

Intonation Units in Spontaneous Speech Evoke a Neural Response

Maya Inbar,^{1,2,3} Shir Genzer,² Anat Perry,² Eitan Grossman,¹ and Ayelet N. Landau^{2,3}

¹Department of Linguistics, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel, ²Department of Psychology, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel, and ³Department of Cognitive and Brain Sciences, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel

Spontaneous speech is produced in chunks called intonation units (IUs). IUs are defined by a set of prosodic cues and presumably occur in all human languages. Recent work has shown that across different grammatical and sociocultural conditions IUs form rhythms of ~ 1 unit per second. Linguistic theory suggests that IUs pace the flow of information in the discourse. As a result, IUs provide a promising and hitherto unexplored theoretical framework for studying the neural mechanisms of communication. In this article, we identify a neural response unique to the boundary defined by the IU. We measured the EEG of human participants (of either sex), who listened to different speakers recounting an emotional life event. We analyzed the speech stimuli linguistically and modeled the EEG response at word offset using a GLM approach. We find that the EEG response to IU-final words differs from the response to IU-nonfinal words even when equating acoustic boundary strength. Finally, we relate our findings to the body of research on rhythmic brain mechanisms in speech processing. We study the unique contribution of IUs and acoustic boundary strength in predicting delta-band EEG. This analysis suggests that IU-related neural activity, which is tightly linked to the classic Closure Positive Shift (CPS), could be a time-locked component that captures the previously characterized delta-band neural speech tracking.

Key words: delta-band speech tracking; electroencephalography; general linear model; intonation units; speech prosody; spontaneous speech processing

Significance Statement

Linguistic communication is central to human experience, and its neural underpinnings are a topic of much research in recent years. Neuroscientific research has benefited from studying human behavior in naturalistic settings, an endeavor that requires explicit models of complex behavior. Usage-based linguistic theory suggests that spoken language is prosodically structured in intonation units. We reveal that the neural system is attuned to intonation units by explicitly modeling their impact on the EEG response beyond mere acoustics. To our understanding, this is the first time this is demonstrated in spontaneous speech under naturalistic conditions and under a theoretical framework that connects the prosodic chunking of speech, on the one hand, with the flow of information during communication, on the other.

Introduction

Research in cognitive neuroscience highlights an important role for temporal interactions between brain activity and speech. Neural activity tracks speech moment by moment at different time scales (Giraud and Poeppel, 2012; Gross et al., 2013; Obleser and Kayser, 2019). Neural speech tracking has been found in the theta band (4–8 Hz). This temporal scale is thought to reflect neural tracking of syllables, which across languages tend to have a similar temporal structure of four to eight syllables per second (Greenberg et al., 2003; Chandrasekaran et al., 2009; Pellegrino et al., 2011; Ding et al., 2017). Neural speech tracking has also been found in the delta band (1–2 Hz); however, in this band there is less understanding of what speech components are being tracked. Studies that investigate speech tracking of natural speech have associated delta-band tracking with prosody (Gross et al., 2013; Park et al., 2015; Keitel et al., 2017, 2018; Teoh et al.,

Received Feb. 8, 2023; revised Aug. 16, 2023; accepted Aug. 29, 2023.

Author contributions: M.I., E.G., and A.N.L. designed research; S.G. and A.P. contributed data; M.I., E.G., and A.N.L. analyzed data; M.I., E.G., and A.N.L. wrote the paper; M.I., S.G., A.P., E.G., and A.N.L. edited the paper.

A.N.L. was supported by the James McDonnell Scholar Award in Understanding Human Cognition, Israel Science Foundation (ISF) Grants 958/16 and 1899/21, HORIZON EUROPE European Research Council Grant 852387, Joy Ventures Research Grant and the Product Academy Award. M.I. was supported by the Humanities Fund PhD Program in Linguistics; Jack, Joseph, and Morton Mandel School for Advanced Studies in the Humanities; and an Azrieli Graduate Studies Fellowship. E.G. was supported by ISF Grant 2765/21. A.N.L. thanks the Mandel Scholion Research Center of the Hebrew University of Jerusalem for its support of the Evolution of Attention research group. We thank Shira Inbar for illustrating the EEG cap; Shlomi Frige for help with linguistic annotation; Meir Horovitz for help with the EEG preprocessing effort; Nir Ofir for developing linear modeling scripts, plotting scripts, and providing comments on the manuscript; and Nadav Matalon, Yael Maschler, and members of the Brain, Attention, and Time lab for discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Maya Inbar at maya.inbar@gmail.com or Ayelet N. Landau at ayelet.landau@mail.huji.ac.il.

<https://doi.org/10.1523/JNEUROSCI.0235-23.2023>

Copyright © 2023 the authors

2019). However, they do not present a detailed analysis of which prosodic modulations or what function they might serve in the language system. Another class of studies investigates tracking of synthesized language, which is structured in time and includes controlled prosodic modulation (Ding et al., 2016; Kaufeld et al., 2020; Bai et al., 2022). The finding in this case demonstrates convincing delta-band neural speech tracking of syntactic structure and meaning, but it raises two main difficulties. First, it is not obvious whether or how abstract linguistic structure in spontaneous speech is structured in time, and given the substantial variation between language systems, it may be hard to generalize this type of finding to a cross-linguistic mechanism. Second, listeners perceive prosodic breaks at syntactic boundaries even when no acoustic cue for such a break exists (Cole et al., 2010; Buxó-Lugo and Watson, 2016). Thus, it is possible that delta-band neural speech tracking in these studies reflects some type of prosodic processing, although no prosodic cues are provided in the stimuli (Gilbert et al., 2015; Glushko et al., 2022; Henke and Meyer, 2021).

Linguistic theory identifies a set of prosodic characteristics that fulfill the purpose of chunking speech cross-linguistically. These characteristics include an acceleration–deceleration dynamic of syllable delivery, resets in the slow modulations of pitch and volume, and at times, pauses. The resulting speech chunks are referred to as intonation units (IUs; Chafe, 1994; Du Bois et al., 1992). Similarly defined chunks in other approaches include intonation(al) phrases (Couper-Kuhlen and Bartch-Weingarten, 2011; Himmelmann et al., 2018; Seifart et al., 2021; Shattuck-Hufnagel and Turk, 1996), intonation groups (Cruttenden, 1997), tone groups (Halliday, 1967), and elementary discourse units (Kibrik, 2019). As IUs are defined by a set of auditory cues, attuned listeners can identify them in languages they do not understand (Chafe, 1994; Himmelmann et al., 2018). However, discourse-oriented qualitative linguistic research suggests a functional role of IUs; upon analyzing the progression of natural speech in context, IUs appear to pace the progression of the discourse in such a way that each IU contains a maximum of one new idea (Chafe, 1987, 1994, 2018; Himmelmann et al., 2018; Matsumoto, 2000; Pawley and Syder, 2000). An example transcript in Extended Data 1-1 allows to appreciate the progression of IUs in a spontaneously recounted personal narrative (Labov and Waletzky, 1967). IUs are also a useful resource for sequence organization in conversation and for constructing different speech actions (Bögels and Torreira, 2015; Ford and Thompson, 1996; Gravano and Hirschberg, 2011; Selting, 2010; cf. Szczepek-Reed, 2010). In addition, we found that sequences of IUs form a ~ 1 Hz rhythm in six languages from around the world, despite substantial variation in the grammatical structures and sociocultural profiles of these languages (Inbar et al., 2020; Stehwen and Meyer, 2022; cf. Tilsen and Arvaniti, 2013). We propose that studying the neural response to IUs might enrich our models of speech processing. They capture prosodic variation that is presumably cross-linguistically associated with the structuring of the discourse in spontaneous speech, and share temporal structure with previously found delta-band neural speech tracking.

Here, we study the neural response to IUs using EEG and a general linear model (GLM) statistical framework (Fig. 1). We use linguistic theory to identify IUs in spontaneous speech. Our analytic approach is similar to the traditional event-related potentials in EEG research (Smith and Kutas, 2015). However, instead of experimentally triggering the brain to synthesized speech, we use naturally occurring words as triggers and

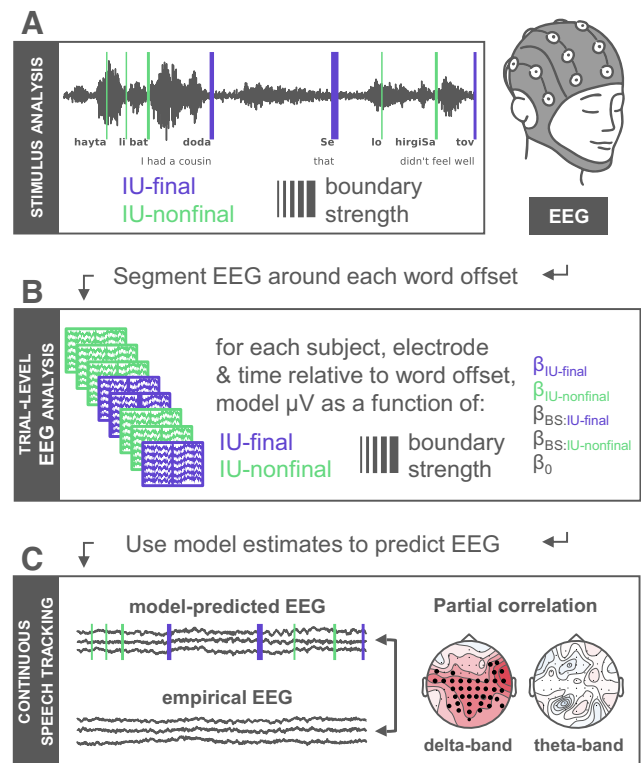


Figure 1. Methodology overview. **A**, Spontaneous speech material in Hebrew was analyzed linguistically and acoustically (Audio 1). Each orthographic word was time stamped (vertical lines relative to the audio waveform at word offsets), annotated as to whether it closed an IU or not (purple and green, respectively), and assigned an acoustic-based boundary strength score (illustrated by line width). Participants listened to this speech material while their EEG was recorded. **B**, We segmented the EEG around word offset times and modeled it as a function of IU closure and BS. Models were fitted to broadband EEG as well as delta- and theta-band EEG. **C**, We used subsets of model estimates to predict the continuous EEG response to specific stimulus parameters. In each band, we studied the unique predictive accuracy of IU closure and boundary strength, as measured by partial correlation.

investigate the neural response to each word as a function of the labeling of IUs. In addition, we add an objective measure of acoustic boundary strength (BS) at the word level to complement the subjective labeling of IUs. This measure enables us to characterize the IU impact on neural response while equating acoustic variation. Characterizing the neural response to IUs above and beyond acoustic boundary strength enables inference on the added cognitive and perceptual processes that are involved in the processing of speech. In addition, we hypothesized that from a continuous perspective, the time course of neural responses to IUs would give rise to delta-band neural speech tracking. We use the estimated responses from the GLM model fitted to band-limited EEG to predict unseen continuous EEG responses to our stimuli. Both IU closure and boundary strength contributed uniquely to predicting delta-band EEG but not theta-band EEG.

Materials and Methods

We analyzed EEG recordings of participants listening to different speakers describing an emotional life event. This data were collected as part of an independent project and summarized in detail in Genzer et al. (2022).

Participants

Genzer et al. (2022) recorded EEGs from 57 Hebrew-speaking undergraduate students from the Hebrew University of Jerusalem. The participants

Table 1. Summary information of the speech recordings and annotations

No.	Duration (s)	Speaker gender	Median F0	No. of words	No. IUs	No. of clause-ending words	Words/s	IUs/s	Words/IU
1	168.64	F	179.07	337	139	40 (5 IU-nonfinal)	2.00	0.82	2.42
2	149.64	F	149.43	371	163	52 (0)	2.48	1.09	2.28
3	125.64	M	118.54	287	129	36 (1)	2.28	1.03	2.22
4	146.44	F	167.00	314	163	37 (0)	2.14	1.11	1.92
5	172.6	M	131.52	299	154	32 (2)	1.73	0.89	1.94
6	135.56	F	200.42	229	109	37 (0)	1.69	0.80	2.10
7	170.36	F	180.06	383	172	59 (1)	2.25	1.01	2.23
8	228.68	M	87.81	506	231	55 (1)	2.21	1.01	2.19
9	120.88	M	91.57	303	129	34 (1)	2.51	1.07	2.34

F, Female; M, Male.

received monetary compensation at a rate of 40 Israeli new shekels per hour (~\$15) or course credit. All participants reported normal or corrected-to-normal visual acuity and had no history of psychiatric or neurological disorders. Participants gave their informed consent before the experimental session. The study was approved by the Institutional Review Board of Ethical Conduct at the Hebrew University of Jerusalem. We excluded from further analysis the data of five participants because of recording failures and two participants because of excessive amounts of artifacts (see below, EEG preprocessing), resulting in a final sample of 50 participants (28 female), age 23.7 ± 1.89 years (mean \pm SD).

Stimuli

Each participant listened, through speakers, to three of nine stories from an Israeli Empathic Accuracy stimulus set (Jospe et al., 2020). In this stimulus set, Hebrew speakers described emotional life events as they sat in front of a professional recorder. The duration of the stories was between 2:01 and 3:48 min, with an average duration of 2:43 min. The nine stories were grouped into sets of three, and the sets were of approximately equal duration (range, 454–520 s). The assignment of story set to participant was random and counterbalanced. Details on experimental design and EEG acquisition can be found in Genzer et al. (2022).

Expert stimulus annotation

Two trained native Hebrew-speaking annotators transcribed the nine stories that served as stimuli and segmented them into IUs according to the criteria devised by Chafe (1994) and Du Bois et al. (1992). The segmentation process involves close listening to the rhythmic and melodic cues for IU boundaries as well as performing acoustic analyses in Praat software (Boersma and Weenink, 2022) for the extraction of pitch contours, which are used to support perceived resets in pitch. We previously described this process in Inbar et al. (2020). The annotators checked each other's segmentations and reached consensus at ambiguous points.

In addition, two native Hebrew-speaking linguists annotated clauses in the recordings and reached a consensus version reflecting a traditional approach to Hebrew syntax (Ravid, 1999). Extended Data 1-1 includes a detailed definition of clausehood with examples.

Next, we time stamped each orthographic word in the transcription relative to the recording and coded whether it appeared at the end of an IU or not and whether it appeared at the end of a clause or not, according to the two segmentation processes described above. Note that this coding does not represent hierarchy in clause structure, and it does not distinguish between words that close a single clause and words that are the end point of two or three clauses (e.g., when the final constituent of a clause includes a relative clause). The time stamping was done in Praat (Boersma and Weenink, 2022), imported ELAN software (The Language Archive, 2022), and exported from there into a single text file including all the words from all recordings. We analyzed relations between annotations and computed descriptive summary information about each recording with the aid of custom-written scripts in R software (R Core Team, 2022). Table 1 presents summary information about the recordings.

Acoustic boundary strength score

Each orthographic word was given a score that indicated the strength of the boundary at its offset. The score was estimated using a published

algorithm, the Wavelet Prosody Toolkit (Sun et al., 2017). This algorithm derives fundamental frequency and intensity information from the speech audio signal and duration information from time-stamped word annotations. Thus, the algorithm operates precisely on the signals that correspond to the main cues for an IU boundary. These signals are summed to one and decomposed to several scales using a continuous wavelet transform. The algorithm then finds peaks and troughs in the output of the continuous wavelet transform. Based on these peaks and troughs, the algorithm defines for each annotated word (see above) a prominence value and a boundary strength value (of which we only used the boundary strength value). In general, the prominence value of a word is defined as the strongest peak within the word. The boundary strength value of a word is defined as the strongest trough between two peaks, namely, between the strongest peak within the word and the strongest peak within the next word.

We provided as input to the algorithm the speech audio files and the time-stamped word annotations and applied the algorithm with default settings. Our procedure deviated from that described in Sun et al. (2017) in a single respect; because we annotated audible breaths in the stories as single-word IUs, in our application of the algorithm we also assigned boundary strength scores to breaths.

Relation between IU closure, boundary strength scores, pause durations, and clause closure

To quantify the difference in boundary strength scores between IU-final words and IU-nonfinal words we fitted a mixed-effects linear model, predicting boundary strength score from IU closure condition. The model included a by-story random intercept as well as a by-story random slope for the effect of IU closure condition (IU-final and IU-nonfinal) to account for slight variations between the stories. The explanatory power of the model related to the fixed effect alone is 0.30. The intercept of the model, corresponding to the average boundary strength score of IU-nonfinal words, is estimated at 0.16, 95% CI [0.13, 0.18]. IU-final words have, on average, a significantly larger boundary strength score, with a difference estimated at 0.55, 95% CI [0.49, 0.61], $t_{(3023)} = 17.07$, $p < 0.001$.

To quantify the relationship between boundary strength scores and pause durations we fitted a mixed-effects linear model, predicting the standardized boundary strength scores from the standardized durations of the pauses from each word to the next. The model included a by-story random intercept as well as a by-story random slope for the effect of pause duration to account for slight variations between the stories. The total explanatory power of the model is 0.45 (conditional R^2) and the part related to the fixed effect alone (marginal R^2) is 0.36. The intercept of the model, corresponding to the average boundary strength score for the mean pause duration (mean and not zero because the variables were standardized), is estimated at 0.01, 95% CI [−0.04, 0.07]. The pause to the next word is significantly and positively related to the boundary strength score ($\beta = 0.67$, 95% CI [0.46, 0.88], $t_{(3014)} = 6.25$, $p < 0.001$).

To quantify the difference in boundary strength scores between clause-final and clause-nonfinal words we fitted a mixed-effects linear model to the data of IU-final words only, predicting boundary strength score from clause closure condition. We modeled only the data of IU-final words because there were nearly no words that ended clauses and

were IU-nonfinal (Table 1). The model included a by-story random intercept as well as a by-story random slope for the effect of clause (C) closure condition (C-final and C-nonfinal), to account for slight variations between the stories. The explanatory power of the model related to the fixed effect alone is 0.02. The intercept of the model, corresponding to the average boundary strength score of C-nonfinal words (within IU-final words), is estimated at 0.75, 95% CI [0.72, 0.79]. C-final words have, on average, a significantly weaker boundary strength score, with a difference estimated at -0.19 , 95% CI $[-0.29, -0.09]$, $t_{(1383)} = -3.69$, $p = 0.008$.

In all three models, including as control variables F0 (as measured by the median across the recording) and overall speech rate (as measured by words per second) did not change these results. Mixed-effect models were estimated using REML and nlptwrap optimizer with lme4 (Bates et al., 2015) in R software (R Core Team, 2022). Conditional and marginal R^2 were calculated with the aid of the MuMIn package (<https://cran.r-project.org/package=MuMIn>). The 95% CIs and p values were computed using a Wald t distribution approximation with the aid of the report package (<https://github.com/easystats/report>).

EEG preprocessing

We processed the 64-channel EEG in MATLAB (version R2018b, MathWorks) using the FieldTrip toolbox (Oostenveld et al., 2011) and custom scripts. We rereferenced the signal in the EEG channels to the average signal from the mastoid electrodes. We computed bipolar derivations of the horizontal and vertical pairs of electrodes around the eyes and appended the resulting horizontal and vertical EOG signals to the EEG channels. We segmented the data to blocks according to the presented clips, starting from 2 s before the story began until 2 s after the story ended. We removed slow drifts in the EEG using spline interpolation as implemented in the function `msbackadj` (part of the MATLAB Bioinformatics toolbox) and similarly to Ofir and Landau (2022). The window size was 4 s, with a step of 0.75 s between two consecutive windows. The median of each window was taken as the baseline value. We detected large artifacts in the EEG semiautomatically with the aid of the function `ft_artifact_zvalue`, padding each artifact with 500 ms on either side. After visually inspecting the detected artifacts, we substituted them with zeros. We detected and removed excessively bad channels by visual inspection. We then ran independent component analysis (ICA) using the infomax algorithm (maximum 1024 training steps, following principal component analysis, and saving 20 components). We identified components corresponding to eye- and muscle-related activity and removed them from the data. We replaced rejected electrodes with the plain average of all neighbors as implemented in `ft_channelrepair`. Next, we defined trials as epochs from 1000 ms before to 1000 ms after each word offset. Trials that overlapped with the semiautomatically detected artifacts were discarded. We removed the mean from each epoch as a baseline. We removed remaining artifactual trials after visual inspection guided by the summary statistics in the `ft_rejectvisual` function. We excluded from further analysis the data of two participants because of excessive amounts of artifacts that affected >15% of their trials. In the final sample of 50 participants, the average number of rejected ICA components was 2.74 ± 1.32 (mean \pm SD), the average number of interpolated channels was 1.1 ± 1.47 (mean \pm SD; maximum 6), and the average number of remaining trials was 966.5 ± 72.72 (mean \pm SD; range, 854–1072).

For visualizing the effect of IU closure (IU-final vs IU-nonfinal) for mean boundary strength (see Fig. 3C,D), the time domain data were processed as follows. After we removed artifactual trials, we standardized the per-trial boundary strength scores of each subject. Next, for each IU closure condition we selected trials with a boundary strength z score between ± 0.5 . This resulted in 138.46 ± 11.30 and 138.7 ± 9.9 (mean \pm SD) trials per participant per IU closure condition, (IU-final and IU-nonfinal, respectively; a two-tailed paired-samples t test indicated no significant difference between the trial counts, $t_{(49)} = -0.168$, $p = 0.87$). We averaged the data in each trial over two subsets of electrodes (six electrodes for the negative cluster, F8, FT8, T8, F6, FC6, C6; seven electrodes for the positive cluster, Cz, CPz, Pz, CP1, CP2, P1, P2), based on previous literature and in accord with the significant clusters we found

in the statistical modeling. Next, we applied a finite impulse response (FIR) windowed-sinc zero-phase low-pass filter, with 20 Hz as the cutoff frequency. Finally, we averaged the trials within IU closure condition and within participant, and then across participants.

For visualizing the effect of boundary strength within IU-final words (see Fig. 3G), the time-domain data were processed as follows. First, for IU-final words only, we binned the entire sample of boundary strength scores into quartiles. We averaged the data in each IU-final trial over a subset of electrodes (18 electrodes: Fz, F1, F2, F3, F4, F5, F6, F7, F8, FCz, FC1, FC2, FC3, FC4, FC5, FC6, FC7, FC8), based on the significant cluster we found in the statistical modeling. Next, we applied an FIR windowed-sinc zero-phase low-pass filter, with 20 Hz as the cutoff frequency. Finally, we averaged the trials within quartile and within-participant, and then across participants.

Statistical modeling

We analyzed the 2 s windows around each word offset using the hierarchical mass univariate general linear modeling approach implemented in the LIMO EEG toolbox (Pernet et al., 2011). At the first level, we modeled the EEG amplitude at each electrode and time point combination per participant using the following formula:

$$\mu V = \beta_1 \cdot IU_{final} + \beta_2 \cdot IU_{nonfinal} + \beta_3 \cdot BS \cdot IU_{final} + \beta_4 \cdot BS \cdot IU_{nonfinal} + \beta_0$$

This model includes a predictor for IU closure condition (IU-final vs IU-nonfinal) and a continuous linear predictor for boundary strength (BS), with a different slope depending on the IU closure condition. The predictor IU-final was one for all words that closed an IU and zero otherwise. The predictor IU-nonfinal was one for all words that did not close an IU and zero otherwise. Boundary strength scores were standardized across both IU closure conditions per participant so that boundary strength 0 is the mean boundary strength across all words. The model parameters, then, have the following interpretation: (1) β_1 , the effect of a word with mean boundary strength that closes an IU; (2) β_2 , the effect of a word with mean boundary strength that does not close an IU; (3) β_3 , the effect of increasing the boundary strength by 1 SD for IU-final words; (4) β_4 , the effect of increasing the boundary strength by 1 SD for IU-nonfinal words; and (5) β_0 , the predicted EEG amplitude around word offset after regressing out all other effects.

In total, 64×1025 models (65,600, 1 for each electrode and time point combination) were fitted to the data of each participant. The model parameters were estimated using a weighted least-squares method for each participant separately (Pernet et al., 2022). We assessed the extent of collinearity among predictor variables using variance inflation factors (VIFs). Specifically, we calculated the VIFs using each participant's weighted predictor matrix and found that the maximum VIF for any given predictor and participant did not exceed 1.55. After the model parameters were estimated for each time point and electrode for each participant, we computed two contrasts, that is, linear combinations of parameters that capture comparisons of interest, hereafter referred to as effects. The main contrast of interest was the IU closure effect, which we computed as the difference in $\beta_1 - \beta_2$ (responses to IU-final words minus responses to IU-nonfinal words, when the boundary strength equals the mean for that dataset). We computed an additional contrast, the difference in $\beta_3 - \beta_4$, to capture any difference between the slope of the regression against boundary strength for IU-final versus IU-nonfinal words. Finally, we tested the effect of boundary strength separately for IU-final (β_3) and IU-nonfinal (β_4) words.

As the four effects were computed at each electrode and time point for each participant, a single participant is represented by four spatio-temporal matrices of size 64×1025 , one per effect. Every element in each matrix, that is, every electrode time point combination of a participant, is a single observation in the statistical significance test of a specific effect. The significance of the contrasts is estimated at the group level, using a clustering approach to correct for multiple comparisons and following the LIMO EEG toolbox conventions (Pernet et al., 2011). For the two contrasts (IU closure effect and the difference in boundary strength

slopes between the different levels of the IU closure condition) we used an alpha of 0.05. For the effects of boundary strength within IU-final and IU-nonfinal trials we used an alpha of 0.025, to correct for *post hoc* comparisons.

Prediction of unseen continuous EEG data from GLM estimates

We used a subject-based leave-one-out cross-validation approach to test how well IU closure and boundary strength predict the continuous EEG signal independently of one another. To predict the continuous EEG of each participant, we first averaged the estimated model parameters across all other participants (i.e., β_1 , β_2 , β_3 , β_4 , and β_0 parameters for each electrode and time point combination). Then, we matrix multiplied these group-average parameter estimates with the left-out participant's predictor matrix (i.e., the IU closure annotations and boundary strength scores of each word in the story presented to this participant). This resulted in 2 s model-predicted EEG responses, centered on the offset of each word.

Each model-predicted EEG segment was embedded in a separate zero-filled EEG time course (at full story length) according to the original offset times of each word in the recording. All time courses were then summed. Note that the stage of matrix multiplication was done for a subset of predictors at a time, allowing to predict separate EEG responses for the different effects of interest (e.g., IU closure). Then, for each participant and effect, we calculated the partial Pearson correlation of the model-predicted continuous EEG with participants' empirical continuous EEG, conditioning out the predicted EEG based on the remaining effects. We averaged the resulting partial correlation maps across subjects to obtain a group-average map of the unique predictive accuracy of each effect.

We used a nonparametric permutation approach to assess whether the unique predictive accuracy of each effect exceeds chance level. Specifically, we compared the observed group-average maps of predictive accuracy per effect to a set of 100 group-average surrogate maps. Surrogate maps were constructed along the following lines: We circularly shifted each participant's predictor matrix by a random number 100 times. We used the original word offset times to predict the continuous EEG according to the shifted predictor matrix. This preserves the sequence of annotations but starts the sequences at random word offset relative to the original recording. We calculated the partial Pearson correlation of the empirical continuous EEG with the model-predicted continuous EEG, conditioning out the model-predicted EEG based on the remaining effects, and averaged the result across participants. For a given effect, we calculated the proportion of times the absolute values of the surrogate partial correlation maps exceeded the absolute values of the observed correlation map. This procedure yields a scalp topography of p values. We corrected p values for multiple comparisons across channels ensuring that on average, the false discovery rate will not exceed 5% (Genovese et al., 2002; Yekutieli and Benjamini, 2001).

Neural speech tracking in the delta and theta bands

We repeated the analysis with a focus on EEG activity in two frequency bands of interest, the delta band, between 0.8 and 1.1 Hz, and the theta band, between 3.5 and 5 Hz. These cutoff frequencies conform with recent neural speech tracking literature (Kaufeld et al., 2020), and in the case of the delta band, correspond to the actual minimal and maximal IU rate in our stimuli (Table 1). We filtered the empirical EEG in the two frequency bands using forward and reverse second-order Butterworth filters. For each bandpassed empirical EEG dataset, we fit the GLM model described earlier (see above, Statistical modeling). We used the coefficients estimated on the bandpassed data to compute the unique predictive accuracy of each speech effect in a given band as previously described (see above, Prediction of unseen continuous EEG data from GLM estimates).

The added effect of clause closure

Statistical modeling. To evaluate the added effect of clause closure we ran an additional statistical model. This model included two predictors. As in the previous model, boundary strength was modeled as a continuous linear predictor. Boundary strength scores were standardized across all words per participant such that boundary strength 0 is the mean

boundary strength across all words. In addition, we used a three-level categorical predictor to simultaneously code for IU closure and clause closure. This was achieved by coding whether a word both closes an IU and a clause, closes an IU but not a clause, or does not close neither an IU nor a clause. Words that end a clause but do not close an IU were too rare to allow for a meaningful estimation of the neural response to them (Table 1; see Fig. 5B) and so we did not include them in the analyses. This model was implemented for each electrode and time point combination per participants using the following formula:

$$\mu V = \beta_1 \cdot IU_{final} \cdot C_{final} + \beta_2 \cdot IU_{final} \cdot C_{nonfinal} + \beta_3 \cdot IU_{nonfinal} \cdot C_{nonfinal} + \beta_4 \cdot BS + \beta_0.$$

The model parameters, then, have the following interpretation: (1) β_1 , the effect of a word with mean BS that closes an IU and ends a clause; (2) β_2 , the effect of a word with mean BS that closes an IU and does not end a clause; (3) β_3 , the effect of a word with mean BS that does not close an IU or end a clause; (4) β_4 , the effect of increasing the boundary strength by 1 SD; and (5) β_0 , the predicted EEG amplitude around word offset after regressing out all other effects.

As for the original model, in total, 64×1025 models (65,600, 1 for each electrode and time point combination) were fitted to the data of each participant. The model parameters were estimated using a weighted least-squares method for each participant separately. After the model parameters were estimated for each time point and electrode for each participant, we computed two contrasts. The main contrast of interest was the IU closure effect, which we computed as the difference $(\beta_1 + \beta_2) / 2 - \beta_3$ (responses to IU-final words, either that end a clause or not, minus responses to IU-nonfinal words, when the boundary strength equals the mean for that dataset). We computed an additional contrast, the difference in $\beta_1 - \beta_2$, to capture any difference between words that end a clause and words that do not end a clause within IU-final words only. Finally, we tested the effect of boundary strength. As the three effects were computed at each electrode and time point for each participant, a single participant is represented by three spatiotemporal matrices of size 64×1025 , one per effect. Every element in each matrix, that is, every electrode-time point pair of a participant, is a single observation in the statistical significance test of a specific effect. The significance of the contrasts is estimated at the group level using an alpha of 0.05 and a clustering approach to correct for multiple comparisons.

Neural speech tracking in the delta and theta bands. To study the added effect of clausehood in capturing delta- and theta-band EEG activity we fit the GLM model (see above, The added effect of clause closure, Statistical modeling) to bandpassed EEG data. As before, we used forward and reverse second-order Butterworth filters, and the cutoff frequencies were 0.8 to 1.1 Hz and 3.5 to 5 Hz for the delta and theta bands, respectively. We used the coefficients estimated on the bandpassed data to compute the unique predictive accuracy of each speech effect in a given band as previously described (see above, Prediction of unseen continuous EEG data from GLM estimates).

Data availability

Custom scripts producing the analyses and figures are available in the Open Science Framework (OSF) repository: <https://osf.io/v8mtg>. Raw EEG data is available in the OSF repository: osf.io/k7bmw. For access to the stimulus set, contact Anat Perry.

Results

We analyzed EEG recordings of participants listening to different speakers describing an emotional life event. We transcribed the stories and manually identified IUs within each story. Each word was labeled as either closing an IU (IU-final) or not (IU-nonfinal). Additionally, we also computed for each word a boundary strength score based on an analysis of the speech envelope, fundamental frequency, and word duration information (see above, Materials and Methods). The distribution of boundary strength

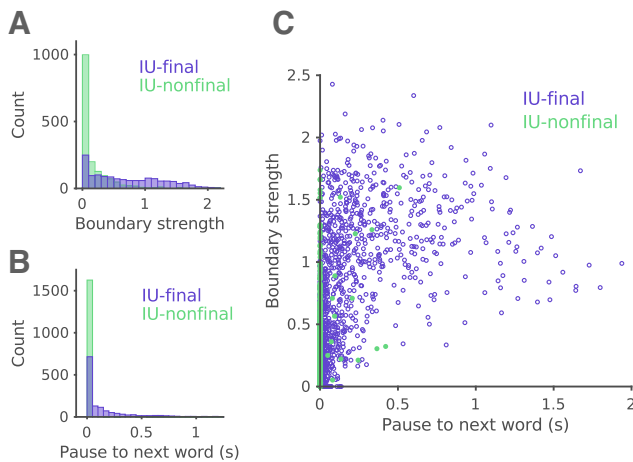


Figure 2. Relationship between boundary strength, pauses, and IU closure. **A**, Distribution of boundary strength scores in IU-final and IU-nonfinal words (purple and green, respectively). Histogram bins span 0.1 units of boundary strength. **B**, Distribution of pauses from each word to the next in IU-final and IU-nonfinal words (purple and green, respectively). Histogram bins span 0.05 s. **C**, Distribution of boundary strength scores relative to pause duration in IU-final and IU-nonfinal words (purple and green, respectively).

scores across the two different word types, IU-final and IU-nonfinal, is presented in Figure 2A. As expected, words that close an IU tend to have stronger boundaries ($\beta = 0.55$, $SE = 0.03$, $p < 0.001$). For comparison, we also present the distribution of pauses across the different word types and the relation between pause duration and boundary strength score (Fig. 2B,C). It is noteworthy that pause duration accounts for at most 44.5% of the overall variance in boundary strength scores, and that 51.48% of IU-final words are followed by a pause of 50 ms or less (see above, Materials and Methods).

EEG amplitude depends on IU closure and boundary strength

We modeled the contribution of two factors to the EEG response at a word offset, namely, whether a word closed an IU or not (IU-final and IU-nonfinal, respectively) and its boundary strength. To this end, we used a hierarchical mass univariate GLM approach (Pernet et al., 2011; see above, Materials and Methods). In this framework, at the first level, the data of each participant is modeled using the IU closure predictor (a categorical predictor, IU-final or IU-nonfinal) and the boundary strength predictor (a continuous predictor, standardized across both levels of the categorical predictor). The effect of boundary strength is modeled separately in each level of the IU closure condition to accommodate the possibility that the effect differs between IU-final and IU-nonfinal words (see above, Materials and Methods). Beta coefficients are estimated for each electrode and time point. At the second level, beta estimates are entered into a nonparametric cluster-based analysis to evaluate the significance of these effects at the group level over time and scalp location.

This analysis revealed an effect of IU closure and an effect of boundary strength in IU-final words on the EEG response around word offset. First, for words with equivalent and average boundary strength, words that close an IU are associated with a centroparietal positivity preceded by a right anterior negativity compared with words that do not close an IU. This was supported by the significant effect of IU closure (Fig. 3A–D; negative cluster extending from -76 to 225 ms following word offset, $p = 0.009$; positive cluster extending from 133 to 504 ms following

word offset, $p < 0.0001$). There was an additional positive cluster preceding the right anterior negativity, ~ 500 ms before word offset (Fig. 3A; cluster extending from -545 until -449 ms relative to word offset, $p = 0.006$). It is noteworthy that the IU closure contrast provided equivalent results when performed on non-standardized variables (not shown). In essence, this means that the effect of IU closure is also apparent when comparing IU-final and IU-nonfinal words with zero boundary strength rather than average boundary strength.

Next, for words that close an IU, stronger boundaries elicited a larger anterior negativity (Fig. 3E–G; *post hoc* test for boundary strength effect, two clusters extending from 119 to 311 ms and from 324 to 400 ms following word offset, $p < 0.0001$ and $p = 0.004$, respectively). Here too there was a positive cluster preceding the negativity ~ 200 ms before word offset (Fig. 3E; cluster extending from -197 until -141 ms relative to word offset, $p = 0.006$). For words that do not close an IU, the effect of boundary strength was only trending, perhaps because of the overall skewed distribution of boundary strength scores within IU-nonfinal words (Fig. 2A). Finally, we tested whether the effect of boundary strength differed between IU-final and IU-nonfinal words and found no significant difference (see above, Materials and Methods).

Delta band in continuous EEG tracks intonation units

Previous studies that characterized the temporal structure of IU sequences showed they form an ~ 1 Hz rhythm (Chafe 1994; Jun, 2005; Chafe, 2018; Inbar et al., 2020; Stehewien and Meyer, 2022). From a continuous perspective, the time course of responses to IUs should therefore give rise to neural activity at that temporal scale. Previous studies investigating ongoing speech have, in fact, demonstrated neural tracking of speech within the delta band (< 2 Hz). To link our findings to the neural speech tracking literature we repeated the modeling effort from above on bandpassed EEG data in two bands of interest, delta (0.8 – 1.1 Hz) and theta (3.5 – 5 Hz). We used the fitted model estimates and subsets of the model parameters to predict (i.e., reconstruct) the ongoing band-limited neural response to the stories. We measured the unique predictive accuracy of every subset of parameters, that is, the correlation between unseen empirical EEG and the predicted signals, while conditioning out the predicted response based on the rest of the predictors (partial correlation, measured in Pearson's r). Finally, to assess the unique predictive accuracy of IU closure and boundary strength in each band, we compared the observed partial correlation value in each channel to a permutation distribution of partial correlation values.

This analysis revealed that both IU closure and boundary strength contribute uniquely to predicting delta-band EEG but not theta-band EEG (Fig. 4). Specifically, the modeled response to the IU closure condition, which considers whether each word in the sequence closed an IU or not, allowed to predict continuous delta-band EEG activity in lateral to centroparietal electrodes. The predictive accuracy of IU closure was unique relative to the expected response given a sequence of word offsets and also relative to the expected response given their boundary strength scores. The modulation of the EEG signal based on the boundary strength score allowed to predict continuous delta-band EEG activity in anterior–lateral electrodes. The predictive accuracy of boundary strength was unique relative to the expected response at word offset and also relative to the expected response given the IU closure condition.

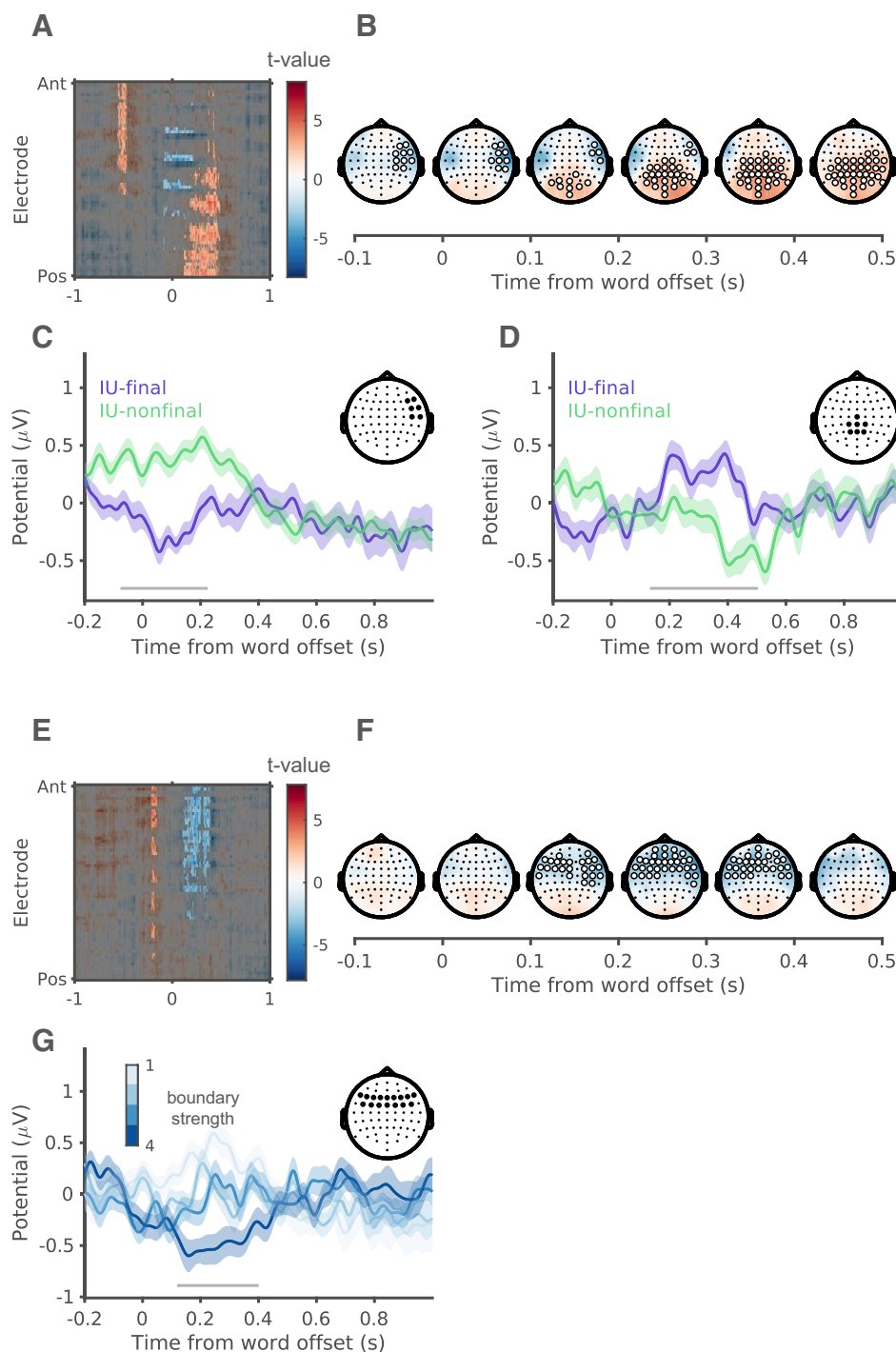


Figure 3. EEG amplitude depends on IU closure and boundary strength. **A**, The full spatiotemporal results for the IU closure effect. A t value is calculated for the observed difference between IU-final and IU-nonfinal EEG responses at each electrode and time point in a 2 s window around word offset. Cluster-based bootstrap tests reveal significant clusters at the group level (unmasked). **B**, Topographical distributions of the IU closure contrast t values, from -100 ms to 500 ms around word offset in 100 ms steps. Electrodes in significant clusters are highlighted in white if they span 15 consecutive ms. **C**, ERP traces in response to IU-final and IU-nonfinal words with mean boundary strength score (purple and green, respectively), illustrating the right anterior negative cluster. Shaded ribbons correspond to ± 1 SEM. Inset, Topography showing the EEG electrodes that were used to visualize the ERP traces. Bottom, The horizontal gray bar marks the time points over which this cluster was significant in the GLM model. **D**, ERP traces in response to IU-final and IU-nonfinal words with mean boundary strength score (purple and green, respectively), illustrating the centroparietal positive cluster. Shaded ribbons correspond to ± 1 SEM. Inset, Topography showing the EEG electrodes that were used to visualize the ERP traces. Bottom, The horizontal gray bar marks the time points over which this cluster was significant in the GLM model. **E**, The full spatiotemporal results for the effect of acoustic boundary strength within IU-final words. A t value is calculated for the estimated change in EEG response against zero at each electrode and time point in a 2 s window around word offset. Cluster-based bootstrap tests reveal significant clusters at the group level (unmasked). **F**, Topographical distributions of the effect of boundary strength within IU-final words from -100 ms to 500 ms around word offset in 100 ms steps. Electrodes in significant clusters are highlighted in white if they span 15 consecutive milliseconds. **G**, ERP traces in response to IU-final words with different levels of boundary strength, illustrating the anterior negative cluster. Four different levels are presented, corresponding to quartiles of boundary strength scores within IU-final words. Shaded ribbons correspond to ± 1 SEM. Inset, Topography showing the EEG electrodes that were used to visualize the ERP traces. Bottom, The horizontal gray bar marks the time points over which this cluster was significant in the GLM model.

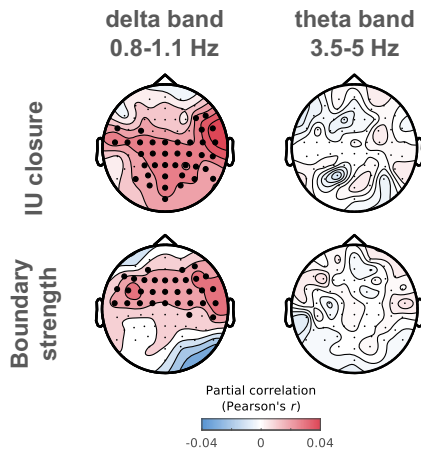


Figure 4. Relation of IU closure and boundary strength to delta- and theta-band EEG activity. Group average partial correlation maps of the unique predictive accuracy of IU closure (top) and boundary strength (bottom), when predicting left-out EEG data in the delta (left) and theta (right) bands. The electrodes in bold denote where the predictive accuracy was significantly higher than chance (corrected for multiple comparisons).

Clause closure and IU closure independently modulate EEG amplitude and delta-band activity

Chafe (1994) found that a large percentage of the IUs in his sample contained a full single clause. Similar observations were made in a variety of similar accounts (Givón, 2018; Pawley and Syder, 2000). Clauses can be said to assert the idea of an event or state. A relatively well-accepted definition of a clause is a constituent with a predicate and all the complements and modifiers related to it. Hence, its definition relies on syntactic terms in contrast to IUs, which are defined based on prosodic cues. Clauses are central to many theoretical approaches to language (Croft, 2022; Haspelmath, 2020; Thompson, 2019; Thompson and Couper-Kuhlen, 2005; cf. Laury et al., 2019). In a final set of analyses, we assessed to what extent the neural responses characterized above for IUs may be attributed to a clause-centered account. Under the assumption that both levels of description capture functionally relevant levels of processing, we modeled both clause closure and IU closure together. This enables an evaluation of their unique and joint impact on the neural response. We identified clauses in the spoken stories and labeled each word as either ending a clause or not (see above, Materials and Methods). The total of 23:38 min of spontaneous Hebrew speech that included 1039 IUs contained 518 clauses, of which 144 were produced as a single IU. Clause-ending words nearly always closed an IU as well (Table 1; Fig. 5B), yet they were characterized by an overall weaker boundary strength score compared with IU-final words that did not end clauses (Fig. 5A; $\beta = -0.19$, SE = 0.05, $p = 0.008$).

We repeated the modeling effort with a three-level categorical predictor and a boundary strength predictor. The categorical predictor coded whether a word closes both an IU and a clause, closes an IU but not a clause, or does not close either an IU or a clause (see above, Materials and Methods). Note that with the spontaneous speech material, because of the rarity of words that end clauses but are IU-nonfinal (Fig. 5B), it is impossible to study the full interaction between IU closure and clause closure in modulating the neural response. The results obtained with this model were highly consistent with the previous modeling effort. Stronger boundaries elicited a larger anterior negativity (Fig. 6A, middle; cluster extending from 90 to 469 ms following word offset, $p < 0.0001$). There was also a positive cluster preceding the negativity

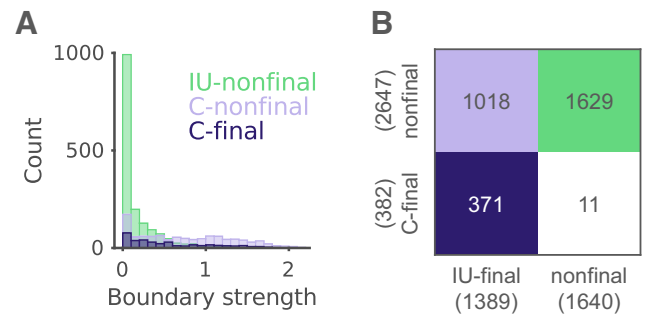


Figure 5. Relationship among boundary strength, clause closure, and IU closure. **A**, Distribution of boundary strength scores in IU-final and IU-nonfinal words (purple and green, respectively), and within IU-final words, of C-final and C-nonfinal words (dark and light purple, respectively). Histogram bins span 0.1 units of boundary strength. **B**, Word counts in each closure condition. We excluded from further analyses the rare words that ended clauses but appear at IU-nonfinal positions.

~200 ms before word offset (cluster extending from –227 until –137 ms relative to word offset, $p = 0.005$). Importantly, even when accounting for possible effects of clause closure, words that close an IU are associated with a centroparietal positivity preceded by a right anterior negativity compared with words that do not close an IU (Fig. 6A, top; negative cluster extending from –74–209 ms following word offset, $p = 0.008$; positive cluster extending from 174 to 484 ms following word offset, $p < 0.0001$). There was an additional positive cluster preceding the negativity, extending from –570 until –441 ms relative to word offset, $p = 0.005$. Clause-ending words are associated with a centroparietal positivity over and above that associated with IU closure only (Fig. 6A, bottom; cluster extending from 283 to 465 ms following word offset, $p = 0.003$).

The results suggest that the brain is attuned to both constructs, the IU and the clause. The earlier anterior aspects of the EEG response are robustly unique to IU processing, whereas the centroparietal later aspects of the response might be shared by both constructs. Finally, clause closure also allowed to predict unique delta-band EEG activity, yet to a lesser extent than IU closure and only in a few posterior electrodes (Fig. 6B).

Discussion

We set out to study the neural response to IUs with EEG. IUs are prosodically defined units that serve as a window onto communicative functions of the language system. Qualitative linguistic research on different languages suggests that they pace the progression of spontaneous discourse and serve as a resource for organizing conversational sequences. IUs have been documented in a variety of languages and appear to form rhythms of around 1 IU per second. The results of the current study suggest that the neural system is attuned to these units, a finding that opens new paths for investigating the neural substrates of language from a usage-based linguistics perspective (Tomasello, 1998, 2003).

The EEG response at IU closure includes a negative deflection at right anterior electrodes, starting as soon as the last word in the IU ends and lasting ~200 ms. IU closure is further characterized by a centroparietal positive deflection between 150 and 500 ms after the last word in the IU. Within words that close an IU we found that stronger acoustic boundaries elicit a larger anterior negativity between 100 and 400 ms. This was also the direction of the trend in words that do not close an IU.

Within the ERP literature, previous work using synthesized speech has identified the Closure Positive Shift (CPS) in response

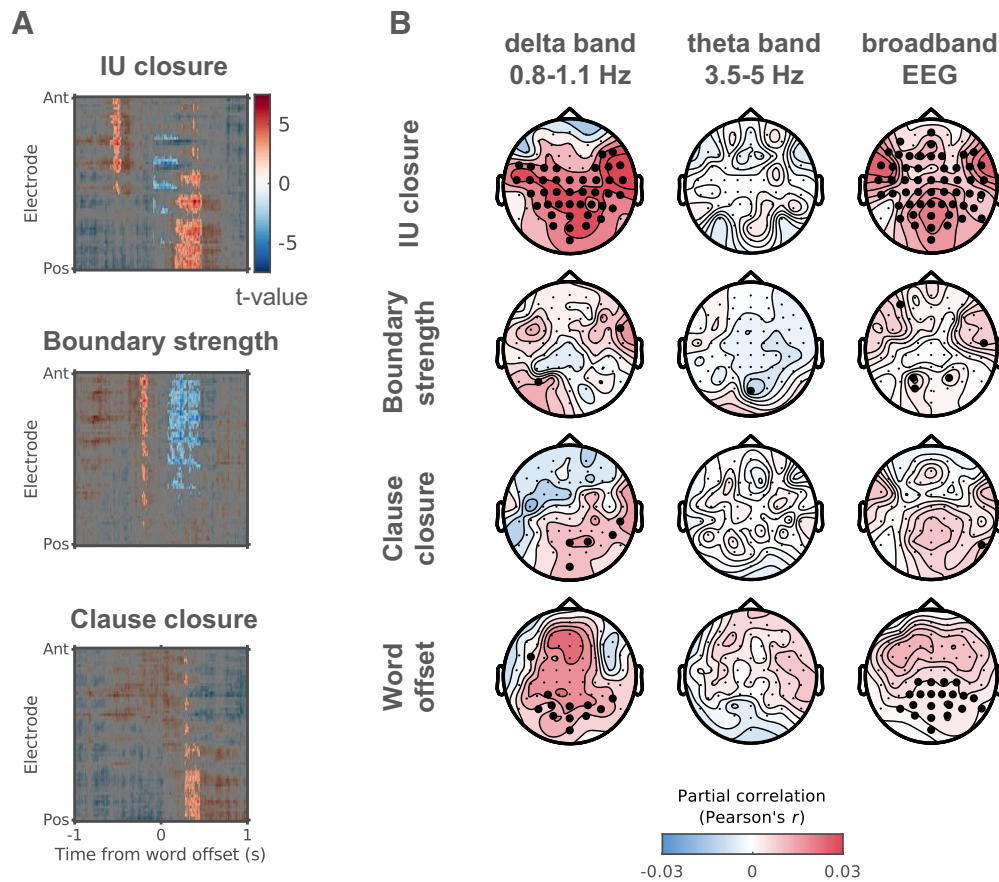


Figure 6. EEG amplitude and delta-band activity depend on IU closure, clause closure, and boundary strength. **A**, The full spatiotemporal results for a model that accounts for both IU and clause closure as well as acoustic boundary strength. A t value is calculated for every contrast at each electrode and time point in a 2 s window around word offset. Cluster-based bootstrap tests reveal significant clusters at the group level (unmasked). **B**, Group average partial correlation maps of the unique predictive accuracy of all effect combinations in a model that accounts for clause closure. We calculate unique predictive accuracy in left-out EEG data in the delta band (left), theta band (middle), and broadband EEG (right). The electrodes in bold denote where the predictive accuracy was significantly higher than chance (corrected for multiple comparisons).

to prosodic phrase boundaries in several languages from different phylogenetic units (families) and geographical areas, including German, Dutch, English, Swedish, Japanese, Chinese, and Korean (Bögels et al., 2011; Steinhauer, 2003; Steinhauer et al., 1999). Prosodic boundaries in these studies rely on cues that are nearly identical to those defining IUs (cf. Chafe, 1994 with Steinhauer et al., 1999). Several studies investigated the prosodic, syntactic, and contextual conditions in which the CPS emerges. A pause in the speech signal is not necessary for the emergence of the CPS (Holzgrefe-Lang et al., 2016; Itzhak et al., 2010; Steinhauer et al., 1999), nor is any acoustic boundary cue necessary whatsoever if a boundary is predictable on the basis of syntactic structure alone (Itzhak et al., 2010). Correspondingly, a small CPS may emerge while reading written sentences at the position of a comma (Steinhauer, 2003). In addition, responses to acoustically identical prosodic boundaries are modulated by the contextual predictability of the boundary (Kerkhofs et al., 2007). Together, these studies suggest that the CPS reflects a structuring of the input rather than a solely bottom-up response to the acoustic boundary cues giving rise to the prosodic phrasing. Nonetheless, a CPS also emerges in the absence of linguistic content at varying levels, suggesting some dependence on acoustics. A CPS may emerge at the closure of phrases that lack content words (jabberwocky sentences), semantics and syntax (pseudo sentences), and even phonological content (hummed sentences; Pannekamp et al., 2005). The neural response to IUs tightly corresponds to the CPS component previously described for phrase

boundaries. Our findings add to this literature in several ways. First, we describe neural responses while participants listened to speech in naturalistic conditions, in contrast to the isolated, constructed sentences that are typically used as stimuli in the CPS literature. Second, our work identifies what appear to be two different components within the classical CPS, a late anterior negativity that is dependent on boundary strength and a centroparietal positivity and early and right anterior negativity that are not dependent on boundary strength. In this regard, we obtained results that differ from those of Pauker (2013) in which both anterior negativity and centroparietal positivity did depend on boundary strength. The difference might stem from any of different experimental procedures, including the different language of presentation, radically different stimulus types, and a different operationalization of boundary strength. Here, we relied on rich natural speech material and operationalized boundary strength using a measure that includes prosodic information beyond that considered by Pauker (2013) (i.e., pitch modulation). Note that the author suggested that further study of the negativity is required (p. 170), and the results of the current study bear out this suggestion and point to what appear to be two subcomponents within the CPS.

IU boundaries are defined auditorily and hence depend (albeit subjectively) on acoustic properties. On the other hand, in spontaneous speech IUs assume a role of organizing the speech stream in units—pacing ideas and conversational actions. We attempt to address this complex manifestation of prosodic phrase

boundaries in naturally occurring spontaneous speech. First, we implement an algorithm for measuring acoustic-based boundary strength as a continuous variable. Next, we quantify the relation of boundary strength to expert transcription of natural speech to IUs. Finally, although previous studies that characterized EEG at prosodic phrase boundaries tackle the acoustic and linguistic facets separately, with our modeling approach we do so in tandem. Our results suggest that to the extent that listeners are capable of listening to speech in a foreign language attentively enough, non-speakers of a language would show a boundary strength effect but not an IU closure effect. Moreover, our work highlights the importance of IUs when considering both low-level processes and high-level linguistic representations.

In a separate model, we studied the shared impact of IUs and one such high-level linguistic representation, namely, the clause, on the neural response. The clause is one type of syntactic entity that is studied under many theoretical frameworks and has been linked to IUs in accounts that are focused on language use (Chafe, 1994; Givón, 2018; Pawley and Syder, 2000). We contend that research of great depth is required to identify the other syntactic patterns that speakers use to compose meaning or perform an action in communication (Ford et al., 2003; Hopper, 1987; Ono and Thompson, 1995; Schegloff, 1979). Such research is required to be able to study the possible impact of other syntactic entities on the neural response beyond clauses in a meaningful way. Specifically, a growing body of research relying on spontaneous, interactional speech provides evidence against the utility of traditionally assumed syntactic categories in explaining the structure of language as it is used in day-to-day interactions (Thompson and Ono, 2020, their footnote 1). When a full understanding of syntactic structuring in speech exists, our current approach would enable further investigation of the brain processes associated with IU processing in light of such syntactic structuring. Nonetheless, we were able to enrich our study of IUs by the inclusion of a basic syntactic entity, the clause, and provide the following insights. First, we were able to quantify the rate of clause closure in spontaneous Hebrew speech. Second, we measured the relation between IU closure and clause closure in spontaneous Hebrew speech and provided evidence for the tight dependency between clause closure and IU closure, but not the other way around. That is, a point of IU closure does not necessarily mean a clause has ended. This, we believe, impacts the ability to isolate the processing of clause closure independently of prosodic phrasing. Finally, we found independent effects for IU closure and clause closure. To the best of our knowledge, this is the first study of the impact of the two on the neural response while listening to spontaneous speech and controlling on a trial-by-trial level for variation in acoustic boundary strength.

Both IU closure and stronger acoustic boundaries within words that close an IU were associated with a positive deflection that preceded word offset by hundreds of milliseconds (~ 500 ms in the former, and 200 ms in the latter). Especially in the case of the IU closure contrast, such an early response might pertain to processes related to word onset, as the duration of words in our stimuli was 373 ± 214 ms (mean \pm SD). For example, there is evidence that preceding and following word onset, listeners' brains are engaged in next-word prediction and surprise (or lack of surprise) at the incoming word (Goldstein et al., 2022). The current study focused on responses to word offsets in accord with the mainstream CPS literature, but future studies may investigate this putative word onset response and clarify whether

there is, for example, a different engagement in word prediction depending on whether a word closes an IU or not. Another possible source for the positive deflection preceding word offset is the existence of systematic differences in pitch movements in IU-final words compared with other word positions. Whether final pitch movements were sufficiently accounted for by the acoustic boundary strength score remains to be tested.

In a final step, we related our results from the GLM model to the body of research on rhythmic brain mechanisms in speech processing (Ding et al., 2016; Giraud and Poeppel, 2012; Gross et al., 2013; Kaufeld et al., 2020; Keitel et al., 2017, 2018; Park et al., 2015). We used the estimated responses from our model to predict unseen continuous EEG and assess whether the unique predictive accuracy of the different effects exceeds chance level. This analysis suggests that IU-related neural activity contributes to the previously characterized delta-band neural speech tracking. We note that a link between delta-band neural speech tracking and the CPS has been previously conceptualized (Meyer et al., 2017, 2020) but to the best of our knowledge has never been explicitly tested. Therefore, in this final step, the current study also bridges two fruitful perspectives on language processing—the single evoked response and the continuous process. We demonstrate that characterizing evoked responses locked to naturally occurring events can equip the analysis of continuous speech with powerful theoretical models. This approach can readily advance the study of neural mechanisms of ongoing cognition in different domains by incorporating explicitly modeled events. Within the domain of speech, the focus on IUs embraces the natural covariation of acoustic boundary strength and the structuring of speech. By doing so we promote a view of the language system as an embodied system whereby abstract structure supervenes on perceptual modulations in time (Kreiner and Eviatar, 2014).

Neural speech tracking in the delta band has previously been studied via different time-varying representations of speech. Most frequently, the speech envelope is used (Gross et al., 2013; Kaufeld et al., 2020; Keitel et al., 2017, 2018; Park et al., 2015), but delta-band EEG activity also tracks the relative fundamental frequency and the spectral content of speech more generally (Teoh et al., 2019). These acoustic signals are tightly related to the theoretical construct of the IU as they capture the major cues for IUs. However, they likely capture additional important speech constructs, for example, accentuation. For this reason, our analyses included the modeling of boundary strength scores, which quantify linguistically relevant acoustic variation that specifically coincides with prosodic phrase boundaries (Sun et al., 2017). Future studies may clarify the ways in which the tracking of IUs characterized here interacts with time-varying acoustic predictors. This would require a different approach to modeling the data (temporal response function; Crosse et al., 2016; Theunissen et al., 2001). We would like to highlight that the explicit modeling of IUs offers a connection to accounts of how language is used in interaction and more specifically a possible handle to studying attentional dynamics during speech. A related topic that especially interests us is the purported role of IUs in pacing new information in discourse. In this context, we note that neural speech tracking has been tied in different ways to models of semantic context (Brodbeck et al., 2022; Broderick et al., 2019), and future work may shed light on the relation between these models and IUs, and their joint contribution to neural speech tracking.

References

- Bai F, Meyer AS, Martin AE (2022) Neural dynamics differentially encode phrases and sentences during spoken language comprehension. *PLoS Biol* 20:e3001713.
- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Soft* 67:1–48.
- Boersma P, Weenik D (2022) Praat: doing phonetics by computer [computer software], Version 6.2.14. Retrieved 9 June 2022 from <http://www.praat.org/>.
- Bögels S, Torreira F (2015) Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *J Phonetics* 52:46–57.
- Bögels S, Schriefers H, Vonk W, Chwilla DJ (2011) Prosodic breaks in sentence processing investigated by event-related potentials. *Lang Linguist Compass* 5:424–440.
- Brodbeck C, Bhattasali S, Cruz Heredia AAL, Resnik P, Simon JZ, Lau E (2022) Parallel processing in speech perception with local and global representations of linguistic context. *Elife* 11:e72056.
- Broderick MP, Anderson AJ, Lalor EC (2019) Semantic context enhances the early auditory encoding of natural speech. *J Neurosci* 39:7564–7575.
- Buxó-Lugo A, Watson DG (2016) Evidence for the influence of syntax on prosodic parsing. *J Mem Lang* 90:1–13.
- Chafe W (1987) Cognitive constraints on information flow. In: *Coherence and grounding in discourse*. (Tomlin RS, ed), pp 21–51. Amsterdam: John Benjamins.
- Chafe W (1994) *Discourse, consciousness and time: the flow and displacement of conscious experience in speaking and writing*. Chicago: Chicago UP.
- Chafe W (2018) *Thought-based linguistics*. Cambridge, UK: Cambridge UP.
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436.
- Cole J, Mo Y, Baek S (2010) The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Lang Cogn Process* 25:1141–1177.
- Couper-Kuhlen E, Barth-Weingarten D (2011) A system for transcribing talk-in-interaction: GAT 2. English translation and adaptation of Selting, M. et al. (2009). *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion* 12:1–51.
- Croft W (2022) *Morphosyntax*. Cambridge, UK: Cambridge UP.
- Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front Hum Neurosci* 10:604.
- Cruttenden A (1997) *Intonation*. (2nd ed.). Cambridge, UK: Cambridge UP.
- 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.
- Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D (2017) Temporal modulations in speech and music. *Neurosci Biobehav Rev* 81:181–187.
- Du Bois JW, Cumming S, Schuetze-Coburn S, Paolino D (1992) *Discourse transcription*. Santa Barbara Papers in Linguistics 4. University of California, Santa Barbara.
- Ford CE, Thompson SA (1996) Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: *Interaction and Grammar* (Ochs E, Schegloff EA, Thompson SA, eds), pp 134–184. Cambridge: Cambridge UP.
- Ford CE, Fox BA, Thompson SA (2003) Social interaction and grammar. In: *The new psychology of language: cognitive and functional approaches to language structure* (Tomasello M, ed), pp 119–143. Mahwah, NJ: Erlbaum.
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Genzer S, Ong DC, Zaki J, Perry A (2022) Mu rhythm suppression over sensorimotor regions is associated with greater empathic accuracy. *Soc Cogn Affect Neurosci* 17:788–801.
- Gilbert AC, Boucher VJ, Jemel B (2015) The perceptual chunking of speech: a demonstration using ERPs. *Brain Res* 1603:101–113.
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Givón T (2018) Toward a discourse definition of syntax: The communicative correlates of grammar. In: *On understanding grammar* (rev. ed.). Amsterdam: John Benjamins.
- Glushko A, Poeppel D, Steinhauer K (2022) Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Sci Rep* 12:14759.
- Goldstein A, et al. (2022) Shared computational principles for language processing in humans and deep language models. *Nat Neurosci* 25:369–380.
- Gravano A, Hirschberg J (2011) Turn-taking cues in task-oriented dialogue. *Computer Speech Language* 25:601–634.
- Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. *J Phon* 31:465–485.
- 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.
- Halliday MAK (1967). *Intonation and grammar in British English*. The Hague: Mouton.
- Haspelmath M (2020) On the difference between a clause and a sentence. In: *VAProsy jazykoznanija: megasbornik nanostatej* (Kibrik AA, Seménova KP, Sičinava DV, Tatevosov SG, Urmancieva AJ, eds), pp 597–605. Moscow: Buki Vedi.
- Henke L, Meyer L (2021) Endogenous oscillations time-constrain linguistic segmentation: cycling the garden path. *Cereb Cortex* 31:4289–4299.
- Himmelfmann NP, Sandler M, Strunk J, Unterladstetter V (2018) On the universality of intonational phrases: a cross-linguistic interrater study. *Phonology* 35:207–245.
- Holzgrete-Lang J, Wellmann C, Petrone C, Råling R, Truckenbrodt H, Höhle B, Wartenburger I (2016) How pitch change and final lengthening cue boundary perception in German: converging evidence from ERPs and prosodic judgements. *Lang Cogn Neurosci* 31:904–920.
- Hopper PJ (1987) Emergent grammar. *BLS* 13:139–157.
- Inbar M, Grossman E, Landau AN (2020) Sequences of intonation units form a ~1 Hz rhythm. *Sci Rep* 10:15846–15849.
- Itzhak I, Pauker E, Drury JE, Baum SR, Steinhauer K (2010) Event-related potentials show online influence of lexical biases on prosodic processing. *Neuroreport* 21:8–13.
- Jospe K, Genzer S, Klein Selle N, Ong D, Zaki J, Perry A (2020) The contribution of linguistic and visual cues to physiological synchrony and empathic accuracy. *Cortex* 132:296–308.
- Jun S (2005) Prosodic typology. In: *Prosodic typology: the phonology of intonation and phrasing*, pp 430–458. Oxford, UK: Oxford UP.
- Kaufeld G, Bosker HR, ten Oever S, Alday PM, Meyer AS, Martin AE (2020) Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *J Neurosci* 40:9467–9475.
- Keitel A, Ince RAA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage* 147:32–42.
- Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biol* 16:e2004473.
- Kerkhofs R, Vonk W, Schriefers H, Chwilla DJ (2007) Discourse, syntax, and prosody: the brain reveals an immediate interaction. *J Cogn Neurosci* 19:1421–1434.
- Kibrik AA (2019) Upper Kuskokwim (Athabaskan, Alaska). *Anthropol Linguist* 61:141–182.
- Kreiner H, Eviatar Z (2014) The missing link in the embodiment of syntax: prosody. *Brain Lang* 137:91–102.
- Labov W, Waletzky J (1967) Narrative analysis: oral versions of personal experience. In: *Essays on the verbal and visual arts*. (Helm J, ed), pp 12–44. Seattle: Washington UP.
- Laury R, Ono T, Suzuki R (2019) Questioning the clause as a crosslinguistic unit in grammar and interaction. *Studies in Language* 43:364–401.
- Matsumoto K (2000) Intonation units, clauses and preferred argument structure in conversational Japanese. *Lang Sci* 22:63–86.
- Meyer L, Henry MJ, Gaston P, Schmuck N, Friederici AD (2017) Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cereb Cortex* 27:4293–4302.
- Meyer L, Sun Y, Martin AE (2020) “Entraining” to speech, generating language? *Lang Cogn Neurosci* 35:1138–1148.
- Obleser J, Kayser C (2019) Neural entrainment and attentional selection in the listening brain. *Trends Cogn Sci* 23:913–926.
- Ofir N, Landau AN (2022) Neural signatures of evidence accumulation in temporal decisions. *Curr Biol* 32:4093–4100.e6.

- Ono T, Thompson SA (1995) What can conversation tell us about syntax? In: *Alternative linguistics: descriptive and theoretical modes* (Davis PW, ed), pp 213–272. Amsterdam: John Benjamins.
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
- Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD (2005) Prosody-driven sentence processing: an event-related brain potential study. *J Cogn Neurosci* 17:407–421.
- Park H, Ince RAA, Schyns PG, Thut G, Gross J (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr Biol* 25:1649–1653.
- Pauker E (2013) How multiple prosodic boundaries of varying sizes influence syntactic parsing: Behavioral and ERP evidence [dissertation]. McGill University.
- Pawley A, Syder FH (2000) The one-clause-at-a-time hypothesis. In: *Perspectives on fluency* (Riggenbach H, ed), pp 163–199. Ann Arbor: Michigan UP.
- Pellegrino F, Coupé C, Marisco E (2011) A cross-language perspective on speech information rate. *Language* 87:539–558.
- Pernet C, Chauveau N, Gaspar C, Rousselet G (2011) LIMO EEG: a toolbox for hierarchical linear modeling of electroencephalographic data. *Comput Intell Neurosci* 2011:831409–831411.
- Pernet C, Suay Mas I, Rousselet G, Martinez R, Wilcox R, Delorme A (2022) Electroencephalography robust statistical linear modelling using a single weight per trial. *Aperture Neuro* 2022.
- R Core Team (2022) R: A Language and Environment for Statistical Computing [Computer software], Version 4.2.0. Retrieved 9 June 2022 from <https://www.R-project.org/>.
- Ravid D (1999) Hataxbir vehora'to ('Syntax and its teaching'). Tel-Aviv: Center for Educational Technology.
- Schegloff EA (1979) The relevance of repair to syntax-for-conversation. In: *Syntax and semantics* (Givón T, ed), pp 261–286. New York: Academic Press.
- Seifart F, Strunk J, Danielsen S, Hartmann I, Pakendorf B, Wichmann S, Witzlack-Makarevich A, Himmelmann NP, Bickel B (2021) The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguist Vanguard* 7:20190063.
- Selting M (2010) Prosody in interaction: state of the art. In: *Prosody in interaction* (Barth-Weingarten D, Reber E, Selting M, eds), pp 3–40. Amsterdam: John Benjamins.
- Shattuck-Hufnagel S, Turk AE (1996) A prosody tutorial for investigators of auditory sentence processing. *J Psycholinguist Res* 25:193–247.
- Smith NJ, Kutas M (2015) Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology* 52:157–168.
- Stehwien S, Meyer L (2022) Short-term periodicity of prosodic phrasing: corpus-based evidence. *Proc Speech Prosody* 2022:693–698.
- Steinhauer K (2003) Electrophysiological correlates of prosody and punctuation. *Brain Lang* 86:142–164.
- Steinhauer K, Alter K, Friederici AD (1999) Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci* 2:191–196.
- Suni A, Šimko J, Aalto D, Vainio M (2017) Hierarchical representation and estimation of prosody using continuous wavelet transform. *Computer Speech Language* 45:123–136.
- Szcszek-Reed B (2010) Intonation phrases in natural conversation: A participants' category? In: *Prosody in interaction* (Barth-Weingarten D, Reber E, Selting M, eds), pp 191–212. Amsterdam: John Benjamins.
- Teoh ES, Cappelloni MS, Lalor EC (2019) Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *Eur J Neurosci* 50:3831–3842.
- The Language Archive (2022) ELAN [Computer software]. Version 6.3. Retrieved 9 June 2022. Available at <https://archive.mpi.nl/tla/elan>.
- Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12:289–316.
- Thompson SA (2019) Understanding 'clause' as an emergent 'unit' in everyday conversation. *Studies in Language* 43:254–280.
- Thompson SA, Couper-Kuhlen E (2005) The clause as a locus of grammar and interaction. *Discourse Studies* 7:481–505.
- Thompson SA, Ono T (2020) Introduction. In: *The 'noun phrase' across languages*. (Ono T, Thompson SA, eds), pp 1–8. Amsterdam: John Benjamins.
- Tilsen S, Arvaniti A (2013) Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *J Acoust Soc Am* 134:628–639.
- Tomasello M (Ed.) (1998) *The new psychology of language: cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum.
- Tomasello M (Ed.) (2003) *The new psychology of language: cognitive and functional approaches to language structure* (Volume 2). Mahwah, NJ: Erlbaum.
- Yekutieli D, Benjamini Y (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 29:1165–1188.