1017/s0048577298000183.

Yan, X., Maeda, Y., Lv, J., Ginther, A., 2016. Elicited imitation as a measure of second language proficiency: a narrative review and meta-analysis. Lang. Test. 33, 497–528. doi:10.1177/0265532215594643.

Zora, H., Rudner, M., Montell Magnusson, A.K., 2020. Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. Eur. J. Neurosci. 51, 2236–2249. doi:10.1111/ejn.14658.