

Research Articles | Behavioral/Cognitive

Emotional Content Enhances Attentional Tracking of Conversations through Context-Sensitive Modulation of Auditory-Language Networks

<https://doi.org/10.1523/JNEUROSCI.0506-26.2026>

Received: 24 March 2026

Revised: 31 August 2026

Accepted: 23 September 2026

Copyright © 2026 Heikkinen et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Emotional Content Enhances Attentional Tracking of Conversations through Context-Sensitive Modulation of Auditory-Language Networks

abbreviated title: Neural Tracking of Emotional Conversations

Pyry Heikkinen^{1,4*}, Santeri Lepistö^{1*}, Patrik Wikman^{1,2,3,4}, Emmi Koskinen⁵, Anssi Peräkylä⁵,
Niklas Ravaja¹, & Ville Johannes Harjunen¹

¹ *Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland*

² *Department of Neuroscience, Georgetown University, Washington D.C., USA*

³ *Advanced Magnetic Imaging Centre, Aalto NeuroImaging, Aalto University, Espoo, Finland*

⁴ *Finnish Centre of Excellence in Music, Mind, Body, and Brain, Cognitive Brain Research Unit, Department of Psychology, Faculty of Medicine, University of Helsinki, Helsinki, Finland*

⁵ *Faculty of Social Sciences, University of Helsinki, Helsinki, Finland*

*corresponding authors: pyry.heikkinen@helsinki.fi, santeri.lepisto@helsinki.fi

University of Helsinki, PO Box 21, Haartmaninkatu 3, 00014, Helsinki, Finland.

Number of pages: 43; Number of figures: 5; Number of tables: 4

Abstract word count: 191; Introduction word count: 649; Discussion word count: 1402

Conflict of interest: The authors declare no competing financial interests.

Acknowledgements

The research has been funded by The Society of Swedish Literature in Finland, and Research Council of Finland, [Project number: 356592]. We thank Mariel Wuolio for her help in preparing materials, participant recruitment and assistance in MEG measurements. We are also grateful to Markus Söderman for his assistance with the preparation of the stimulus material and study design, and Jan Lindström for his role in the conceptualization and planning of the project. Finally, we thank students Leevi Lintunen, Jenna Jakonen, Aino Innanen and Aarni Arvilommi for their contributions in producing the study's stimulus material.

Data Availability Statement

The datasets and analysis materials supporting the findings of this study are openly available on the Open Science Framework (<https://osf.io/9eyuf/>). The repository includes processed data, analysis scripts, and supplementary materials necessary to reproduce the reported results.

Abstract

Engagement in conversation often requires continuous monitoring of relevant speech in a continuously evolving auditory environment. Emotional salience of semantic content may influence this process, but its role in attentional tracking of continuous speech remains

unclear. We address this question using an audiovisual “Cocktail-party” paradigm combined with source-localized magnetoencephalography and encoding–decoding analyses. Participants (18 female, 14 male) selectively attended to one dialogue while ignoring a concurrent distractor dialogue, with both emotional salience and semantic coherence of the conversations manipulated independently. Decoding analyses revealed that negatively valenced, high-arousal, and socially relevant semantic content strengthened neural speech tracking, but only for attended dialogues. This effect depended on semantic structure, emerging earlier in semantically coherent than incoherent dialogues. Encoding analyses further showed latency- and region-specific effects across temporal, parietal, and frontal cortical regions. Emotional salience and coherence jointly modulated mid-latency (100 - 250 ms) responses, whereas late-latency (270 - 450 ms) responses were driven primarily by emotional salience. These findings demonstrate that emotional meaning enhances neural prioritization of attended speech and semantic coherence shapes the temporal dynamics of this enhancement. Together, these results indicate context-sensitive emotional modulation of auditory–language processing during selective listening of continuous speech.

Significance Statement

Understanding how the brain prioritizes and selectively processes information in naturalistic, continuously evolving audiovisual scenes such as conversations is a central challenge in neuroscience. Here, we show that conversations conveying negatively valenced, high-arousal, and socially relevant content enhance neural tracking of attended speech in a noisy cocktail-party setting. This enhanced tracking depends on the semantic structure of the conversation, emerging earlier for coherent dialogues that are easy to follow. Furthermore, we show that emotional content strengthens mechanisms related to speech comprehension and sharpens segregation of relevant from irrelevant talk. These findings demonstrate that emotional meaning and semantic continuity jointly shape cortical processing of continuous speech, providing new insight into how the brain selects and processes meaningful information during a conversation.

Introduction

When engaging in a conversation in noisy settings, attention plays a critical role in determining how speech is processed and represented in the brain. Emotional content, such as highly arousing and negative information with social significance, may bias this process by enhancing sensory processing of stimulus characteristics (Fischer et al., 2025; Lang et al., 1998; Sander et al., 2003; Yiend, 2010). However, it remains unclear whether emotional semantic content of speech modulates attentional selection by influencing neural tracking of naturalistic conversation.

At the neural level, much of what we know about emotional speech and language processing comes from studies examining responses to emotional words and prosody. These local emotional cues influence multiple processing stages, from early sensory encoding (e.g., N1) and salience detection (e.g., P2) to later evaluative stages (e.g., N400) (Kotz & Paulmann, 2011; Pinheiro, 2025; Schirmer & Kotz, 2006). Their processing is further shaped by linguistic context, with emotionally congruent contexts facilitating emotional word processing through semantic predictability (Bayer et al., 2010; J. Ding et al., 2015, 2016; Fields & Kuperberg, 2012; Martín-Loeches et al., 2012; Nieuwland & Van Berkum, 2006; van Berkum et al., 2003; Zhang et al., 2021).

In everyday communication, emotional salience unfolds across continuous talk rather than isolated utterances (Hoemann et al., 2025; Mar, 2011). In such settings, the brain integrates information at multiple representational levels, from rapid acoustic fluctuations to slower semantic and contextual structure (Brodbeck et al., 2022; Zion Golumbic et al., 2012; Zion Golumbic et al., 2013). This integration depends on coordination between auditory, language, and higher-order cognitive networks linking sensory input with semantic representations and contextual knowledge (Fedorenko et al., 2024; Ji et al., 2019; Yamashita et al., 2025). Neural responses to continuous speech reflect this organization, progressing from early spectrotemporal encoding to later lexical–semantic integration and contextual predictability (Broderick et al., 2018; Zion Golumbic et al., 2013).

In environments where relevant speech must be segregated from background noise, attention enhances neural tracking of the speech stream (N. Ding & Simon, 2012; Fiedler et al., 2019; O’Sullivan et al., 2019; Zion Golumbic et al., 2013). Segregating competing speech streams relies on integrating information across multiple timescales to support processing of the task-relevant stream (Brodbeck et al., 2022; Power et al., 2012; Teoh et al., 2022). Recent work shows that neural tracking of conversation fluctuates across turns and engages a distributed cortical network supporting auditory, linguistic, and social-cognitive processing (Wikman et al., 2021, 2024). Together, these findings raise the possibility that emotional meaning may influence neural tracking of conversational speech by interacting with distributed cortical systems that track the unfolding semantic context.

The present study examined whether emotional salience enhances neural tracking of conversational speech during selective listening. During magnetoencephalography (MEG) recording, participants listened to target dialogues while ignoring a competing speech stream in an audiovisual cocktail-party paradigm. We manipulated both the emotional salience of dialogue content and the semantic coherence of dialogue turns. The emotional-content condition comprised negatively valenced, highly arousing, and socially relevant dialogues. This profile is common in natural communicative settings (e.g., gossip, rumors; Berger & Milkman, 2012; Feinberg et al., 2012; Robbins & Karan, 2020), and combinations of these features are shown to promote attentional prioritization in sensory processing (Astudillo et al., 2018; Fischer et al., 2025; Schacht & Vrtička, 2018).

To characterize neural speech tracking, we combined speech envelope reconstruction (SER) with temporal response function (TRF) analyses. SER reconstructs the acoustic speech envelope from multivariate neural activity, whereas TRFs model stimulus–

response relationships across time lags to estimate how fluctuations in the speech envelope drive cortical activity (Crosse et al., 2016; Kaufman & Zion Golumbic, 2023).

We hypothesized that negatively valenced, highly arousing, and socially relevant conversational content would enhance neural tracking of attended conversational speech, and that dialogue coherence would modulate the temporal dynamics of this emotional enhancement. Finally, we predicted that emotional salience and dialogue coherence would differentially modulate neural responses across response latencies and cortical regions involved in language comprehension.

Materials and Methods

Participants

Thirty-two healthy volunteers (18 women, 14 men, age range 20–50 years, $M = 30.22$, $SD = 7.98$) took part in the study. Participants, recruited through university mailing lists and campus outreach, were asked during recruitment to report any diagnosed neurological or psychiatric disorders or use of medications affecting the central nervous system, all of which constituted exclusion criteria. Participants were also required to have normal or corrected-to-normal vision and hearing. All participants provided written informed consent before the experiment and received a reward valued at approximately 30 euros. The research protocol was approved by the University of Helsinki Research Ethics Committee of the Faculty of Medicine, and the study followed the ethical principles outlined in the World Medical Association's Declaration of Helsinki. One participant was excluded due to data-quality issues, leaving a final sample of $N = 31$. For this participant, envelope-tracking quality-control diagnostics indicated unusable model output: SER performance was at or near zero across the tested regularization values, and TRF estimates showed atypically low-amplitude, highly stereotyped waveform patterns across regions in all conditions.

Sensitivity Power Analysis

Given the final sample size ($N = 31$), we conducted a simulation-based sensitivity analysis for the planned SER and TRF analyses. The analysis estimated the minimum magnitudes of the principal SER and TRF effects detectable with 80% power at $\alpha = .05$. Simulations preserved the observed variances and shrinkage-estimated within-participant correlation structures while varying the target effect. For each candidate magnitude, 5,000 datasets with 31 participants were analyzed using the corresponding Type III repeated-measures ANOVA. Minimum detectable effects were reported in the original measurement units and as generalized eta squared (η_G^2). Analyses were conducted in R 4.6.1 using Superpower (Lakens and Caldwell, 2021). For SER, simulations reproduced the $2 \times 2 \times 2$ Attention $\times$ Emotion $\times$ Coherence design, and for TRFs, they reproduced the Emotion $\times$ Coherence $\times$ ROI designs used for the respective latency windows. The 80%-power minimum detectable effect size for both the SER Attention $\times$ Emotion interaction and the Attention $\times$ Emotion $\times$ Coherence interaction was $\eta_G^2 = .033$. For the analysed midlatency TRF response the minimum detectable effect size at 80% power was $\eta_G^2 = .0029$. For the late-latency response, the minimum detectable effect size at 80% power was $\eta_G^2 = .0109$.

Experimental Design

The participants completed an audiovisual “cocktail-party” experiment during MEG recording (Figure 1). After a 4-minute resting-state recording, participants performed two experimental runs consisting of 16 trials each. In each trial, two dialogues with six lines were presented simultaneously: an audiovisual target dialogue and an audio-only distractor dialogue, with audio presented from the same speakers concurrently. Participants were instructed to focus exclusively on the target dialogue. Interactants in the dialogues did not exhibit bodily expressions (e.g., facial expressions, hand gestures, or postural movement). Crucially, to focus the manipulation on semantic content, the speech was delivered with flat prosody, minimizing variation in stress, rhythm, and loudness.

The experiment followed a 2 x 2 x 2 factorial design, manipulating attentional engagement (target vs. distractor), and two dimensions of semantic content in both target and distractor dialogues: the emotional salience of the dialogues (emotional vs. neutral) and turn-to-turn dialogue coherence (narratively coherent dialogue vs. narratively incoherent dialogue). Emotional dialogues featured negatively valenced topics, whereas neutral dialogues described affectively non-salient activities (details of the dialogue material are provided in the Dialogue Set Construction section). Coherent dialogues maintained contextually related conversational turns, while incoherent dialogues were generated by mixing lines from different dialogues. Each trial within a run formed a different factorial pairing, resulting in 16 unique target–distractor combinations. For example, both runs included a trial in which an emotionally incoherent target dialogue was presented alongside a neutral coherent distractor dialogue. The trial order was pseudo-randomized.

Stimulus Material

Dialogue Set Construction

The stimulus material comprised a set of 128 spoken dialogues developed specifically for the present study, with conversations varying in emotional salience (emotional vs. neutral) and coherence (coherent vs. incoherent). All dialogues were scripted by two research assistants and the last author. Each consisted of a six-line exchange of approximately 30 seconds, with semantic content manipulated according to the intended experimental condition. For each participant, a subset of 32 target dialogues and 32 distractor dialogues was selected from this larger stimulus set. Each dialogue was recorded by two different male–female speaker pairs (Dyads A and B), yielding two versions of each dialogue with different speakers. The speaker pair was counterbalanced across participants: for half of the participants, Dyad A performed the target dialogues and Dyad B the distractor dialogues, whereas for the other half, this assignment was reversed. The gender of the active speaker also differed between the target and distractor dialogues.

The level of emotional salience was manipulated by constructing dialogues around two topic clusters that differed in their apparent emotional salience. Emotional dialogues were about negatively valenced, emotionally arousing events with high social relevance (e.g., tragic accident, fear- or disgust-related experience, physical conflict, mental health issue or illness, harassment, social exclusion), whereas neutral dialogues centered on neutral, everyday events with low emotional arousal and low social significance (e.g., grocery shopping, attending a lecture, cooking, working, talking about the weather, having lunch, drinking coffee). The emotional salience of the stimuli was validated in a separate sample (see below).

In each dialogue, information concerning the topic unfolded across the six conversational turns (i.e., each turn introduced new emotionally salient or neutral content). The level of semantic coherence was manipulated by varying the semantic relatedness of the interactants' utterances (turn-to-turn contextual continuity). In coherent dialogues, interactants responded appropriately to each other's speech, whereas in incoherent dialogues, the turns were not semantically connected. Incoherent dialogues were constructed by recombining utterances from different coherent dialogues to form new dialogues with unrelated successive turns (e.g., one interactant makes a comment about gardening, the other responds with an unrelated remark about shopping). The resulting coherence manipulation was independently validated using dialogue ratings and natural language processing-based measures of semantic dispersion (see below).

Independent Dialogue Validation Study

To ensure the dialogue-materials were in accordance with our experimental design, we conducted a separate dialogue validation study administered online using a custom jsPsych-based questionnaire (de Leeuw, 2015). The dialogue validation study utilized the full set of 128 dialogues, with each participant viewing and evaluating a subset of 16 dialogues

comprising four dialogues from each experimental condition (emotional coherent, emotional incoherent, neutral coherent, and neutral incoherent). Participants were primarily recruited through university campus outreach. The validation study sample included 31 participants (14 female, 16 male, and 1 reporting “other/prefer not to disclose”), with ages ranging from 22 to 70 years ($M = 32.71$, $SD = 8.68$).

Participants rated each dialogue on five dimensions—arousal, valence, social relevance, visual imagery, and semantic coherence—using 9-point Likert-type scales. Arousal and valence were assessed using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994). Social relevance was assessed with the question “To what extent did the dialogue concern people, relationships, or social interaction?” (1 = not at all, 9 = very much). Visual imagery was assessed with the question “How easy was it to imagine the events described in the dialogue?” (1 = very difficult, 9 = very easy). Finally, perceived coherence was assessed using the statement “Did the dialogue progress coherently” (1 = strongly disagree, 9 = strongly agree).

Lexical and Semantic Feature Extraction

To examine if the dialogue categories differed on their linguistic and semantic properties, we extracted the lexical and semantic features of each dialogue. These variables were used first to examine any possible differences between the categories, and then in follow-up analyses testing whether they accounted for observed effects in SER.

We used lexical-control variables derived from LASTU, a psycholinguistic search tool and lexical database for Finnish stimuli (Itkonen et al., 2024). The dialogue transcripts were lemmatized (returned to base form) with STANZA (Qi et al., 2020), and each word was matched to the corresponding LASTU entry. Dialogue-level lexical variables were then calculated by averaging token-level values across all matched lexical tokens within each dialogue.

Mean word frequency was calculated as the average log-transformed lemma frequency across matched tokens in the dialogue, with higher values indicating more frequent occurrence in the Finnish Internet Parsebank corpus (Luotolahti et al., 2015). Mean word length was calculated as the mean character count of words within a dialogue. Mean word ambiguity was calculated as the average LASTU lemma-ambiguity value, with higher values indicating words associated with more possible lexical or morphological interpretations. Mean morphological feature count was calculated as the average number of morphological feature tags assigned to the actual words in the dialogue, indexing the degree of morphological complexity of the dialogue content.

In addition to the above lexical features, dialogue level-narrative density and semantic spread were used to characterise the overall narrative structure of the dialogues. Narrative density indexed the extent to which a dialogue contained explicit event-structure cues, including references to actions or events, causal or temporal relations, goals or intentions, and human participants. Higher values therefore indicated dialogues with denser event-based and causal-temporal content.

Finally, semantic dispersion was computed using sentence embeddings generated with the Finnish Sentence-BERT paraphrase model (Kanerva et al., 2024). This model is a Finnish sentence-transformer trained from FinBERT and produces sentence-level embeddings suitable for estimating semantic similarity between sentences (Reimers & Gurevych, 2019; Virtanen et al., 2019). Semantic dispersion was then calculated as the average cosine distance among the turn embeddings within a dialogue, providing a dialogue-level index of how semantically spread out the turns were.

Stimulus Presentation

The stimulus was presented with the Presentation® software (Neurobehavioral Systems Inc.). The target dialogues were presented on a projected screen (67 * 38 cm) at a viewing distance of approximately 130 cm (from the participant's eyes). The projector's resolution

was 1920 * 1080 (Full HD), with a refresh rate of 120 Hz. Before each target dialogue, a gray screen with the instruction “Pay attention to the conversation” was shown. Subsequently, target and distractor dialogues were presented line by line. After each line, a gray screen with a fixation cross was presented. For each line, the stimulus volume was set to a comfortable level based on the target, and the paired distractor was attenuated by 10 decibels based on the root mean square of the amplitude, and each distractor line was set to begin 250 ms before the onset of the corresponding target line.

Stimulus Processing

The broadband amplitude envelope of each dialogue was obtained as the absolute value of the Hilbert transform. Audio envelopes were band-pass filtered (0.5–15 Hz, 4th-order Butterworth filter), and sampling rate was matched to MEG data at 128 Hz. Finally, the envelopes were segmented into speech epochs for each line by detecting silent gaps of ≥ 1 s where the envelope remained below 1% of the maximum amplitude. As the distractor dialogue lines started 250 ms prior to the targets, the start of distractor epochs was cut to align with the onset of target lines.

Behavioral Measures

After each trial, participants completed two questionnaires with a VPIxx response device. To evaluate whether participants attended the target dialogues, they answered four true–false statements about the content of the target dialogue. The statements were designed to measure participants' detection of central semantic information conveyed in the dialogues, such as recognizing mentioned people, places, objects, or actions (e.g., “The interactant asked whether the other person had an umbrella”; “The interactant is looking for a job”; “The interactant said that their lecturers change often”). Each true–false statement was constructed with its correct answer randomly set as either true or false. After responding, participants received feedback (e.g., “3 out of 4 correct”). The proportion of correct responses was used to quantify recall accuracy (number of correct responses divided by four). Second, participants rated how emotionally charged the dialogue was on a three-point scale (1 = “Not at all”, 2 = “Somewhat”, 3 = “Highly”).

MEG and MRI Data Acquisition

For source localization, structural MRI data were acquired using a 1.5 T Siemens Avanto scanner (Erlangen, Germany) with a T1-weighted 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (voxel size = $1 \times 1 \times 1$ mm³). Brain activity during the experiment was recorded at 1 kHz with a 306-channel MEGIN Triux system (204 planar gradiometers, 102 magnetometers) at the BioMag Laboratory, Helsinki University Hospital. Horizontal and vertical electro-oculograms (EOG), as well as electrocardiograms (ECG), were recorded to detect eye movements and cardiac artifacts. The head positions of the participants were continuously measured using 5 head position indicator (HPI) coils placed around the scalp. To facilitate accurate co-registration between the MEG data and a structural magnetic resonance image (MRI), 100 evenly divided scalp points were digitized (Polhemus).

Data Processing

MEG Preprocessing

MEG signals were first preprocessed using MNE-Python 1.5 (Gramfort et al., 2013). Raw recordings were denoised using MNE MaxFilter (Taulu & Simola, 2006) with temporal signal space separation (tSSS, 20-second sliding window). The head position of each participant was spatially aligned to the start of run 1, with movement compensation based on the recorded continuous HPI coil information. Noisy and flat channels were automatically marked, and the data were resampled to 128 Hz. Data were high-pass filtered at 0.5 Hz and low-pass filtered at 15 Hz using zero-phase FIR filters. The 0.5–15 Hz passband was chosen to reduce slow drifts while retaining low-frequency neural activity and speech-envelope modulations relevant for continuous speech tracking, including delta- and theta-range

fluctuations and adjacent syllabic/onset-related structure. Similar low-frequency passbands extending to 15 Hz have been used in TRF and stimulus-reconstruction analyses of continuous speech (Crosse et al., 2016; Di Liberto et al., 2015).

Bad channel selection was manually confirmed, and channels were interpolated with spherical-spline interpolation. Next, ocular and cardiac artefacts were removed via the MNE FastICA algorithm by flagging components correlated with the EOG and ECG, respectively, and manually confirming the ICA exclusion. Next, MEG data were epoched from each dialogue line onset to offset for both target and distractor dialogues separately. Finally, the data were grouped based on emotional salience and semantic coherence, yielding separate neural data and audio envelope epochs for each Attention-Emotional Salience-Coherence combination.

Distributed Source Localisation

To estimate the cortical generators of the measured MEG signal, distributed minimum-norm source reconstruction (Gramfort et al., 2014; Hämäläinen & Ilmoniemi, 1994) was performed with each participant's FreeSurfer reconstructed surfaces. The surfaces were coordinate aligned to MEG digitization points, and a single-shell boundary element head model (BEM, conductivity 0.3 S/m) was computed from the inner skull surface. The cortical source spaces were defined at oct6 resolution (~5 mm spacing) on the white matter surface. Noise covariance matrices were estimated from 4-minute-long empty-room recordings. Minimum-norm inverse operators were built (depth weighting = 0.2; SNR = 1/9) with dipole orientation constrained to the cortical surface normal using a loose orientation constraint (0.2). MEG data were projected to source space separately for each MEG epoch, yielding one cortical source current estimate time-locked to each stimulus combination.

Encoding and Decoding Models

Encoding and decoding modelling were used to examine how distributed cortical activity tracked the auditory envelopes of the presented dialogues, utilizing temporal response functions (TRFs) and speech envelope reconstruction (SER; Crosse et al., 2016; Lalor & Foxe, 2010). By using speech envelope reconstruction (SER), it is possible to assess the fidelity with which the cortical activity follows the speech signal, whereas temporal response functions (TRFs) characterize the temporo-spatial dynamics of stimulus-response relationship (Crosse et al., 2016). The temporal lags considered in the models between speech signal and brain activity were restricted to be between -200 – 600 milliseconds. For both approaches, the *receptive field estimation* function in MNE-Python was used, with the default lagged ridge-regression approach, with 4-fold crossvalidation (regularisation-parameters tested between 1 to 1×10^6 ; SI Figure S1). Separate models were fit for each Attention - Emotional salience - Dialogue coherence combination. Model performance was quantified as the mean Pearson correlation between model predictions and held-out test data across cross-validation folds. The final selection was based on the highest overall correlation value, averaged across conditions. For the encoding models, a regularization value of $\lambda = 10^4$ provided the best performance and was therefore used in the final analyses. The same λ was used for SER to keep regularization matched across encoding and decoding models.

Speech Envelope Reconstruction (SER)

To reduce computational strain, the neural input data was first condensed to anatomically-functionally defined cortical areas, based on the human connectome project's multi-modal parcellation (Glasser et al., 2016) by calculating the average time-course within each cortical area (with sign-flip of activations in the non-dominant orientations with > 90 degree deflection). Reconstruction accuracy was quantified as Pearson's product-moment correlation coefficient of the reconstructed and true envelopes across cross-validation folds.

To examine how speech tracking changed across dialogue turns, we used the envelope reconstructions of the held-out data from the overall cross-validated SER analysis.

The reconstructed and actual envelopes were segmented according to the dialogue-turn boundaries, and turn-wise SER accuracy was calculated as the Pearson correlation between the reconstructed and actual envelopes within each turn. Correlation coefficients were Fisher z-transformed before statistical analysis. Overall SER accuracy and turn-wise SER trajectories were analyzed separately.

Temporal Response Functions (TRFs)

TRFs were used to model how changes in the speech envelope drove cortical responses over time with the envelope as the input and the neural source activity at each vertex as the response. Regression coefficients were averaged across folds to yield source-by-time TRFs for each subject and condition and were subsequently averaged based on the Glasser cortical parcellation. From this partition, the auditory, language and default-mode related regions located in the lateral temporal, parietal and frontal areas were chosen for further analysis. As the grand averaged responses displayed a distinct two-peak response pattern, the data were analysed in two separate time windows of interest. For the earlier window, a local positive peak within 100–250 ms was identified for each network and condition separately. The mean amplitude within a 50 ms window around this peak was computed, capturing the mid-latency response while accommodating latency differences across cortical regions. The later activation was quantified as the overall mean across 270–450 ms time window, due to less pronounced curvature in the source waveforms.

Statistical Analyses

All statistical analyses were performed in R (version 4.5.1). Statistical significance was assessed at $\alpha = 0.05$ and effect sizes are reported as generalized eta squared (η_G^2). Post hoc comparisons were conducted when applicable.

Dialogue-Level Validation

To validate the differences between stimulus categories and characterize possible differences between them in lexical or semantic features, we conducted dialogue-level control analyses comparing factors of emotional content (emotional vs. neutral) and coherence (coherent vs. incoherent). The analyses were conducted at the dialogue level, with each of the 128 dialogues acting as one observation. The used questionnaire-derived variables were dialogue-level mean ratings of arousal, valence, social relevance, visual imagery, and perceived coherence, and the used lexical variables were mean log word frequency, mean word length, mean word ambiguity, and mean morphological feature count. Finally, we used semantic variables of narrative density and embedding-based semantic dispersion across dialogue turns. Continuous outcome variables were standardized before modelling. Type-III tests were used for the analyses, with false-discovery-rate correction applied across the set of tests (Benjamini & Hochberg, 1995).

Behavioral Analysis

For the MEG study participants, emotional intensity ratings were analyzed using generalized estimating equations (GEE) with an ordinal link function. Recall accuracy and speech envelope reconstruction (SER) accuracy were analyzed using repeated-measures ANOVAs with within-subject factors of Attention (attended, distractor), Emotional Salience (emotional, neutral), and Dialogue Coherence (coherent, incoherent).

SER and TRF Analysis

To assess temporal changes in neural tracking across conversational turns, SER trajectories were analyzed using repeated-measures ANOVAs with Turn included as an additional within-subject factor. When the assumption of sphericity was violated, Greenhouse–Geisser corrections were applied. Multiple comparisons were corrected using the Holm-correction.

To control for possible effects of lexical or semantic features in speech tracking, we tested if any of the ancillary lexical or discourse-level variables accounted for the observed effects in target SER. For this, a baseline categorical model predicting dialogue-level target SER from emotional-content condition, coherence, and their interaction was compared with

models additionally including either lexical or semantic-level predictors. The model fits were evaluated using Akaike information criterion (AIC) and Bayesian information criterion (BIC; Akaike, 1974; Schwarz, 1978).

Next, we examined whether the independently rated affective-social dimensions provided a stronger continuous account of target SER than the categorical manipulation. These models tested arousal, valence, social relevance, and their interactions and were compared with the planned model with categorical factors of emotion and coherence. The model fits were evaluated using AIC and BIC. Together, these follow-up analyses test whether the categorical SER effect persists after accounting for measured stimulus differences and clarify the relation of SER to the dialogues' affective-social profile.

Finally, temporal response function (TRF) amplitudes were averaged within predefined latency windows (mid-latency: 100–250 ms; late-latency: 270–450 ms) and analyzed using repeated-measures ANOVAs with factors Emotional Salience, Dialogue Coherence, Hemisphere, and Region of Interest (ROI). Multiple comparisons in topographical analyses were controlled using false discovery rate (FDR) correction.

Results

The results are presented in four stages. We first report the dialogue-level validation and lexical–semantic control analyses, followed by behavioral results from the MEG study. We then examine the effects of emotional salience and dialogue coherence on neural speech tracking. Finally, TRF analyses characterize the temporal and cortical organization of emotional enhancement in neural speech tracking.

Dialogue Validation and Lexical-Semantic Control Analyses

Independent ratings confirmed the intended differences in both emotional-content and coherent categories (SI Figure S2A). Emotional-content dialogues were rated as more arousing, $F(1, 124) = 1122.70$, $p < .001$, lower in valence, $F(1, 124) = 991.47$, $p < .001$, and more socially relevant, $F(1, 124) = 106.99$, $p < .001$, than neutral dialogues. Coherent dialogues were rated as more coherent than incoherent dialogues, $F(1, 124) = 2738.30$, $p < .001$. Perceived coherence also showed an Emotion $\times$ Coherence interaction, $F(1, 124) = 17.33$, $p < .001$, reflecting a somewhat smaller coherent–incoherent difference for emotional-content dialogues.

The categories also differed in several lexical and semantic dimensions (SI Figure S2B-C). Coherent dialogues were easier to imagine than incoherent dialogues, $F(1, 124) = 132.51$, $p < .001$, and emotional-content dialogues were slightly more imageable than neutral dialogues, $F(1, 124) = 5.70$, $p = .041$. Semantic dispersion was higher for incoherent than coherent dialogues, $F(1, 124) = 199.52$, $p < .001$, and also varied with emotional content, $F(1, 124) = 41.71$, $p < .001$, with an Emotion $\times$ Coherence interaction, $F(1, 124) = 22.78$, $p < .001$. Narrative density was higher in emotional-content dialogues, $F(1, 124) = 5.23$, $p = .049$. Neutral dialogues contained higher-frequency, $F(1, 124) = 15.65$, $p < .001$, and longer words, $F(1, 124) = 8.04$, $p = .013$; word ambiguity and morphological feature count showed no reliable effects. As several ancillary lexical and discourse-level properties differed between categories, we ran follow-up analyses testing whether these differences accounted for the effects observed in neural data.

Behavioral Results

As part of the main MEG experiment, each dialogue was followed by two behavioral measures: a rating of perceived emotional load and true–false questions assessing recall accuracy of the dialogue content.

Emotional dialogues ($M = 2.40$, $SD = 0.69$) were perceived as more affectively charged, when compared to neutral ones ($M = 1.06$, $SD = 0.24$), ordinal GEE estimate = 14.14, $SE = 0.14$, $p < .001$. This indicated that participants were sensitive to the emotionality embedded in the conversation content. Neither dialogue coherence, ordinal GEE estimate =

0.08, $SE = 0.04$, $p = .06$, nor the Emotion $\times$ Coherence interaction, ordinal GEE estimate = -0.003, $SE = 0.04$, $p = .93$, significantly affected perceived emotional load.

Recall accuracy for the coherent dialogue content was higher ($M = 0.89$, $SD = 0.18$) than for incoherent ones ($M = 0.80$, $SD = 0.21$), implying that semantic continuity had facilitatory effect on memory performance, $F(1, 30) = 53.73$, $p < .001$, $\eta_G^2 = 0.199$. Emotional salience had no significant effect on recall accuracy, $F(1, 30) = 0.81$, $p = .376$, $\eta_G^2 = .004$, and there was no Emotion $\times$ Coherence interaction, $F(1, 30) = 0.08$, $p = .779$, $\eta_G^2 < .001$.

Emotional Salience Enhances Neural Tracking of Attended Speech

SER showed better accuracy for attended, compared to distractor speech, demonstrating a robust attentional enhancement in neural speech tracking, rm-ANOVA: $F(1, 30) = 408.79$, $p < .001$, $\eta_G^2 = 0.56$ (Table 1). SER accuracy was also higher for emotional than for neutral dialogues, $F(1, 30) = 26.66$, $p < .001$, $\eta_G^2 = 0.06$, whereas coherence had no impact on overall tracking accuracy ($p = .83$). Importantly, the influence of emotional salience depended on attentional engagement: emotional content enhanced SER only for attended speech, regardless of dialogue coherence (Figure 2A), Attention $\times$ Emotion $\times$ Coherence interaction: $F(1, 30) = 6.17$, $p = 0.019$, $\eta_G^2 = 0.02$, (SI Table S1).

We next tested whether the differences in lexical and semantic properties of the dialogue categories accounted for the emotional-content effect on target SER. The categorical Emotion $\times$ Coherence model provided the best fit with the data, AIC = -434.08, BIC = -419.82. Adding lexical controls did not improve model fit, AIC = -428.97, BIC = -403.30, nor did adding semantic-control variables, AIC = -427.45, BIC = -401.78. In both control models, the emotional-content effect remained stable (SI Table S2).

Because emotional-content dialogues were found to be lower in valence, and higher in arousal and social relevance, we next asked whether target SER was primarily associated with any one of these dimensions or with their combination, and whether the dimensional account outperformed the planned categorical Emotion $\times$ Coherence model. The continuous model showed a significant arousal $\times$ valence $\times$ social relevance interaction, $b = 0.0155$, $SE = 0.0076$, $t = 2.03$, $p = .044$ (SI Table S2). However, it provided a less parsimonious account than the categorical emotional-content model, AIC = -426.95, BIC = -398.43, compared with AIC = -434.08, BIC = -419.82. Thus, we retain the categorical model as our main focus when interpreting the findings.

Emotional Salience Modulates Temporal Dynamics of Speech Tracking

As neural tracking of audio-visual conversation stimuli has been shown to vary according to the unfolding dialogue (Wikman et al., 2021, 2024), we further examined how neural tracking unfolds across conversational turns. Emotional salience affected the temporal evolution of SER and was modulated by dialogue coherence, rm-ANOVA: $F(4.23, 126.98) = 4.58$, $p = 0.001$, $\eta_G^2 = 0.03$ (Table 2). Specifically, Emotional salience selectively increased the parabolic arc of the SER trajectory in coherent dialogues, whereas dialogue coherence modulated the timing of emotional effects, with coherent emotional dialogues showing an earlier peak, compared to a more sustained increase in incoherent dialogues (Figure 2B, SI Table S1).

Emotional Modulation of Cortical Responses To Continuous Speech

To examine how emotional salience affects the temporal and topographical organisation of neural speech processing, we analysed TRFs for emotionally salient and neutral dialogues in both coherent and incoherent conditions. Grand-averaged TRFs exhibited a distinct two-peak structure, with the strongest activations in temporal-frontal regions (Figure 3A). Based on these patterns, lateral cortical regions associated with speech and narrative comprehension were chosen for further analysis (Figure 3B, see SI Figure S3 for corresponding right-hemisphere waveforms; Fedorenko et al. 2024; Glasser et al, 2016; Ji et al, 2019; Simony et al. 2016).

In the mid-latency TRF window (100–250 ms), the effect of emotional salience on neural responses varied across cortical regions, depending on the dialogue coherence (Emotion x Coherence x ROI: $F(4.11, 123.21) = 4.96, p < 0.001$; Table 3). Under incoherence, Emotional salience increased response amplitudes in temporal and parietal regions, while coherent dialogues showed emotional enhancement primarily in the frontal regions (Figure 4, see SI Figure S4 for full ROI distributions). Thus, emotional-content modulation of mid-latency auditory–language responses depended on dialogue coherence and cortical region.

In the late-latency window (270–450 ms), emotionally salient speech produced stronger responses than neutral speech, $F(1, 30) = 7.56, p = .01, \eta_G^2 = 0.007$ (Table 4), independent of dialogue coherence ($p > .05$). Overall, late TRFs exhibited a broad enhancement for emotional speech in the temporal lobe that was less dependent on dialogue coherence than the mid-latency responses (Figure 5, see SI Figure S5 for full ROI distributions). Together, these findings indicate that emotional meaning influences attentional tracking of conversations at two stages of speech comprehension: affecting both earlier mid-latency auditory–language encoding and later evaluative processing of socio-emotional meaning (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006).

Discussion

Using a combined encoding–decoding approach, we show that emotional semantic speech content robustly enhances sustained attentional engagement in conversation in naturalistic cocktail-party settings. By reconstructing the speech envelope from neural activity, we demonstrate that emotional dialogues with negatively valenced, high-arousal and socially significant content enhance neural tracking of attended speech, with a distinct nonlinear temporal trajectory that is shaped by the semantic structure of the unfolding dialogue. Control analyses confirmed that the observed SER effects were not explained by ancillary lexical or semantic differences between the dialogue conditions, such as word typicality, word length, or narrative complexity. Complementary speech-to-brain encoding further revealed context-dependent and context-general effects of emotional salience across frontal, temporal, and parietal brain regions. Together, these findings show that emotional significance conveyed through speech content enhances the neural tracking of attended conversation, while the evolving semantic context shapes the temporal and cortical expression of this enhancement. These effects are observed in a distributed cortical network involved in auditory, semantic, and socio-cognitive processing.

Using listeners' brain activity to reconstruct concurrent conversational speech, we found that emotional salience selectively enhanced neural tracking of attended, but not ignored, conversations. While enhanced tracking of attended speech is well established (Ding & Simon, 2012; Fiedler et al., 2019; O'Sullivan et al., 2019; Zion Golumbic et al., 2013), our results extend this work by showing that attentional tracking is further modulated by higher-order socio-emotional meaning, particularly when speech conveys negatively valenced, highly arousing, and socially relevant information. The absence of a corresponding emotional-content effect for ignored speech indicates that this enhancement depended on the attended status of the dialogue and is therefore contingent on task relevance. This selectivity was mirrored in participants' subjective ratings, as emotionally salient dialogues were consistently perceived as more affectively intense. Independent validation showed that emotional dialogues combined higher arousal, lower valence, and greater social relevance, suggesting that their emotional significance was carried by a conjoint affective–social profile. Hence, the enhanced neural tracking may not be attributable to negative valence alone (with a possible contribution of negativity bias), but rather to the broader combination of negative valence, heightened arousal, and social relevance. Consistent with this, target SER was related to the combined pattern of arousal, valence, and social relevance. However, this

dimensional model did not explain the observed SER better than the categorical emotion contrast, which provided the better overall account.

Complementary TRF analyses showed that this modulation differed across response latencies and cortical regions. At mid-latency (100 - 250 ms), emotionally salient content enhanced responses in auditory, posterior language, and default-mode regions during semantically incoherent dialogues, whereas coherent dialogues showed stronger effects in inferior frontal language and default-mode regions. Thus, emotional-content modulation of mid-latency auditory–language responses depended on contextual structure. At later latencies (270–450 ms), emotional content elicited broader enhancements across temporal language and default mode regions regardless of coherence, suggesting a sustained effect of emotional salience during semantic and socio-cognitive integration. Together, these findings indicate that emotional content modulates speech processing at multiple stages through interactions between auditory–language networks and default mode regions supporting semantic comprehension and social cognition (Brodbeck et al., 2022; Citron, 2012; Fedorenko et al., 2024; Mar, 2011; Yamashita et al., 2025).

SER accuracy for attended dialogues evolved nonlinearly across dialogue turns. Importantly, this temporal profile depended on both the emotional salience of the speech and the dialogue’s semantic structure. In emotionally salient coherent dialogues, neural tracking showed a relatively early rise followed by attenuation, whereas emotionally salient incoherent dialogues exhibited a more sustained increase with a later peak. Similar nonlinear changes in attentional tracking of audiovisual speech have been reported previously (Wikman et al., 2021, 2024) and interpreted within an expansion–renormalization framework. In this framework, processing of behaviorally relevant information initially recruits additional neural resources before becoming more selective as an effective task-relevant representation is established (Kilgard, 2012; Wikman et al., 2021, 2024). Importantly, the present findings extend this account by showing that the precise temporal profile varies with both emotional salience and semantic coherence of the content. Emotionally salient sensory events have been associated with attentional prioritization, as they are more likely to offer behaviorally significant information (Astudillo et al., 2018; Fischer et al., 2025; Schacht & Vrtička, 2018). Emotional conversation content may therefore recruit broader processing resources. Consistent with this, emotionally arousing narrative moments have been associated with greater functional integration across large-scale neural networks (Park et al., 2025). Within an expansion–renormalization account and in accordance with our observations, semantic continuity may determine how quickly this enhanced processing stabilizes. In coherent emotional dialogues, successive turns contribute to the same evolving context, potentially allowing enhanced processing to stabilize more rapidly to a more efficient state. In turn, when the content is incoherent, successive turns do not support a stable dialogue-level representation, potentially requiring enhanced processing for longer, which could lead to the more sustained SER increase we observed.

To further examine why semantic coherence may have altered the time course of the emotional enhancement, we turn to predictive-processing theories. Here, our finding that coherence modulated the effect of emotional-content in the mid-latency TRFs is of particular relevance. Predictive accounts propose that regularities in ongoing auditory input generate expectations about upcoming information, with violations producing error-related modulation (Winkler et al., 2009; Schröger et al., 2015). Such responses emerge around 150 ms, where prediction and attention interact in auditory processing, overlapping with our 100–250 ms TRF window (Schröger et al., 2015). Within this window, emotional-content effects were stronger in temporal and parietal regions under incoherence. This pattern is compatible with emotional salience enhancing error-related processing when attended speech conflicts with the evolving context. This increased responsiveness in lower-level auditory processing could then delay stabilization of the enhanced processing state, thereby contributing to the more

sustained SER enhancement under incoherence. Because the TRFs did not directly quantify prediction error, this interpretation should be regarded as consistent with, rather than direct evidence of, predictive processing.

Taken together, our findings show that emotionally salient content enhances neural processing of attended conversation, strengthening speech tracking and modulating cortical responses across distributed temporal, parietal, and frontal regions. Importantly, this enhancement was not temporally or contextually uniform, as semantic continuity shaped both how neural tracking of emotional content developed across dialogue turns and how it affected cortical responses. Accordingly, emotional significance may transiently enhance processing across auditory, language, and control systems, while semantic continuity shapes how this enhanced processing evolves across the conversational arc (Clark, 2013; Diederer & Fletcher, 2021; Park et al., 2025; Song et al., 2021). This combination of preferential processing and context-sensitive modulation may be particularly important in real-time interaction, where comprehension and response preparation must proceed in parallel across speaker switches of only 100–200 ms (Bögels et al., 2015; Brehm & Meyer, 2021; Heldner & Edlund, 2010; Magyari, 2022; Stivers et al., 2009). Emotional salience may therefore contribute to the rapid coordination required during emotionally engaging interaction (Stevanovic & Peräkylä, 2015).

Limitations

In the present study, emotional salience was operationalized as a combination of negative valence, heightened arousal, and high social relevance. This region of the emotional salience space is both empirically grounded and theoretically meaningful, but the study leaves open whether similar effects emerge for other types of emotionally salient conversational content (e.g., positively valenced, high-arousal, socially relevant content). Future research should therefore examine other regions of this space to further delineate the scope of the emotional enhancement effect. Furthermore, emotional expressivity was limited to the semantic content of the dialogues, allowing us to examine how content of speech modulates neural response associated with attentional engagement. However, human expressivity depends on the integration of various expressive modalities, such as paralinguistic features like changes in intonations, as well as communicative gestures like hand movements that likely co-occur with effects presented here. Future work should therefore extend the inquiry to conversations where the participants actively take part in the dialogue and are more free to use their communicative repertoire as a whole.

Conclusions

Our findings show that emotional salience dynamically shapes attentional engagement in conversational speech in multi-speaker environments. Emotional semantic content enhances neural tracking of attended speech and modulates cortical responses to continuous speech. Furthermore, the interaction between emotional significance and semantic continuity demonstrates that cortical responses to socially and motivationally significant cues are shaped by the evolving conversational context. Rather than acting as a simple amplifier of sensory encoding, emotional salience actively recruits and distributes neural resources, reshaping language processing to meet changing conversational demands and enabling flexible adjustment between stable comprehension and adaptive responsiveness.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>

- 669 Astudillo, C., Muñoz, K., & Maldonado, P. E. (2018). Emotional content modulates
 670 attentional visual orientation during free viewing of natural images. *Frontiers in*
 671 *Human Neuroscience*, 12, Article 459. <https://doi.org/10.3389/fnhum.2018.00459>
- 672 Bayer, M., Sommer, W., & Schacht, A. (2010). Reading emotional words within sentences:
 673 The impact of arousal and valence on event-related potentials. *International Journal*
 674 *of Psychophysiology*, 78(3), 299–307. <https://doi.org/10.1016/j.ijpsycho.2010.09.004>
- 675 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and
 676 powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series*
 677 *B (Methodological)*, 57(1), 289–300. [https://doi.org/10.1111/j.2517-](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
 678 [6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
- 679 Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing*
 680 *Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- 681 Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning
 682 occur midway through an incoming question in conversation. *Scientific Reports*, 5(1),
 683 12881. <https://doi.org/10.1038/srep12881>
- 684 Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and
 685 the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*,
 686 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- 687 Brehm, L., & Meyer, A. S. (2021). Planning when to say: Dissociating cue use in utterance
 688 initiation using cross-validation. *Journal of Experimental Psychology: General*,
 689 150(9), 1772–1799. <https://doi.org/10.1037/xge0001012>
- 690 Brodbeck, C., Bhattasali, S., Cruz Heredia, A. A., Resnik, P., Simon, J. Z., & Lau, E. (2022).
 691 Parallel processing in speech perception with local and global representations of
 692 linguistic context. *eLife*, 11, e72056. <https://doi.org/10.7554/eLife.72056>
- 693 Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).
 694 Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension
 695 of Natural, Narrative Speech. *Current Biology*, 28(5), 803–809.e3.
 696 <https://doi.org/10.1016/j.cub.2018.01.080>
- 697 Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with
 698 Two Ears. *The Journal of the Acoustical Society of America*, 25(5), 975–979.
 699 <https://doi.org/10.1121/1.1907229>
- 700 Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of
 701 recent electrophysiological and hemodynamic neuroimaging studies. *Brain and*
 702 *Language, Neurobiology of Language* 2010, 122(3), 211–226.
 703 <https://doi.org/10.1016/j.bandl.2011.12.007>
- 704 Clark, A. (2013). The many faces of precision (Replies to commentaries on “Whatever next?
 705 Neural prediction, situated agents, and the future of cognitive science”). *Frontiers in*
 706 *Psychology*, 4, 270. <https://doi.org/10.3389/fpsyg.2013.00270>
- 707 Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal
 708 Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals
 709 to Continuous Stimuli. *Frontiers in Human Neuroscience*, 10.
 710 <https://doi.org/10.3389/fnhum.2016.00604>

- 711 de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in
712 a web browser. *Behavior Research Methods*, 47(1), 1-12. doi:10.3758/s13428-014-
713 0458-y
- 714 Diederer, K. M. J., & Fletcher, P. C. (2021). Dopamine, Prediction Error and Beyond. *The*
715 *Neuroscientist*, 27(1), 30–46. <https://doi.org/10.1177/1073858420907591>
- 716 Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment
717 to speech reflects phoneme-level processing. *Current Biology*, 25(19), 2457–2465.
718 <https://doi.org/10.1016/j.cub.2015.08.030>
- 719 Ding, J., Wang, L., & Yang, Y. (2015). The dynamic influence of emotional words on
720 sentence processing. *Cognitive, Affective, & Behavioral Neuroscience*, 15(1), 55–68.
721 <https://doi.org/10.3758/s13415-014-0315-6>
- 722 Ding, J., Wang, L., & Yang, Y. (2016). The dynamic influence of emotional words on
723 sentence comprehension: An ERP study. *Cognitive, Affective, & Behavioral*
724 *Neuroscience*, 16(3), 433–446. <https://doi.org/10.3758/s13415-016-0403-x>
- 725 Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while
726 listening to competing speakers. *Proceedings of the National Academy of Sciences*,
727 109(29), 11854–11859. <https://doi.org/10.1073/pnas.1205381109>
- 728 Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind
729 within the broader landscape of the human brain. *Nature Reviews Neuroscience*,
730 25(5), 289–312. <https://doi.org/10.1038/s41583-024-00802-4>
- 731 Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational
732 information sharing as prosocial behavior. *Journal of Personality and Social*
733 *Psychology*, 102(5), 1015–1030. <https://doi.org/10.1037/a0026650>
- 734 Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of
735 ignored speech facilitates neural selectivity in acoustically challenging conditions.
736 *NeuroImage*, 186, 33–42. <https://doi.org/10.1016/j.neuroimage.2018.10.057>
- 737 Fields, E. C., & Kuperberg, G. R. (2012). It's All About You: An ERP Study of Emotion and
738 Self-Relevance in Discourse. *NeuroImage*, 62(1), 562–574.
739 <https://doi.org/10.1016/j.neuroimage.2012.05.003>
- 740 Fischer, A., Postin, D., Meiners, L. J., Kulke, L., Vrtička, P., & Schacht, A. (2025).
741 Challenging the negativity bias in affective scene viewing: The role of social content.
742 *Social Cognitive and Affective Neuroscience*, 20(1), Article nsaf108.
743 <https://doi.org/10.1093/scan/nsaf108>
- 744 Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E.,
745 Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van
746 Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*,
747 536(7615), 171–178. <https://doi.org/10.1038/nature18933>
- 748 Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj,
749 R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data
750 analysis with MNE-Python. *Frontiers in Neuroscience*, 7.
751 <https://doi.org/10.3389/fnins.2013.00267>

- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, 86, 446–460. <https://doi.org/10.1016/j.neuroimage.2013.10.027>
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1), 35–42. <https://doi.org/10.1007/BF02512476>
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555–568. <https://doi.org/10.1016/j.wocn.2010.08.002>
- Hoemann, K., Lee, Y., Dussault, È., Devylder, S., Ungar, L. H., Geeraerts, D., & Mesquita, B. (2025). The construction of emotional meaning in language. *Communications Psychology*, 3(1), 99. <https://doi.org/10.1038/s44271-025-00255-0>
- Itkonen, S., Häikiö, T., Vainio, S. et al. LASTU: A psycholinguistic search tool for Finnish lexical stimuli. *Behav Res* 56, 6165–6178 (2024). <https://doi.org/10.3758/s13428-024-02347-x>
- Ji, J. L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., & Cole, M. W. (2019). Mapping the human brain's cortical-subcortical functional network organization. *NeuroImage*, 185, 35–57. <https://doi.org/10.1016/j.neuroimage.2018.10.006>
- Kanerva, J., Ginter, F., Chang, L.-H., Rastas, I., Skantsi, V., Kilpeläinen, J., Kupari, H.-M., Piirto, A., Saarni, J., Sevón, M., & Tarkka, O. (2024). Towards diverse and contextually anchored paraphrase modeling: A dataset and baselines for Finnish. *Natural Language Engineering*, 30(2), 319–353. <https://doi.org/10.1017/S1351324923000086>
- Kaufman, M., & Zion Golumbic, E. (2023). Listening to two speakers: Capacity and tradeoffs in neural speech tracking during Selective and Distributed Attention. *NeuroImage*, 270, 119984. <https://doi.org/10.1016/j.neuroimage.2023.119984>
- Kilgard, M. P. (2012). Harnessing plasticity to understand learning and treat disease. *Trends in Neurosciences*, 35(12), 715–722. <https://doi.org/10.1016/j.tins.2012.09.002>
- Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass*, 5(3), 108–125. <https://doi.org/10.1111/j.1749-818X.2010.00267.x>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4(1), Article 2515245920951503. <https://doi.org/10.1177/2515245920951503>
- Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *The European Journal of Neuroscience*, 31(1), 189–193. <https://doi.org/10.1111/j.1460-9568.2009.07055.x>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, 44(12), 1248–1263. [https://doi.org/10.1016/S0006-3223\(98\)00275-3](https://doi.org/10.1016/S0006-3223(98)00275-3)

- 792 Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., & Ginter, F. (2015). Towards universal
793 web parsebanks. In Proceedings of the Third International Conference on
794 Dependency Linguistics (Depling 2015) (pp. 211–220). Uppsala University.
- 795 Magyari, L. (2022). Predictions in Conversation. In J. Gervain, G. Csibra, & K. Kovács
796 (Eds.), *A Life in Cognition: Studies in Cognitive Science in Honor of Csaba Pléh* (pp.
797 59–75). Springer International Publishing. [https://doi.org/10.1007/978-3-030-66175-](https://doi.org/10.1007/978-3-030-66175-5_5)
798 [5_5](https://doi.org/10.1007/978-3-030-66175-5_5)
- 799 Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual*
800 *Review of Psychology*, 62(Volume 62, 2011), 103–134.
801 <https://doi.org/10.1146/annurev-psych-120709-145406>
- 802 Martín-Loeches, M., Fernández, A., Schacht, A., Sommer, W., Casado, P., Jiménez-Ortega,
803 L., & Fondevila, S. (2012). The influence of emotional words on sentence processing:
804 Electrophysiological and behavioral evidence. *Neuropsychologia*, 50(14), 3262–
805 3272. <https://doi.org/10.1016/j.neuropsychologia.2012.09.010>
- 806 Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400
807 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience*, 18(7),
808 1098–1111. <https://doi.org/10.1162/jocn.2006.18.7.1098>
- 809 O’Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G. M., Sheth, S. A., Mehta, A.
810 D., & Mesgarani, N. (2019). Hierarchical Encoding of Attended Auditory Objects in
811 Multi-talker Speech Perception. *Neuron*, 104(6), 1195–1209.e3.
812 <https://doi.org/10.1016/j.neuron.2019.09.007>
- 813 Park, J. S., Gollapudi, K., Ke, J., Nau, M., Pappas, I., & Leong, Y. C. (2025). Emotional
814 arousal enhances narrative memories through functional integration of large-scale
815 brain networks. *Nature Human Behaviour*, 1–20. [https://doi.org/10.1038/s41562-025-](https://doi.org/10.1038/s41562-025-02315-1)
816 [02315-1](https://doi.org/10.1038/s41562-025-02315-1)
- 817 Pinheiro, A. P. (2025). Behind a Voice There is a Speaker: Why Vocal Emotion Research
818 Needs to Become ‘Personal.’ *Affective Science*, 6(3), 562–574.
819 <https://doi.org/10.1007/s42761-025-00317-w>
- 820 Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the
821 cocktail party? A late locus of selective attention to natural speech. *European Journal*
822 *of Neuroscience*, 35(9), 1497–1503. [https://doi.org/10.1111/j.1460-](https://doi.org/10.1111/j.1460-9568.2012.08060.x)
823 [9568.2012.08060.x](https://doi.org/10.1111/j.1460-9568.2012.08060.x)
- 824 Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural
825 language processing toolkit for many human languages. In Proceedings of the 58th
826 Annual Meeting of the Association for Computational Linguistics: System
827 Demonstrations (pp. 101–108). Association for Computational Linguistics.
828 <https://doi.org/10.18653/v1/2020.acl-demos.14>
- 829 Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese
830 BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in
831 Natural Language Processing and the 9th International Joint Conference on Natural
832 Language Processing (EMNLP-IJCNLP) (pp. 3982–3992). Association for
833 Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>

- Robbins, M. L., & Karan, A. (2020). Who gossips and how in everyday life? *Social Psychological and Personality Science*, 11(2), 185–195. <https://doi.org/10.1177/1948550619837000>
- Sander, D., Grafman, J., & Zalla, T. (2003). The Human Amygdala: An Evolved System for Relevance Detection. *Reviews in the Neurosciences*, 14(4). <https://doi.org/10.1515/REVNEURO.2003.14.4.303>
- Schacht, A., & Vrtička, P. (2018). Spatiotemporal pattern of appraising social and emotional relevance: Evidence from event-related brain potentials. *Cognitive, Affective, & Behavioral Neuroscience*, 18(6), 1172–1187. <https://doi.org/10.3758/s13415-018-0629-x>
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24–30. <https://doi.org/10.1016/j.tics.2005.11.009>
- Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, 41(5), 641–664. <https://doi.org/10.1111/ejn.12816>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Song, H., Finn, E. S., & Rosenberg, M. D. (2021). Neural signatures of attentional engagement during narratives and its consequences for event memory. *Proceedings of the National Academy of Sciences of the United States of America*, 118(33), e2021905118. <https://doi.org/10.1073/pnas.2021905118>
- Stevanovic, M., & Peräkylä, A. (2015). Experience sharing, emotional reciprocity, and turn-taking. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00450>
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine & Biology*, 51(7), 1759. <https://doi.org/10.1088/0031-9155/51/7/008>
- Teoh, E. S., Ahmed, F., & Lalor, E. C. (2022). Attention Differentially Affects Acoustic and Phonetic Feature Encoding in a Multispeaker Environment. *Journal of Neuroscience*, 42(4), 682–691. <https://doi.org/10.1523/JNEUROSCI.1455-20.2021>
- van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17(3), 701–718. [https://doi.org/10.1016/S0926-6410\(03\)00196-4](https://doi.org/10.1016/S0926-6410(03)00196-4)
- Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., Ginter, F., & Pyysalo, S. (2019). Multilingual is not enough: BERT for Finnish. *arXiv preprint arXiv:1912.07076*.

- Wikman, P., Sahari, E., Salmela, V., Leminen, A., Leminen, M., Laine, M., & Alho, K. (2021). Breaking down the cocktail party: Attentional modulation of cerebral audiovisual speech processing. *NeuroImage*, 224, 117365. <https://doi.org/10.1016/j.neuroimage.2020.117365>
- Wikman, P., Salmela, V., Sjöblom, E., Leminen, M., Laine, M., & Alho, K. (2024). Attention to audiovisual speech shapes neural processing through feedback-feedforward loops between different nodes of the speech network. *PLOS Biology*, 22(3), e3002534. <https://doi.org/10.1371/journal.pbio.3002534>
- Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13(12), 532–540. <https://doi.org/10.1016/j.tics.2009.09.003>
- Yamashita, M., Kubo, R., & Nishimoto, S. (2025). Conversational content is organized across multiple timescales in the brain. *Nature Human Behaviour*, 9(10), 2066–2078. <https://doi.org/10.1038/s41562-025-02231-4>
- Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*, 24(1), 3–47. <https://doi.org/10.1080/02699930903205698>
- Zhang, Q., Ding, J., Zhang, Z., Yang, X., & Yang, Y. (2021). The effect of congruent emotional context in emotional word processing during discourse comprehension. *Journal of Neurolinguistics*, 59, 100989. <https://doi.org/10.1016/j.jneuroling.2021.100989>
- Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language, Neurobiology of Language* 2010, 122(3), 151–161. <https://doi.org/10.1016/j.bandl.2011.12.010>
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron*, 77(5), 980–991. <https://doi.org/10.1016/j.neuron.2012.12.037>

Figure Legends

Figure 1. The experimental design and analysis framework. (A) Example stimulus materials. The left panel shows an example of an audiovisual target dialogue, including the video stimulus and transcript, illustrating an emotionally salient coherent dialogue. The right panel shows a corresponding audio-only distractor dialogue, exemplifying a neutral incoherent conversation. (B) Structure of a single trial, in which each target dialogue type was presented with each distractor type. Trials are shown grouped by target condition for illustration; actual presentation order was pseudorandomized (Methods). (C) Schematic overview of the modelling approaches. Speech envelopes of the target (orange) and distractor (blue) dialogues for each condition were used separately to predict how neural activity responded to changes in the speech envelope with temporal response functions (TRFs). In the speech envelope reconstruction (SER) analysis, neural activity was used to

reconstruct the corresponding speech envelopes. Note that presented reconstructions and TRF waveforms are for illustrative purposes only.

Figure 2. Emotional salience enhances and temporally modulates neural speech tracking. (A) Mean SER accuracy (Fisher's z-transformed) for target and distractor speech as a function of emotional salience and dialogue coherence. Boxplots show empirical distributions, with individual observations overlaid. Diamonds indicate estimated marginal means with 95% confidence intervals. Horizontal bars indicate Holm-adjusted pairwise comparisons; exact p values are shown for significant contrasts and n.s. for the corresponding nonsignificant distractor contrasts. (B) Turn-by-turn trajectories of SER accuracy for attended speech. Points and solid lines show observed means with standard error bars. Dashed lines indicate cubic model predictions, illustrating an earlier peak for coherent emotional dialogues followed by attenuation and a later, more sustained increase for incoherent emotional dialogues.

Figure 3. Temporal response functions reveal latency- and region-dependent cortical responses to continuous speech. (A) Grand-averaged TRF waveform in the auditory cortex for both hemispheres across all conditions. The grey highlights denote the mid-latency (100-250 ms) and late-latency (270-450 ms) windows from which mean activations were extracted for statistical analysis. The parcel-level topographies of TRF responses within the chosen time-windows are depicted on the right. (B) Regions of interest (ROIs) chosen for statistical analysis, and left hemisphere TRF-waveforms for all conditions in the target dialogues.

Figure 4. Emotional salience and dialogue coherence modulate mid-latency cortical responses. (A) Topographical distribution of significant differences between emotional and neutral dialogues in the mid-latency window (100–250 ms), shown separately for coherent and incoherent dialogues. Maps show FDR-corrected effects, and color intensity represents the strength of emotional enhancement (Emotional > Neutral) as t -ratios. (B) Mid-latency TRF amplitudes averaged across hemispheres in regions of interest showing significant condition differences for emotional coherent (EC), emotional incoherent (EI), neutral coherent (NC), and neutral incoherent (NI) dialogues. Boxplots and overlaid points show the empirical distributions across participants ($n = 31$). Diamonds indicate estimated marginal means with 95% confidence intervals. Horizontal bars indicate significant FDR-adjusted pairwise contrasts between conditions.

Figure 5. Emotional salience enhances late-latency cortical responses. (A) Topographical distribution of significant differences between emotional and neutral dialogues in the late-latency window (270–450 ms), collapsed across dialogue coherence. The map shows FDR-corrected effects, and color intensity represents the strength of emotional enhancement (Emotional > Neutral) as t -ratios. (B) Late-latency TRF amplitudes averaged across hemispheres in regions of interest showing significant emotional–neutral differences, collapsed across dialogue coherence. Boxplots and overlaid points show the empirical distributions across participants ($n = 31$). Diamonds indicate estimated marginal means with 95% confidence intervals. Horizontal bars indicate significant FDR-adjusted emotional–neutral contrasts within each region of interest.

Table Legends

Table 1. Repeated-Measures Analysis of Variance for Overall Speech Envelope Reconstruction Accuracy. Type III repeated-measures ANOVA testing the within-participant effects of Attention (attended, distractor), Emotional Salience (emotional, neutral),

and Dialogue Coherence (coherent, incoherent) on overall speech envelope reconstruction (SER) accuracy. Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Table 2. Turn-Wise Repeated-Measures Analysis of Variance for Attended Speech Envelope Reconstruction Accuracy. Type III repeated-measures ANOVA testing the within-participant effects of Emotional Salience, Dialogue Coherence, and conversational Turn on SER accuracy for attended speech. Greenhouse-Geisser-corrected degrees of freedom are reported where the assumption of sphericity was violated. Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Table 3. Repeated-Measures Analysis of Variance for Mid-Latency TRF Amplitudes. Type III repeated-measures ANOVA testing the within-participant effects of Emotional salience, Dialogue Coherence, Region of Interest (ROI), and Hemisphere on temporal response function (TRF) amplitudes in the mid-latency window (100–250 ms). Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Table 4. Repeated-Measures Analysis of Variance for Late-Latency TRF Amplitudes. Type III repeated-measures ANOVA testing the within-participant effects of Emotional Salience, Dialogue Coherence, Region of Interest (ROI), and Hemisphere on temporal response function (TRF) amplitudes in the late-latency window (270–450 ms). Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Table 1. Repeated-Measures Analysis of Variance for Overall Speech Envelope Reconstruction Accuracy. Type III repeated-measures ANOVA testing the within-participant effects of Attention (attended, distractor), Emotional Salience (emotional, neutral), and Dialogue Coherence (coherent, incoherent) on overall speech envelope reconstruction (SER) accuracy. Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Effect	<i>df</i>	<i>F</i>	η^2_G	<i>p</i>
Attention	1, 30	408.79	.56	< .001
Emotion	1, 30	26.66	.06	< .001
Coherence	1, 30	0.05	.00	.831
Attention × Emotion	1, 30	13.11	.05	.001
Attention × Coherence	1, 30	0.36	.00	.555
Emotion × Coherence	1, 30	2.31	.01	.139
Attention × Emotion × Coherence	1, 30	6.17	.02	.019

Table 2. Turn-Wise Repeated-Measures Analysis of Variance for Attended Speech Envelope Reconstruction Accuracy. Type III repeated-measures ANOVA testing the within-participant effects of Emotional Salience, Dialogue Coherence, and conversational Turn on SER accuracy for attended speech. Greenhouse-Geisser-corrected degrees of freedom are reported where the assumption of sphericity was violated. Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

Effect	<i>df</i>	<i>F</i>	η^2_G	<i>p</i>
Emotion	1.00, 30.00	31.04	.06	< .001
Coherence	1.00, 30.00	0.21	.00	.652
Turn	4.02, 120.56	8.47	.04	< .001
Emotion × Coherence	1.00, 30.00	0.24	.00	.627
Emotion × Turn	4.32, 129.71	3.18	.02	.013
Coherence × Turn	4.33, 129.92	3.33	.02	.010
Emotion × Coherence × Turn	4.23, 126.98	4.58	.03	.001

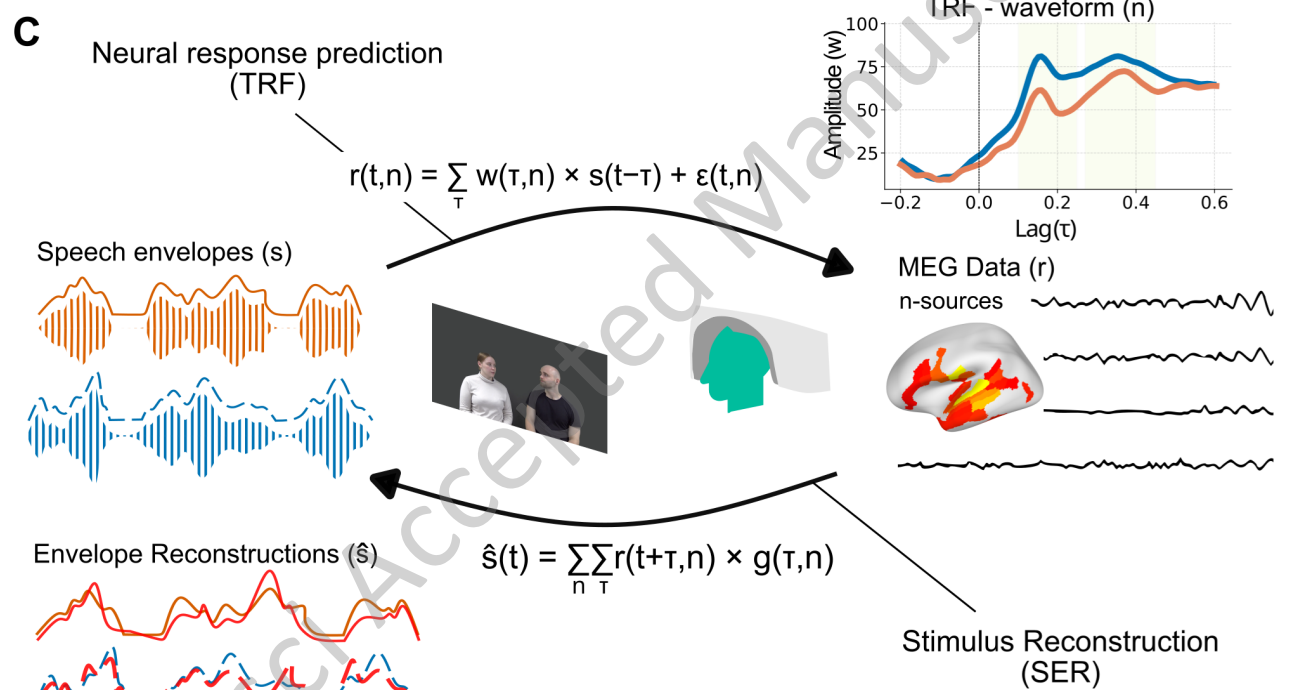
Table 3. Repeated-Measures Analysis of Variance for Mid-Latency TRF Amplitudes. Type III repeated-measures ANOVA testing the within-participant effects of Emotional salience, Dialogue Coherence, Region of Interest (ROI), and Hemisphere on temporal response function (TRF) amplitudes in the mid-latency window (100–250 ms). Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

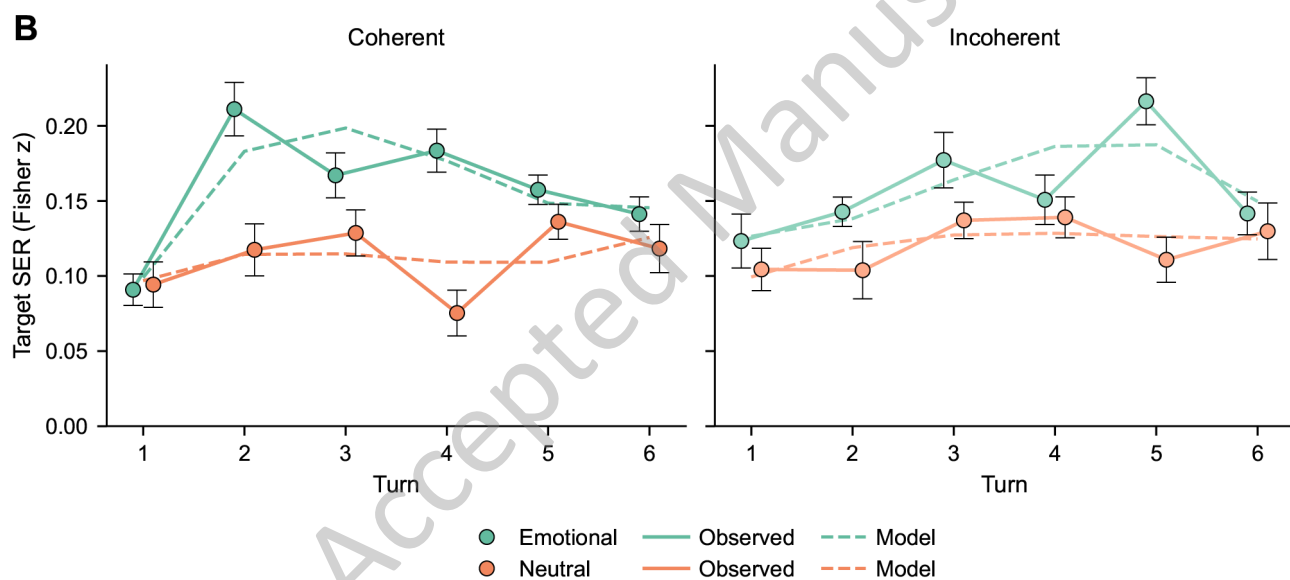
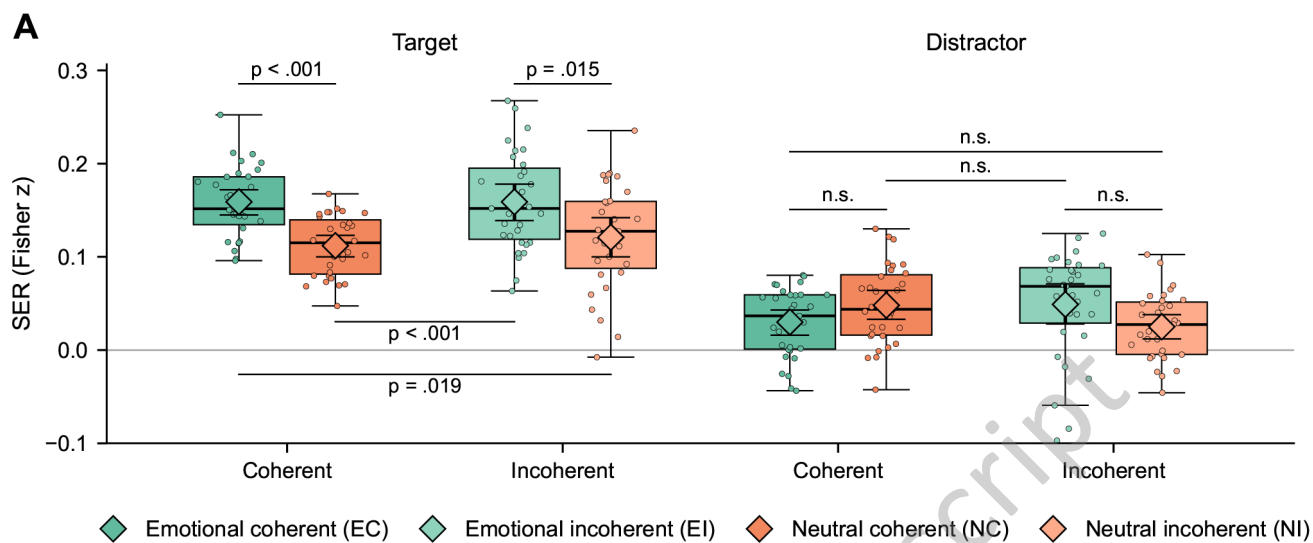
Effect	<i>df</i>	<i>F</i>	η^2_G	<i>p</i>
Emotion	1.00, 30.00	17.17	.014	< .001
Coherence	1.00, 30.00	3.43	.003	.074
ROI	2.16, 64.79	44.02	.234	< .001
Hemisphere	1.00, 30.00	1.68	.002	.205
Emotion × Coherence	1.00, 30.00	6.22	.003	.018
Emotion × ROI	3.50, 105.09	5.73	.006	< .001
Coherence × ROI	3.14, 94.24	1.46	.002	.229
Emotion × Hemisphere	1.00, 30.00	0.00	.000	.949
Coherence × Hemisphere	1.00, 30.00	0.55	.000	.463
ROI × Hemisphere	3.27, 98.12	2.67	.008	.047
Emotion × Coherence × ROI	4.11, 123.21	4.96	.004	< .001
Emotion × Coherence × Hemisphere	1.00, 30.00	1.30	.000	.263
Emotion × ROI × Hemisphere	3.52, 105.52	1.23	.001	.302
Coherence × ROI × Hemisphere	3.25, 97.54	0.32	.000	.826
Emotion × Coherence × ROI × Hemisphere	3.77, 112.98	0.84	.001	.497

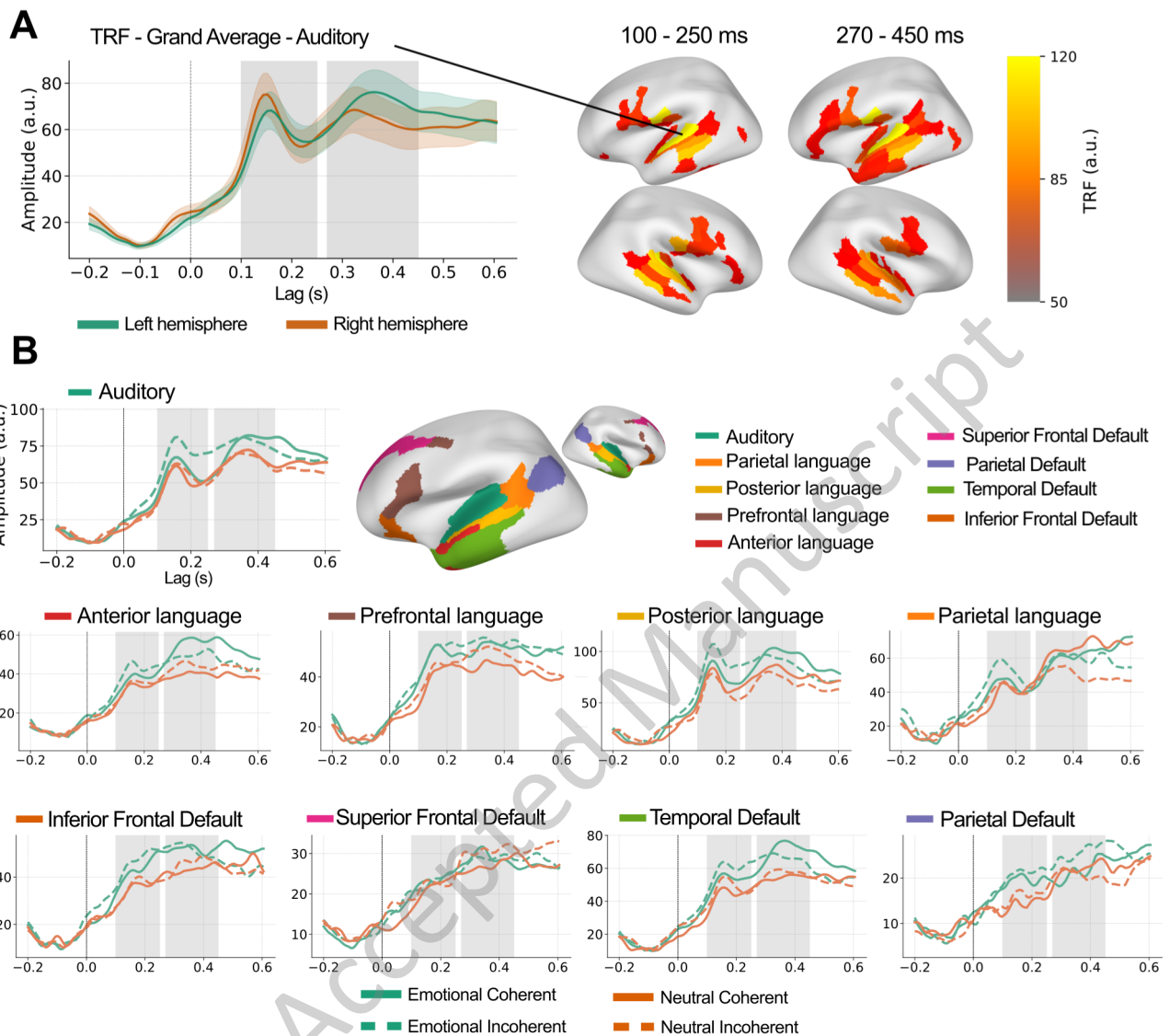
Table 4. Repeated-Measures Analysis of Variance for Late-Latency TRF Amplitudes.

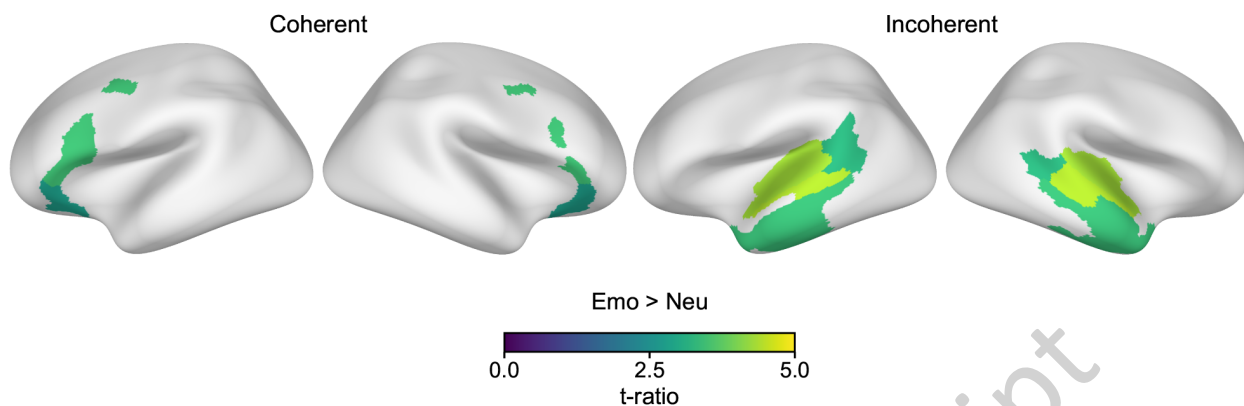
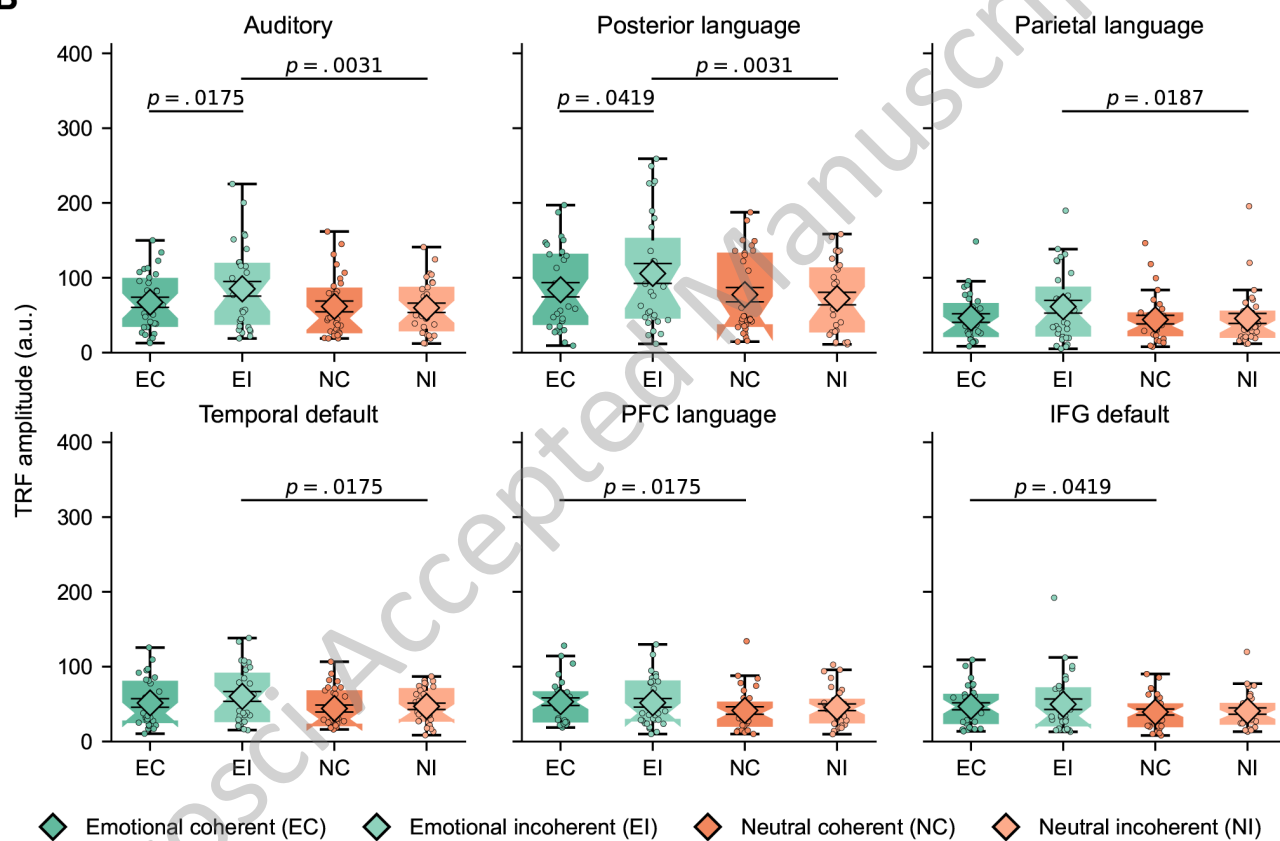
Type III repeated-measures ANOVA testing the within-participant effects of Emotional Salience, Dialogue Coherence, Region of Interest (ROI), and Hemisphere on temporal response function (TRF) amplitudes in the late-latency window (270–450 ms). Generalized eta squared (η^2_G) is reported as the effect size. Exact p values are reported unless $p < .001$.

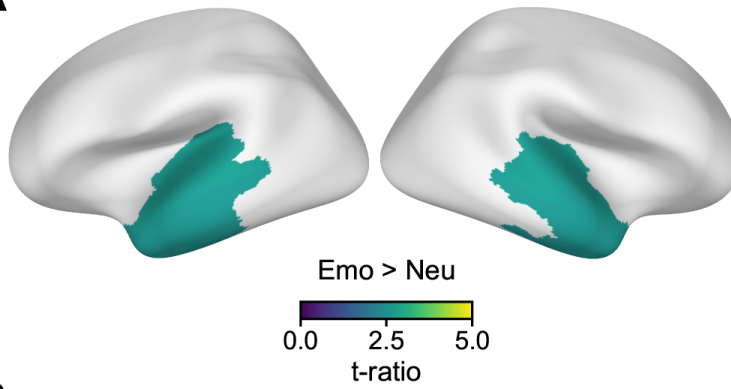
Effect	<i>df</i>	<i>F</i>	η^2_G	<i>p</i>
Emotion	1.00, 30.00	7.56	.007	.010
Coherence	1.00, 30.00	1.25	.001	.273
ROI	2.37, 71.09	34.91	.197	< .001
Hemisphere	1.00, 30.00	9.07	.017	.005
Emotion × Coherence	1.00, 30.00	0.51	.000	.479
Emotion × ROI	3.04, 91.19	3.60	.005	.016
Coherence × ROI	3.22, 96.71	0.48	.001	.708
Emotion × Hemisphere	1.00, 30.00	1.62	.000	.213
Coherence × Hemisphere	1.00, 30.00	3.45	.002	.073
ROI × Hemisphere	3.46, 103.82	2.22	.008	.081
Emotion × Coherence × ROI	4.20, 125.87	0.90	.001	.468
Emotion × Coherence × Hemisphere	1.00, 30.00	1.25	.000	.272
Emotion × ROI × Hemisphere	4.68, 140.25	0.82	.001	.530
Coherence × ROI × Hemisphere	3.97, 119.18	2.14	.002	.080
Emotion × Coherence × ROI × Hemisphere	2.91, 87.29	1.04	.001	.376







A**B**

A**B**