No Risky Bets: The Brain Avoids All-In Predictions During Naturalistic Multitalker Listening.

Giorgio Piazza[a†*], Marco Sala[a†*], Rebecca Guerrini[a], Martin M. Winchester[b,c], Francesca Peressotti[a,d]

[a]*Department of Developmental Psychology and Socialisation (DPSS), University of Padova, Padova, Italy;*

[b]*School of Computer Science and Statistics, University of Dublin, Trinity College, Dublin, Ireland;*

[c]*ADAPT Centre, Trinity College Institute of Neuroscience, Dublin, Ireland;*

[d]*Padua Neuroscience Center, University of Padua, Padua, Italy;*

[†] These authors contributed equally to this work.

*Corresponding authors:

giorgio.piazza@unipd.it;

marco.sala.2@phd.unipd.it;

Via Venezia, 8, 35131 Padova PD, Italy.

Abstract

Speech comprehension requires dealing with variability and uncertainty, especially when we are not familiar with the talker. Listeners achieve robust speech comprehension by tracking acoustic variability across talkers and implementing predictions based on their prior knowledge and contextual information. However, how perceptual adaptation to speech and predictive processing influence each other is still largely unknown.

In this EEG study, we examined how listeners process continuous speech when exposed to a single familiar talker (Single condition) versus multiple unfamiliar talkers (Multi condition), all with native Italian pronunciation. By applying multivariate Temporal Response Function (TRF), we aimed to determine whether increased talker variability in the Multi condition influences phonemic encoding and predictive processing, as indexed by neural responses to cohort entropy (phonological uncertainty among lexical candidates) and semantic surprisal (lexical prediction error).

Results showed increased neural responses to phonemic categories and to cohort entropy in the Multi than in the Single condition. These effects suggest that acoustic variability across talkers increases uncertainty in perceiving individual phonemes, prompting the brain to avoid committing to specific lexical candidates during word processing. Importantly, semantic surprisal responses were comparable between conditions, indicating that the modulation primarily affects early stages of processing. This study provided a critical test for how listeners compensate for increased bottom-up uncertainty by implementing probabilistic predictions during language comprehension. Our findings also offer insights into the flexibility of the human speech processing system in dynamically adapting to variable social and acoustic environments.

1. Introduction

Humans interact daily with many different talkers, most of whom are unfamiliar and vary widely in how they produce speech. For example, during a conference question time, listeners may be exposed to a rapid succession of unfamiliar talkers that differ in accent (regional or non-native accent), voice quality, fundamental frequency, speech rate, and articulation style (etc.). Such variability reveals a fundamental challenge of spoken language comprehension: the acoustic signal is continuous and highly variable, and there is not a deterministic mapping between its physical properties and the abstract linguistic categories it conveys such as phonemes, words, and syntactic structures. The brain must infer abstract representations from a signal that is both transient and shaped by the unique characteristics of each speaker. Yet, listeners typically comprehend speech with remarkable ease. One explanation for this efficiency is that comprehension is not purely reactive but also anticipatory (1, 2). This idea informs a broad range of language processing models, from speech perception to word recognition and sentence or discourse comprehension. Although prediction is widely regarded as a mechanism involved across multiple stages, it remains unclear how different levels of processing interact during spoken language comprehension. Specifically, whether acoustic variability across talkers influences the brain's ability to generate higher-level predictions during word recognition and sentence processing remains an open question.

A common feature of these models is that prediction operates probabilistically: the brain continuously forms graded expectations reflecting the inherent uncertainty in both the speech signal and the linguistic content (2–4). Research in the field of speech perception showed that phonemic categories form distributions across multiple acoustic dimensions, which can vary across talkers (3, 5). Exposure to a talker enables the construction of distributional models of acoustic cues, which allow to predict cue values based not only on what is being said but also on who is saying it. From this perspective, listeners must select the most appropriate generative model for the current talker in order to achieve robust speech perception. This is consistent with empirical evidence showing that speech perception can be influenced by the listeners' expectations about the talker, such as their dialect background (6, 7), ethnicity (8), accent (9), gender (10, 11), and age

(12–14). However, most prior research has concentrated on sociolinguistic contrasts, such as native versus non-native accents or regional differences, rather than on the acoustic variability that occurs among talkers who share similar social or linguistic backgrounds. Understanding how listeners deal with this fine-grained variability is essential to determine to which extent natural speech variability influences higher-level language processing.

Word recognition is also guided by probabilistic expectations: as speech unfolds, lexical candidates that are compatible with the input encountered so far are incrementally pre-activated (4, 15–18). For instance, upon hearing the initial sounds /cap-/, listeners may pre-activate lexical items such as *captain* or *capital*, with activation levels modulated by their relative likelihood in context, even before the full word is available. When additional information becomes available, lexical candidates are dynamically updated: some options are strengthened, while others are suppressed or ruled out entirely, allowing the listener to converge rapidly on the intended word. Moreover, listeners generate probabilistic predictions about upcoming words (19), drawing on context to predict not only word meanings (20–23), but also their grammatical structure (24–28) and even word form (19, 29–32). This is supported by empirical evidence showing that contextual information has a graded influence on processing upcoming words (19, 33–38).

In the present study, we investigated to what extent listeners adapt to variability in native speech, and whether this in turn modulates higher-level predictions during naturalistic listening. EEG signals were recorded from Italian native speakers as they listened to naturalistic continuous speech. Participants listened to two narrative stories under two conditions: in the Single-talker condition (hence Single) the story was narrated by one talker with a canonical Italian pronunciation, allowing listeners to build voice familiarity; in the Multi-talker condition (hence Multi) the story was narrated by 9 different talkers, introducing substantial acoustic variability and reducing voice-specific adaptation, while maintaining canonical Italian pronunciation. This manipulation allowed us to investigate whether processing speech of various yet canonical pronounced voices modulates predictive processing during language comprehension.

To map brain responses to speech and linguistic features, we used the Temporal Response Function (TRF), a method that characterizes how specific stimulus features are reflected in the neural signal (19, 39–43). This approach offers an ecologically valid lens into how the human brain deals with acoustic variability in everyday listening. This study moves beyond controlled manipulations by investigating predictive processing in an underexplored naturalistic multitalker setting, while targeting various features of interest. The TRF approach enables us to isolate and model phonemic encoding, which represents how the brain dynamically represents talkers' individual phonemes as they unfold over time (43, 44). Simultaneously, we disentangle predictive mechanisms operating at both the lexical and sentential levels of language processing: during word recognition, as phonemic input accumulates (within-word), and at the sentential level, as lexical information unfolds incrementally across words. To capture these processes, we used cohort entropy to model lexical uncertainty as phonemes accumulate over time within each word (39), and semantic surprisal to model word-level prediction across sentence context (45). Cohort entropy indicates the degree of competition among lexical candidates compatible with the phoneme sequence presented so far, indexing the extent to which the brain is uncertain about the upcoming phonemes of the current word (i.e., within-word prediction). The TRF-modelled neural response (weights) of cohort entropy typically exhibits fronto-central activity emerging within 250 ms after word onset (39, 46). Semantic surprisal quantifies the degree of unexpectedness associated with a word given its preceding context, reflecting prediction at the sentence or discourse level (i.e., between-word prediction). This measure is typically estimated via computational models such as large language models (e.g., GPT) (37, 46). The TRF-modelled neural response of semantic surprisal reveals a complex with prominent centro-parietal negativity, akin to the classic N400 component associated with prediction error (46, 53–55). These models have shown that the brain dynamically encodes probabilistic lexical information as phonemes and words unfold, supporting hierarchical predictive processing across multiple linguistic levels, from words to syntax and phonemes (19).

In our study, we employed a multivariate TRF model encompassing phonemic categories (probing phonemic encoding), cohort entropy, and semantic surprisal, which was estimated using umBERTo, a transformer-based large language model for Italian.

To control for low-level acoustic differences, we included the acoustic spectrogram as an additional regressor. We built hypotheses of how acoustic variability affects speech processing for the three features of interest. In the Multi condition, listeners may need to accommodate acoustic variability across talkers and this can be reflected in differences in the strength of the neural response to phonemic categories. We assume this to be reflected on enhanced neural responses to phonemic categories in Multi compared to the Single condition. In addition, increased uncertainty in perceiving individual phonemes may prompt the brain to avoid committing to specific lexical candidates during word processing. This would result in stronger neural responses to cohort entropy in the Multi than in the Single condition. Finally, limited talker-specific information in the Multi condition may lead to reduced predictions about upcoming words, leading to reduced neural responses to semantic surprisal in the Multi than the Single condition. Conversely, if acoustic/phonemic uncertainty is fully resolved by lexical selection, we would not expect any effect of condition (Single vs. Multi) on semantic surprisal. Such results would indicate that both lexical selection based on the preceding phonemic context and lexical prediction based on the preceding sentential context are intertwined with bottom-up sensory perception.

2. Results

We used multivariate TRF models to estimate the time-resolved mapping between continuous linguistic features (e.g., phonemes, cohort entropy, and semantic surprisal) and EEG signals. The individual contribution of linguistic features was assessed by comparing model prediction correlations with and without that regressor (correlation gain), to evaluate whether the corresponding feature is reliably encoded in the EEG signals. To measure differences in neural responses across conditions, we applied cluster-based permutation tests (CBPT; 50) to TRF weights across electrodes and time points.

2.1. Phonemic processing

The individual contribution of this regressor yielded prediction correlations higher than zero ($t = 4.510$, $p < .001$), indicating that phonetic features were encoded in the EEG signals. Additionally, we assessed whether TRF responses to phonemic categories (derived via linear projection from phonemic features, see Method) differed between conditions. CBPT results indicated a significant widespread fronto-parietal cluster with stronger (negative) response in the Multi than the Single condition (cluster statistics = 1.955e+03, $p = .005$; Figure 1A-B).

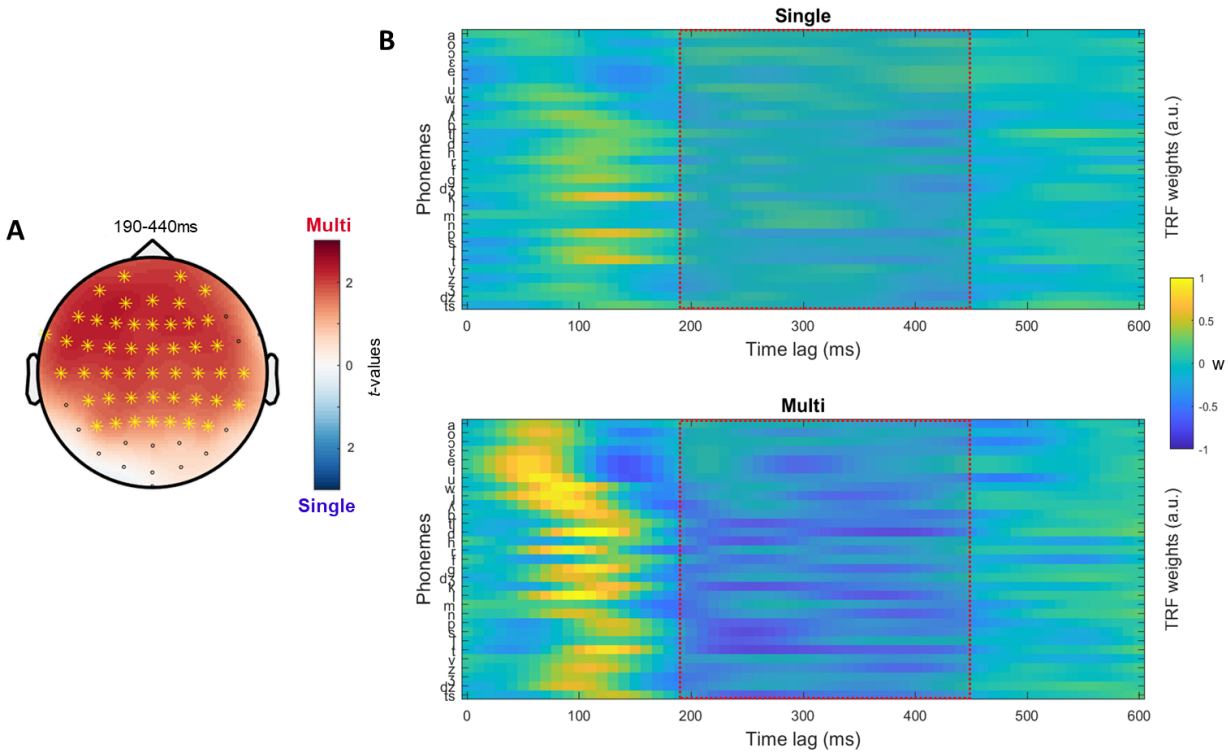

Figure 1. Participants show a stronger response to phonemic categories in the Multi than in the Single condition. **(A)** Significant channels (yellow asterisks) and time-window of Talker comparison resulting from the CBPT (difference between Single-Multi). Red color indicates greater absolute magnitude in the Multi condition. **(B)** Mean TRF weights of the phoneme regressor by Talker condition (Single above and Multi below) for Cz channel at

post-stimulus time latencies from 0 to 600 ms. Each row represents the temporal fluctuation of TRF response to individual phonemes.

2.2. Cohort Entropy

Cohort entropy regressor performance yielded prediction correlations that were higher than zero (t = 4.661, p < .001), indicating that cohort entropy was encoded in the EEG signals. CBPT on the cohort entropy weights highlighted a significant central cluster within the first 260 ms post-stimulus, reflecting a more positive early peak in Multi than Single (cluster statistics = 1.647e+03, p = .012; Figure 2A-B). To explore the neural substrates of the cohort entropy effect, we performed source estimation of TRF weights and statistical significance was assessed by CBPT. The analysis revealed significant clusters over bilateral temporal lobes, including ventral temporal regions, and right inferior frontal cortex peaking in the 70-150 ms time-window (p < .01; see Figure 2C), showing greater amplitude in Multi compared to Single.

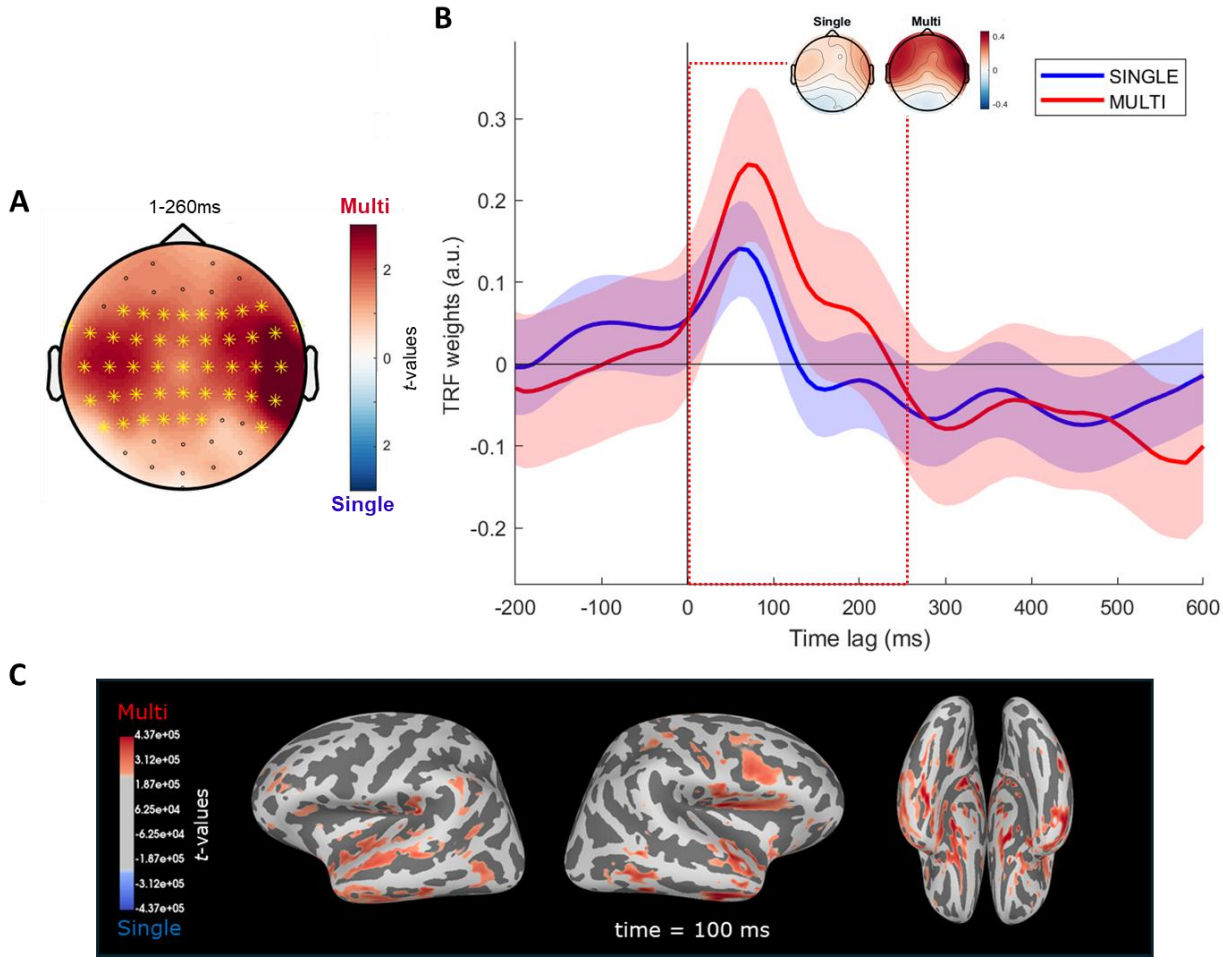

Figure 2. Participants are more sensitive to lexical uncertainty as phonemes unfold when listening to the Multi than the Single condition. **(A)** Significant channels (yellow asterisks) and time-window of Talker comparison resulting from the CBPT (difference between Single-Multi). Red color indicates greater absolute magnitude in the Multi condition. **(B)** Mean TRF weights of the cohort entropy regressor by Talker condition (Single and Multi) for FCz channel at post-stimulus time latencies from -200 to 600 ms. Shaded lines indicate SD across participants (on FCz). **(C)** Source estimation of the cohort entropy TRF weights. Red areas indicate significantly greater activation for the Multi vs Single condition.

2.3. Semantic Surprisal

We observed that semantic surprisal prediction correlation gain was greater than 0 (t = 3.282, p = .002), suggesting that semantic surprisal was encoded in the EEG

signals. CBPT on the semantic surprisal weights did not reveal any significant cluster of differences between the Single and Multi conditions (cluster statistics = 168.272, p = .336; Figure 3A-B).

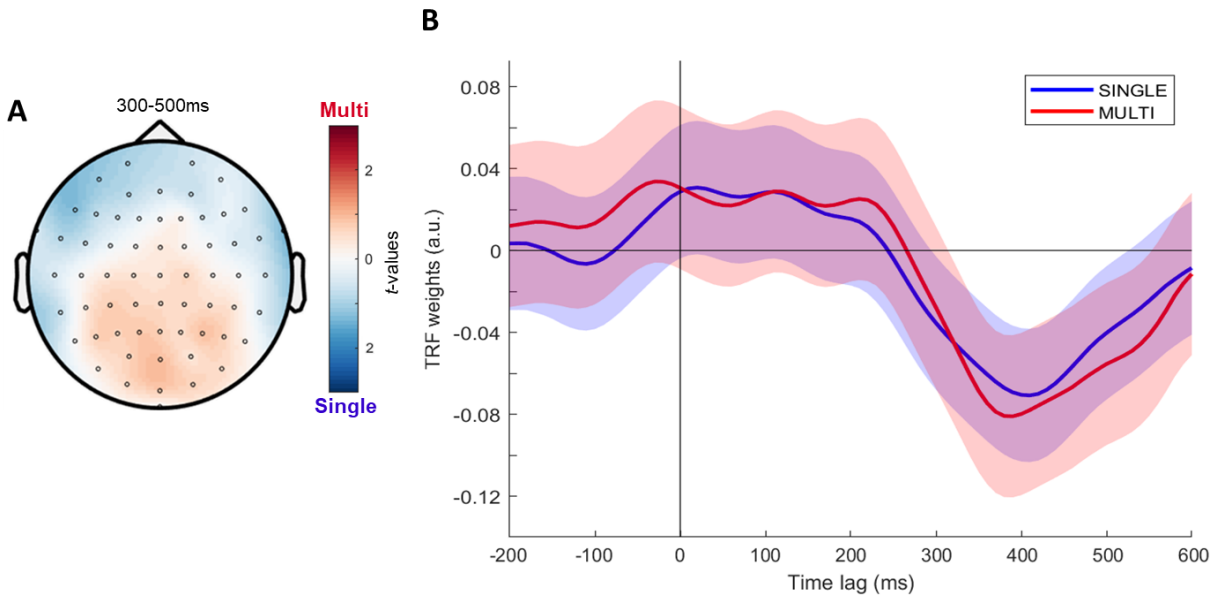

Figure 3. Lexical prediction is not affected by talkers' familiarity. **(A)** Weight topographic distribution in the canonical TRF-N400 time-window (difference between Single-Multi). Red color indicates greater magnitude in the Multi condition. **(B)** Mean TRF weights of the semantic surprisal regressor by Talker condition (Single and Multi) for Pz channel at post-stimulus time latencies from -200 to 600 ms. Shaded lines indicate SD across participants (on Pz).

3. Discussion

In this study, we investigated whether talker variability affects predictive processing during language comprehension using EEG. Participants listened to two stories, either narrated by one talker (Single condition) or by nine different talkers (Multi condition). We found that talker variability influences neural responses during the early stages of speech processing. Our results suggest that listeners generate probabilistic predictions at multiple levels of language processing, and that such predictions are sensitive to

variations in the physical properties of the speech signal, at least up to the stage of lexical access. The enhanced TRF response to phonemic categories in the Multi condition (190–440 ms) compared to the Single condition (Figure 1) suggests that increased talker variability imposes additional computational load during the perception of speech sounds. Building on evidence that phonemic encoding is shaped by how talkers adjust their speech across speech registers (Piazza et al., 2025), this is the first study with continuous natural speech to show that talker variability (with canonical pronunciation) directly modulates the early neural encoding of phonemic categories. These findings provide compelling neural evidence in support of recent theories of speech perception, which propose that listeners dynamically adapt to differences in the realization of speech sounds across talkers. According to the Ideal Adapter Framework (3), speech perception is best understood as an adaptive, probabilistic process in which listeners use talker-specific experiences to build and update internal distributional (statistical) models of acoustic cues. By tracking how these cues vary across different talkers, listeners are able to infer the intended linguistic categories despite variability in the input signal. In our study, when the entire story was narrated by a single talker, participants could accumulate a substantial amount of information about how the talker realized speech sounds. As a result, their internal distributional model likely converged more closely to the actual cue distribution of the talker, thereby facilitating speech perception. In contrast, exposure to multiple talkers reduced the amount of speaker-specific information available, making it harder to build reliable cue distributions. The need to rapidly map uncertain and variable phonemic realizations onto recognizable categories is likely reflected in the stronger neural response to phoneme categories in the Multi than the Single condition. Our results provide novel insights into speech perception, showing that listeners progressively accumulate talker-specific information about the realization of speech sounds, even when all talkers have a canonical native pronunciation. This indicates that perceptual adaptation is a continuous and dynamic process, in which listeners continuously leverage available talker-specific information to refine their internal representations of speech sounds.

At the phoneme-sequence level, the stronger early positive TRF response to cohort entropy in the Multi condition suggests increased uncertainty among lexical candidates compared to the Single condition (Figure 1A-B).This heightened uncertainty

was further reflected in stronger bilateral temporal lobe activation (Figure 2D). Importantly, such an increased uncertainty does not arise from entropy differences per se, as the same stimuli are used in both conditions. Rather, the cohort entropy effect originates from variability in the phonemic realization of the speech signal. During word recognition, listeners incrementally activate lexical candidates consistent with the input heard so far, with activation levels modulated by factors such as contextual likelihood (15, 51). If talker variability reduces the distinctiveness of phonemic cues, making phonemic categorization more challenging, listeners may delay committing to a specific lexical candidate, maintaining multiple interpretations equally active for a longer time.

Maintaining uncertainty about the current lexical item may serve as an adaptive strategy to facilitate word recognition as additional information becomes available over time. To ensure that the observed effect was not solely driven by word onsets, where cohort entropy is typically maximal, we reran the TRF model excluding cohort entropy values at word onset. Notably, the CBPT still revealed a significantly stronger cohort entropy response in the Multi condition compared to the Single condition ($p = .012$), confirming the robustness of this effect beyond words' initial phonemes (see Data and Code Availability). Our results are consistent with prior studies showing that maintaining uncertainty about a linguistic category can be beneficial for speech perception. For example, evidence suggests that later segments of a spoken stimulus might influence the interpretation of earlier segments (i.e., right-context effects in word recognition, 52–55). Despite the limited spatial resolution of EEG, we observed a stronger bilateral temporal response to cohort entropy in the Multi condition compared to the Single condition. Our exploratory source estimation results are consistent with previous research showing an involvement of the temporal lobe in graded lexical access (40, 56–58). Moreover, our results align with (59) dual-stream model of speech processing, which posits that the ventral stream, encompassing regions like the ITG (Inferior Temporal Gyrus) and MTG (Middle Temporal Gyrus), plays a central role in mapping sound to meaning, including processing of lexical cohorts.

Finally, although semantic surprisal was reliably encoded in the EEG signals, we found no evidence that talker variability influences neural responses to semantic surprisal

(Figure 3). This is unexpected, yet not entirely surprising, given that lexical access and prediction of upcoming words likely involve distinct cognitive processes. For instance, (15) distinguished between two types of prediction: *between-item prediction* and *within-item prediction*. Between-items prediction occurs when the activation of a lexical item spreads to similar or semantically related words. Instead, within-items prediction occurs when the incipit of a word pre-activates multiple possible lexical candidates, representing a form of pattern completion (60). In the context of talker variability, between-item predictions may remain stable as they rely on higher-level, context-driven processes that are relatively unaffected by acoustic variability. Conversely, resolving competing lexical candidates, which is closely tied to the processing of phonological cues, might be more susceptible to variability in the physical properties of speech.

In sum, our findings provide compelling evidence that listeners adaptively weight their prior knowledge and sensory input to maximize the likelihood of accurate speech comprehension. During word recognition, the brain delays committing to a specific lexical candidate when facing increased acoustic uncertainty. By doing so, it balances the need for optimizing information processing with the flexibility required for speech perception. This adaptive strategy is reflected in the strong engagement of temporal regions during lexical access. Furthermore, our results indicate that talker variability primarily affects early stages of language processing, whereas lexical-semantic predictions do not seem to be affected. Critically, we provide novel evidence that the brain dynamically adjusts the weighting of parallel lexical hypotheses based on the reliability of the acoustic signal, enabling greater flexibility in speech perception under uncertainty. In such conditions, the brain appears to engage in a less committed, more distributed predictive mode, maintaining a broader set of lexical candidates in parallel rather than prematurely narrowing down to a limited set of likely outcomes. This flexible predictive strategy supports robust comprehension in challenging environments. To conclude, we suggest that successful speech comprehension under uncertain conditions is supported by a system that flexibly integrates bottom-up input with probabilistic top-down predictions, sustaining lexical uncertainty when needed to allow robust, context-sensitive interpretation of spoken language.

5. Method

5.1. Participants

A total of 34 native speakers of Italian, aged between 18-35, were recruited to take part in the experiment (Female = 21). Of the original sample, 4 participants were excluded due to technical problems leaving the final cohort to 30 (M_{age} = 24.1 y.o., SD = 3.52). Participants were recruited from volunteers and students at the University of Padova, provided their informed consent before participating in the experiment, and were paid 25 euros for their participation. The research adhered to the principles outlined in the Declaration of Helsinki. The research protocol was approved by the Ethics Committee for Psychological Research of the University of Padova.

5.2. Material

Continuous Italian speech audios were employed in this study (see Data and Code Availability for stimuli and data). Two travel stories, each approximately 22 minutes long, were pre-recorded by male native Italian speakers (~45 minutes in total; Supporting Information for the transcription of the stories). The stories described trips to Ethiopia and North Korea, with content that was similarly novel to Italian participants. Each story was recorded twice: one for the Single-talker condition and one for the Multi-talker condition (Data and Code Availability for the audio recording files). In the Single-talker condition, the entire story was narrated by one speaker, while in the Multi-talker condition, each section of the story was narrated by a different speaker (9 speakers in total). Speakers were instructed to speak naturally, which led to variations in speech rate (thus duration) across sections. Since previous literature showed that speech rate affects neural encoding of speech (49, 61), we matched the duration of corresponding sections between the two conditions (e.g., Section 2 of the Ethiopia story in Single and Multi conditions). This was achieved by calculating the mean duration of each section pair and applying minimal time-stretching or shortening (~5%) to align the durations, while preserving pitch. The words and number of syllables were identical across conditions, ensuring that duration adjustments roughly corresponded to speech rate normalization. Both stories

were normalized to a consistent average level of 68 dB (see Data and Code Availability to check the audio sounds).

We ran a pilot study on Prolific (www.prolific.com) to collect subjective ratings of audio sounds qualities that potentially affect speech processing (without being the focus of our manipulation). A total of 212 native Italian speakers rated one 2-minute story section on four dimensions: clarity, expressiveness, pleasantness, and naturalness, using a 7-point scale (1 = Not at all, 7 = Completely/Extremely).

We employed the ζ-Overlapping test in R (62–64) with permutation-based inference to compare density distribution of the two conditions. This method is assumption-free and computes the exact probability of observing a test statistic as extreme or more extreme than the one observed by chance (64). Results showed no significant differences between conditions for any of the rated dimensions (all *p*-values > .05; see Supporting Information for visualizing the empirical distribution of the ratings on the evaluated dimensions).

Lastly, other acoustic features of the conditions audio sounds were expected to reflect differences in speakers' voices in a naturalistic communicative scenario. We deliberately did not control for other acoustic features (e.g., pitch, phonemic realization) that naturally varied across speakers and are part of our manipulation to investigate adaptation to various speakers' acoustics. To assess differences in the acoustic salience of speech sounds over time, we modeled non-linear fluctuations in the speech envelope (0-500 ms after each word onset) using Generalized Additive Mixed Models (GAMMs; 65, 66). The model did not reveal a significant difference in amplitude between the two conditions across the analyzed time window. However, the two conditions showed significantly different patterns of speech envelope fluctuations over time ($p = .003$), suggesting distinct temporal dynamics in speech acoustics (see Supplementary Materials). No analysis on the cohort entropy or semantic surprisal levels was performed as the two conditions shared identical words and phonemes (counterbalanced across stories and participants).

5.3. Procedure

All EEG data were collected in dimly lit and sound-proof booths. Stimuli were presented at a sampling rate of 44,100 Hz, monophonically, and at a comfortable volume throw loudspeakers. Participants were asked to listen attentively to two stories while EEG signals were recorded. They were asked to sit upright while looking at a fixation cross, which was presented on the centre of a computer screen right in front of them (at ~80 cm of distance from their eyes). During the experimental session, participants were presented with one story per condition, with counterbalanced order across participants. To avoid any effects derived from specific relations between stories and conditions (e.g., a certain story is more interesting), each story was presented in both Single and Multi conditions across participants. The continuous narration of each story was divided into nine consecutive shorter blocks of ~2'30'' minutes each. At the end of each block, participants were asked 5 true/false comprehension questions (45 questions per story). Experimental sessions lasted ~2 hours including preparation and testing.

5.4 Equipment

Electroencephalography (EEG) data were recorded using a 64 Ag-AgCl electrodes standard setting following the 10-20 international system (two actiCAP 64-channel systems, Brain Products GmbH, Germany) with hardware amplification (BrainAmp DC, Brain Products GmbH, Germany). Signals were bandpass filtered between 0.05 and 500 Hz, digitised using a sampling rate of 1000 Hz, and online referenced to the left mastoid. PsychoPy 2024 Software (67) was employed to present the stimuli. In addition, to ensure perfect synchronisation between the EEG recordings and the audio sounds, we implemented the Focusrite Scarlett 4i4 sound card (3rd gen) and the TriggerBox hardware (BrainProducts) to send triggers (see Supporting Information).

5.5. Analysis

5.5.1. Behavioural data

Behavioural data were analysed to identify and discard those participants with very low accuracy, who did not pay a sustained level of attention throughout the experiment (N=0). Each question could be scored either 0 to 1, which respectively represented wrong

and correct answers. All participants scored above chance (~80% on average, ranging between 67% and 93%). We did not further investigate behavioral data.

5.5.2. EEG pre-processing

EEG signal analyses were performed on MATLAB Software (MathWorks, 2024b), using custom scripts, Fieldtrip toolbox functions (50), EEGLAB (68), and CNSP resources (69). Offline, the data were re-referenced to the average of the two mastoids, resampled to 100 Hz, and band-pass filtered between 0.5 and 8 Hz with a Butterworth zero-phase filter (order 2+2). Noisy channels with variance 3 times larger or smaller than the channels median variance were recalculated by spline interpolating the surrounding clean channels in EEGLAB. We had planned to exclude participants from the analysis if they had more than 30% of rejected data or more than six contaminated electrodes in over 30% of sections. However, no participants or sections were discarded for these reasons. Out of 18 sections, one story section for ten participants and two sections for two participants were excluded due to recording failures or contaminated data.

5.5.3. EEG analysis

The cortical encoding of speech in the Single vs Multi conditions was estimated by measuring forward models (TRF), capturing the linear relationship between continuous stimulus features and the corresponding neural response. TRFs were calculated with the TRF-Toolbox (41), which implements a linear regression mapping multiple stimulus features to one EEG channel at a time. The regression included an L2 Tikhonov regularization with parameter λ, and was solved through the closed formula $\beta=(X^{\top}X+\lambda I)^{-1}X^{\top}y$, where β indicates the regression weights, X the stimulus features, I the identity matrix, and y an EEG channel. The regularization parameter was selected through an exhaustive search on a logarithmic parameter space from 10^{-2} to 10^{8}. This selection was carried out via cross-validation to maximize the EEG prediction correlation averaged across all channels, leading to TRF models that optimally generalize to unseen data. The interaction between stimulus and recorded brain responses is not instantaneous, as a sound stimulus at time t_0 can affect the brain signals for a certain time-window $[t_1, t_1+t_{\text{win}}]$, with $t_1 \geq 0$ and $t_{win} > 0$. The TRF takes this into account by including

multiple time-lags between stimulus and neural signal, providing model weights that can be interpreted in both space (scalp topographies) and time-lag (speech-EEG latencies).

Stimulus and EEG time-series were split into nine folds, each corresponding to one of the nine segments of the stories. Leave-one-out cross-validation procedure was employed to maximize the amount of data used for the model fit, at the cost of additional computational time compared with a single train-test split. Each iteration provided a prediction correlation coefficient (*r*-value) between each feature and the EEG response (per channel). The prediction correlation coefficient is the estimate of how strongly an EEG signal encodes a given set of stimulus features. An *r*-value of 1 would represent perfect correspondence between EEG signal and TRF features, whereas an r-value of 0 would indicate no correlation whatsoever.

5.5.4. Encoding of speech features (TRF regressors)

The neural encoding of linguistic features of interest was estimated by relating those features with the EEG signal with TRFs. The stimulus features considered here were phonemic categories, cohort entropy, and semantic surprisal. The time window used to fit the TRF model was [-200, 600] ms, based on previous research that found this time-lag window to be sufficient to capture the measurable EEG response to the speech envelope (45, 47). The TRF procedure offers two dependent measures that can be studied to infer the neural encoding of the stimulus. First, EEG prediction correlations are derived for each EEG channel, informing us on how informative a feature is for predicting EEG signals at a particular scalp location. Second, the TRF weights (i.e., linear regression weights, where a large weight, positive or negative, indicates a stimulus feature and time-lag of particular importance for predicting the EEG signal). Here, we fitted a multivariate TRF model including several regressors: acoustic spectrogram, phonemic categories, cohort entropy, semantic surprisal. By doing this, we aimed to investigate phonemic perception and high-level predictive processing (both cohort entropy and semantic surprisal) while accounting for acoustic differences across talker conditions (44, 49, 70, 71). EEG and (non-zero) stimulus-vectors were normalized prior to performing the TRF model (41, 69).

Acoustic spectrogram. The spectrogram regressor (Sgram) was implemented in 8 frequency bands ranging between 250 Hz and 8000 Hz. Those bands were defined based on the Greenwood equation that correlates the position of the hair cells in the inner ear to the frequencies that stimulate their corresponding auditory neurons (72).

Phonemic categories. Phonetic alignments of the speech material were obtained through forced alignment, initially performed automatically using WebMAUS (73) and subsequently verified manually with PRAAT (74). The alignments were stored as time-series data, with ones marking phoneme onsets and zeros elsewhere. This time-series representation was 17-dimensional, where the different dimensions corresponded to phonetic articulatory features. Phonetic features indicated whether each phoneme was voiced, voiceless (consonants), plosive, fricative, affricate, nasal, approximant, front, back, central, close, open (vowels), bilabial, labiodental, dento-alveolar, palatal, velar-glottal (Ladefoged, 2006). This way, each phoneme could be described as a particular linear combination of phonetic features. A linear transformation matrix was derived to describe the linear mapping from phonetic features to phonemes (phonemic categories), which we used to rotate TRF weights from phonetic features to the phoneme domain (43, 44, 49).

Cohort entropy. Cohort entropy quantifies the uncertainty in identifying a word as each successive phoneme unfolds. Building upon the cohort model of word recognition (75, 76), this measure reflects the degree of competition among potential lexical candidates activated by a certain phoneme sequence within words (77). As more phonemes are heard, the cohort narrows until the word is uniquely identified (78, 79). If many words are a suitable continuation of a given phoneme sequence, then cohort entropy is high (and vice versa). The cohort entropy is given as

$$-\sum_{w \epsilon C} P(w|C)\ log2\ P(w|C)$$

where C is the cohort of all words compatible with the sequence of phonemes presented so far in the word *w*. Spoken word frequencies and their phonemic compositions, were

extracted from the WikiPronunciationDict, a multilingual pronunciation dictionary that includes Italian, to obtain accurate phonetic transcriptions of words. In our study, we calculated cohort entropy using a tree-based approach (39, 71, 78, 79). By constructing a probabilistic tree structure from this corpus, we were able to model the likelihood of various phoneme sequences and compute entropy values at each phoneme position within words. Entropy values were then coded into a sparse time-vector, where non-zero values represent phoneme onsets until entropy dropped to zero (toward the world's uniqueness point).

Semantic surprisal. For investigating encoding of semantic surprisal, we first calculated its values as the negative logarithm of the probabilities extracted from the Italian transformer-based large language model UmBERTo. It is a model based on BERT and trained on large Italian Corpora (cased model trained on Commoncrawl ITA (~69GB) exploiting 110M parameters and 32k vocabulary size). UmBERTo calculated the probability of each word of the story chuck given its context. Surprisal was given as the negative logarithm of the lexical probabilities of the current word. Surprisal values were then coded into a sparse time-vector, where non-zero entries represent word onsets, and their values the surprise of that word based on the context.

5.6. Statistical analysis

5.6.1. TRF model performances

Before our analyses of interest, we conducted control tests to assess that the model regressors yielded EEG prediction correlations significantly greater than the null model. Specifically, we assessed whether our features of interest were encoded in the EEG signals to some extent. We conducted one-sample *t*-tests (one-tailed) against the null hypothesis (prediction correlations were not greater than 0) for all the features of the TRF model. We measured the unique contribution of phonetic features, cohort entropy, and semantic surprisal on model performance by using the leave-one-out approach (71, 80). This method involves systematically excluding each predictor from the full model to observe changes in predictive accuracy, thereby determining the impact of each regressor. Thus, we assessed whether the residual *r*-values (also called *gain*), averaged

across channels per participant, were greater than 0 (43, 49, 81). In analyzing the prediction correlation gain of semantic surprisal, most of the model data points do not capture the effect of interest, as the TRF-N400 complex typically peaks between 300-500 ms post-stimulus onset (47, 48). To address this temporal dispersion, which substantially dilutes the regressor's contribution, we employed the time-constrained EEG prediction correlation metric proposed by (70). This restricts model evaluation to the time window of interest, in this case 300-500 ms, thereby isolating the TRF model proportion where semantic surprisal is most likely encoded. Model comparison showed a significant contribution for all regressors (see Results).

5.6.3. TRF regression weights

In order to investigate differences in feature-specific encoding across talker conditions, we analyzed the temporal weights of each regressor in the final multivariate (full) model. For each feature, we submitted the TRF weights (including time and channels) to CBPT to identify statistically significant effects (in line with prior TRF studies; 49, 82). CBPT is a non-parametric approach that accounts for multiple comparisons across electrodes and time points (test statistics = paired t-test; cluster statistics = "maxsum"; method = "montecarlo"; correction for multiple comparisons = "cluster"). This data-driven approach allowed us to avoid predefining regions of interest or time windows. This is particularly important for features whose spatiotemporal characteristics might vary or exhibit delayed latency as a function of experimental manipulation, such as speaker variability (49, 83). For the analysis of phoneme representation, we averaged the TRF responses across all phonemic categories to derive a single spatio-temporal vector per participant (per condition). This collapsed response was then tested using CBPT, providing a conservative estimate of phoneme-related encoding over time.

As part of exploratory analysis of the neural substrates of responses to cohort entropy, we performed source localization of TRF weights using a minimum norm estimate approach implemented in MNE-Python (84). Individual forward models were computed using a 3-layer boundary element model (BEM) based on a standard anatomical template (fsaverage). The inverse solution was constrained to the cortical

surface and depth-weighted to reduce superficial bias. The resulting current estimates were normalized using dynamic statistical parametric mapping (dSPM) and statistical significance was assessed via CBPT.

Data Availability

Materials, data and analysis scripts will be made publicly available on OSF upon acceptance of the manuscript.

Acknowledgments

This research was supported by the Italian Ministry of University of Research with the PRIN grant (project code 2022FT8HNC, “Perceiving and predicting multisensory speech: a window on the interplay between sensory and motor processes and brain representations”) awarded to G.P. and F.P. This research was also supported by a PhD grant from the University of Padova (2022–2025) awarded to M.S.

Conflict of interest

The authors declare no competing interests.

CRediT author statement

GP: Conceptualization, Investigation, Data curation, Formal analysis, Visualization, Writing - Original Draft; MS: Conceptualization, Investigation, Data curation, Visualization, Writing - Original Draft; RG: Investigation, Writing - Review & Editing. MMW: Resources, Methodology; FP: Conceptualization, Supervision, Funding acquisition, Writing - Review & Editing.

References

1. M. H. Christiansen, N. Chater, The Now-or-Never bottleneck: A fundamental constraint on language. *Behav. Brain Sci.* **39**, e62–e62 (2016).

2. G. R. Kuperberg, T. F. Jaeger, What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* **31**, 32–59 (2016).

3. D. F. Kleinschmidt, T. F. Jaeger, Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* **122**, 148–203 (2015).

4. D. Norris, J. M. McQueen, Shortlist B: A Bayesian model of continuous speech recognition. *Psychol. Rev.* **115**, 357–395 (2008).

5. K. Weatherholtz, T. F. Jaeger, “Speech Perception and Generalization Across Talkers and Accents” in *Oxford Research Encyclopedia of Linguistics*, (Oxford University Press, 2016).

6. J. Hay, A. Nolan, K. Drager, From fush to feesh: Exemplar priming in speech perception. *Linguist. Rev.* **23** (2006).

7. N. Niedzielski, The Effect of Social Information on the Perception of Sociolinguistic Variables. *J. Lang. Soc. Psychol.* **18**, 62–85 (1999).

8. L. Staum Casasanto, Does social information influence sentence processing? in (2008).

9. K. B. McGowan, Social Expectation Improves Speech Perception in Noise. *Lang. Speech* **58**, 502–521 (2015).

10. M. Babel, G. McGuire, Listener Expectations and Gender Bias in Nonsibilant Fricative Perception. *Phonetica* **70**, 117–151 (2013).

11. E. A. Strand, Uncovering the Role of Gender Stereotypes in Speech Perception. *J.*

Lang. Soc. Psychol. **18**, 86–100 (1999).

12. K. Drager, Speaker Age and Vowel Perception. *Lang. Speech* **54**, 99–121 (2011).

13. J. Hay, P. Warren, K. Drager, Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* **34**, 458–484 (2006).

14. A. Walker, J. Hay, Congruence between 'word age' and 'voice age' facilitates lexical access. *Lab. Phonol.* **2** (2011).

15. F. Huettig, J. Audring, R. Jackendoff, A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition* **224**, 105050–105050 (2022).

16. W. Marslen-Wilson, "Access and integration: Projecting sound onto meaning" in *Lexical Representation and Process*, (The MIT Press, 1989), pp. 3–24.

17. W. D. Marslen-Wilson, Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102 (1987).

18. J. L. McClelland, J. L. Elman, The TRACE model of speech perception. *Cognit. Psychol.* **18**, 1–86 (1986).

19. M. Heilbron, K. Armeni, J.-M. Schoffelen, P. Hagoort, F. P. de Lange, A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci.* **119** (2022).

20. G. T. M. Altmann, Y. Kamide, Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* **73**, 247–264 (1999).

21. G. T. M. Altmann, Y. Kamide, The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *J. Mem. Lang.* **57**, 502–518 (2007).

22. C. G. Chambers, M. K. Tanenhaus, K. M. Eberhard, H. Filip, G. N. Carlson, Circumscribing Referential Domains during Real-Time Language Comprehension.

J. Mem. Lang. **47**, 30–49 (2002).

23. K. D. Federmeier, M. Kutas, A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *J. Mem. Lang.* **41**, 469–495 (1999).

24. M. W. Crocker, Wide-Coverage Probabilistic Sentence Processing. *J. Psycholinguist. Res.* **29**, 647–669 (2000).

25. R. Levy, Expectation-based syntactic comprehension. *Cognition* **106**, 1126–1177 (2008).

26. R. L. Lewis, Falsifying Serial and Parallel Parsing Models: Empirical Conundrums and An Overlooked Paradigm. *J. Psycholinguist. Res.* **29**, 241–248 (2000).

27. A. Staub, C. Clifton, Syntactic prediction in language comprehension: Evidence from either...or. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 425–436 (2006).

28. M. J. Traxler, Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends Cogn. Sci.* **18**, 605–611 (2014).

29. K. A. DeLong, T. P. Urbach, M. Kutas, Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* **8**, 1117–1121 (2005).

30. A. Ito, Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *J. Mem. Lang.* **139**, 104553–104553 (2024).

31. M. Sala, F. Vespignani, L. Casalino, F. Peressotti, I know how you'll say it: evidence of speaker-specific speech prediction. *Psychon. Bull. Rev.* (2024). https://doi.org/10.3758/s13423-024-02488-2.

32. M. Sala, F. Vespignani, S. Gastaldon, L. Casalino, F. Peressotti, In the words of others: ERP evidence of speaker-specific phonological prediction. [Preprint] (2025). Available at: http://biorxiv.org/lookup/doi/10.1101/2025.04.16.648895 [Accessed 8 May 2025].

33. S. L. Frank, Neural language model gradients predict event-related brain potentials in *Proceedings of the Society for Computation in Linguistics 2024*, R. Futrell, C. Mayer, N. Zaslavsky, Eds. (Association for Computational Linguistics, 2024), pp. 316–323.

34. S. L. Frank, L. J. Otten, G. Galli, G. Vigliocco, The ERP response to the amount of information conveyed by words in sentences. *Brain Lang.* **140**, 1–11 (2015).

35. J. Li, R. Futrell, A decomposition of surprisal tracks the N400 and P600 brain potentials. *Proc. Annu. Meet. Cogn. Sci. Soc.* **45** (2023).

36. S. A. McDonald, R. C. Shillcock, Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychol. Sci.* **14**, 648–652 (2003).

37. J. A. Michaelov, M. D. Bardolph, C. K. Van Petten, B. K. Bergen, S. Coulson, Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. *Neurobiol. Lang.* **5**, 107–135 (2024).

38. N. J. Smith, R. Levy, The effect of word predictability on reading time is logarithmic. *Cognition* **128**, 302–319 (2013).

39. C. Brodbeck, L. E. Hong, J. Z. Simon, Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Curr. Biol.* **28**, 3976-3983.e5 (2018).

40. C. Brodbeck, *et al.*, Parallel processing in speech perception with local and global representations of linguistic context. *eLife* **11**, e72056 (2022).

41. M. J. Crosse, G. M. Di Liberto, A. Bednar, E. C. Lalor, The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.* **10** (2016).

42. M. J. Crosse, *et al.*, Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Front. Neurosci.* **15** (2021).

43. G. M. Di Liberto, J. A. O'Sullivan, E. C. Lalor, Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Curr. Biol.* **25**, 2457–2465 (2015).

44. G. M. Di Liberto, *et al.*, Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage* **227**, 117586 (2021).

45. A. Klimovich-Gray, *et al.*, Increased top-down semantic processing in natural speech linked to better reading in dyslexia. *NeuroImage* **273**, 120072 (2023).

46. M. Gillis, J. Vanthornhout, J. Z. Simon, T. Francart, C. Brodbeck, Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics. *J. Neurosci.* **41**, 10316–10329 (2021).

47. M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, E. C. Lalor, Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Curr. Biol. CB* **28**, 803-809.e3 (2018).

48. M. P. Broderick, N. J. Zuk, A. J. Anderson, E. C. Lalor, More than words: Neurophysiological correlates of semantic dissimilarity depend on comprehension of the speech narrative. *Eur. J. Neurosci.* **56**, 5201–5214 (2022).

49. G. Piazza, *et al.*, Are you talking to me? How the choice of speech register impacts listeners' hierarchical encoding of speech. *Imaging Neurosci.* (2025).

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

51. D. Norris, J. M. McQueen, A. Cutler, Prediction, Bayesian inference and feedback in speech recognition. *Lang. Cogn. Neurosci.* **31(1)**, 4–18 (2015).

52. F. Grosjean, The recognition of words after their acoustic offset: Evidence and implications. *Percept. Psychophys.* **38**, 299–310 (1985).

53. E. G. Bard, R. C. Shillcock, G. T. M. Altmann, The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Percept. Psychophys.* **44**, 395–408 (1988).

54. C. M. Connine, D. G. Blasko, M. Hall, Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *J. Mem. Lang.* **30**, 234–250 (1991).

55. D. Dahan, The Time Course of Interpretation in Speech Comprehension. *Curr. Dir. Psychol. Sci.* **19**, 121–126 (2010).

56. N. Mesgarani, C. Cheung, K. Johnson, E. F. Chang, Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science* **343**, 1006–1010 (2014).

57. E. Formisano, F. De Martino, M. Bonte, R. Goebel, “Who” Is Saying “What”? Brain-Based Decoding of Human Voice and Speech. *Science* **322**, 970–973 (2008).

58. L. Gwilliams, T. Linzen, D. Poeppel, A. Marantz, In Spoken Word Recognition, the Future Predicts the Past. *J. Neurosci.* **38**, 7585–7599 (2018).

59. G. Hickok, D. Poeppel, The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).

60. J. B. Falandays, B. Nguyen, M. J. Spivey, Is prediction nothing more than multi-scale pattern completion of the future? *Brain Res.* **1768**, 147578–147578 (2021).

61. E. Verschueren, M. Gillis, L. Decruy, J. Vanthornhout, T. Francart, Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech Rate Manipulation Paradigm. *J. Neurosci.* **42**, 7442–7453 (2022).

62. M. Pastore, Overlapping: a R package for Estimating Overlapping in Empirical Distributions. *J. Open Source Softw.* **3**, 1023 (2018).

63. M. Pastore, A. Calcagnì, Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. *Front. Psychol.* **10** (2019).

64. A. Perugini, G. Calignano, M. Nucci, L. Finos, M. Pastore, How do my distributions differ? Significance testing for the Overlapping Index using Permutation Test. [Preprint] (2024). Available at: https://osf.io/8h4fe_v1 [Accessed 10 April 2025].

65. H. Baayen, S. Vasishth, R. Kliegl, D. Bates, The cave of shadows: Addressing the human factor with generalized additive mixed models. *J. Mem. Lang.* **94**, 206–234 (2017).

66. S. N. Wood, *Generalized Additive Models: An Introduction with R, Second Edition*, 2nd Ed. (Chapman and Hall/CRC, 2017).

67. J. Peirce, *et al.*, PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

68. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).

69. G. M. Di Liberto, *et al.*, A standardised open science framework for sharing and re-analysing neural data acquired to continuous stimuli. *ArXiv* arXiv:2309.07671v3 (2024).

70. A. Chalehchaleh, M. M. Winchester, G. M. Di Liberto, Robust assessment of the cortical encoding of word-level expectations using the temporal response function. *J. Neural Eng.* **22**, 016004 (2025).

71. M. Gillis, J. Vanthornhout, J. Z. Simon, T. Francart, C. Brodbeck, Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics. *J. Neurosci. Off. J. Soc. Neurosci.* **41**, 10316–10329 (2021).

72. D. D. Greenwood, A cochlear frequency-position function for several species--29 years later. *J. Acoust. Soc. Am.* **87**, 2592–2605 (1990).

73. T. Kisler, F. Schiel, H. Sloetjes, Signal processing via web services: the use case

WebMAUS. *Digital Humanities Conference 2012* (2012).

74. P. Boersma, D. Weenink, PRAAT, a system for doing phonetics by computer. *Glot Int.* **5**, 341–345 (2001).

75. M. Taft, G. Hambly, Exploring the Cohort Model of spoken word recognition. *Cognition* **22**, 259–282 (1986).

76. W. D. Marslen-Wilson, Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102 (1987).

77. P. Gaston, C. Brodbeck, C. Phillips, E. Lau, Auditory Word Comprehension Is Less Incremental in Isolated Words. *Neurobiol. Lang.* **4**, 29–52 (2023).

78. L. Gwilliams, J.-R. King, A. Marantz, D. Poeppel, Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nat. Commun.* **13**, 6606 (2022).

79. L. Gwilliams, M. H. Davis, "Extracting Language Content from Speech Sounds: The Information Theoretic Approach" in *Speech Perception*, L. L. Holt, J. E. Peelle, A. B. Coffin, A. N. Popper, R. R. Fay, Eds. (Springer International Publishing, 2022), pp. 113–139.

80. E. S. Teoh, F. Ahmed, E. C. Lalor, Attention Differentially Affects Acoustic and Phonetic Feature Encoding in a Multispeaker Environment. *J. Neurosci.* **42**, 682–691 (2022).

81. G. M. Di Liberto, M. J. Crosse, E. C. Lalor, Cortical Measures of Phoneme-Level Speech Encoding Correlate with the Perceived Clarity of Natural Speech. *eNeuro* **5**, ENEURO.0084-18.2018 (2018).

82. J. Pérez-Navarro, *et al.*, Early language experience modulates the tradeoff between acoustic-temporal and lexico-semantic cortical tracking of speech. *iScience* **27** (2024).

83. V. S. Raghavan, J. O'Sullivan, S. Bickel, A. D. Mehta, N. Mesgarani, Distinct neural encoding of glimpsed and masked speech in multitalker situations. *PLOS Biol.* **21**, e3002128 (2023).

84. A. Gramfort, *et al.*, MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).