Neuropsychologia 186 (2023) 108584

Contents lists available at ScienceDirect

Neuropsychologia

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

The effects of speech masking on neural tracking of acoustic and semantic features of natural speech

Sonia Yasmin [a,*], Vanessa C. Irsik [a], Ingrid S. Johnsrude [a,b], Björn Herrmann [c,d]

[a] Department of Psychology & the Brain and Mind Institute,The University of Western Ontario, London, ON, N6A 3K7, Canada
[b] School of Communication and Speech Disorders,The University of Western Ontario, London, ON, N6A 5B7, Canada
[c] Rotman Research Institute, Baycrest, M6A 2E1, Toronto, ON, Canada
[d] Department of Psychology,University of Toronto, M5S 1A1, Toronto, ON, Canada

ARTICLE INFO

Keywords:
Auditory perception
Acoustic processing
Semantic processing
Neural tracking
Speech processing

ABSTRACT

Listening environments contain background sounds that mask speech and lead to communication challenges. Sensitivity to slow acoustic fluctuations in speech can help segregate speech from background noise. Semantic context can also facilitate speech perception in noise, for example, by enabling prediction of upcoming words. However, not much is known about how different degrees of background masking affect the neural processing of acoustic and semantic features during naturalistic speech listening. In the current electroencephalography (EEG) study, participants listened to engaging, spoken stories masked at different levels of multi-talker babble to investigate how neural activity in response to acoustic and semantic features changes with acoustic challenges, and how such effects relate to speech intelligibility. The pattern of neural response amplitudes associated with both acoustic and semantic speech features across masking levels was U-shaped, such that amplitudes were largest for moderate masking levels. This U-shape may be due to increased attentional focus when speech comprehension is challenging, but manageable. The latency of the neural responses increased linearly with increasing background masking, and neural latency change associated with acoustic processing most closely mirrored the changes in speech intelligibility. Finally, tracking responses related to semantic dissimilarity remained robust until severe speech masking (−3 dB SNR). The current study reveals that neural responses to acoustic features are highly sensitive to background masking and decreasing speech intelligibility, whereas neural responses to semantic features are relatively robust, suggesting that individuals track the meaning of the story well even in moderate background sound.

1. Introduction

Many sound environments in everyday life contain background sounds, such as ambient music or speech, that can mask the target speech signal, resulting in communication challenges (Meyer et al., 2013; Song et al., 2011; Meyer et al., 2013). Segregation of speech from background sound is facilitated by a host of acoustic features such as onset times and harmonicity (Carroll et al., 2011; Flaherty et al., 2021; Kong et al., 2012; Darwin et al., 1995; Darwin, 2008). For example, speech signals fluctuate in amplitude at the semi-regular rate at which syllables, and words are uttered, typically below 10 Hz (Rosen, 1992). Because the amplitude fluctuations in speech and background sound typically differ, tracking amplitude fluctuations of speech provides a means to segregate speech from background sound. Semantic information also facilitates speech-in-noise perception. The semantic context of what has been heard can be used to predict upcoming words and, in turn, improve speech intelligibility in challenging listening conditions (Holt and Bent, 2017; Holliday et al., 2008; Shi, 2014; Zekveld et al., 2011; Davis and Johnsrude, 2007; Miller et al., 1951; Ganong, 1980; Pitt and Samuel, 1993; Norris et al., 2003). This is especially important for individuals with hearing impairments, who experience disproportionate challenges in settings with noisy backgrounds (Henry and Heinz, 2012; Monaghan et al., 2020; Bacon et al., 1998; Alain et al., 2014). Understanding how neural encoding of acoustic and semantic information occurs in different individuals and contexts is an important step towards clinical interventions for hearing loss, which are critically needed. The current study is concerned with how neural encoding of the acoustic amplitude fluctuations and the

* Corresponding author.
E-mail address: svarma6@uwo.ca (S. Yasmin).

https://doi.org/10.1016/j.neuropsychologia.2023.108584
Received 7 February 2023; Received in revised form 30 April 2023; Accepted 8 May 2023
Available online 9 May 2023
0028-3932/

semantic context of speech is affected by different degrees of background masking noise.

Much of the research into the neural processing of acoustic and semantic features of speech has relied on brief, disconnected sentences presented in a repetitive event-related design (Uhrig et al., 2020; Kasparian and Steinhauer, 2016; Handy, 2005; Luck, 2014; Salmelin, 2007; Picton, 2013; Pratarelli, 1995; Lovrich et al., 1988; Connolly et al., 1992; Connolly and Phillips, 1994). However, speech in everyday life is typically more continuous (Jefferson, 1978; Ochs and Capps, 1996; Pasupathi et al., 2002; Ochs and Capps, 1996), requires the integration of words into a larger semantic context and topical thread (Ehrlich and Rayner, 1981; Hale, 2001; Frank, 2013; Smith and Levy, 2013), and may be intrinsically motivating for a listener to comprehend. Listeners may thus engage differently with continuous speech compared to disconnected sentences, and the recruited neural mechanisms may thus also differ.

We have recently shown that listeners are absorbed by and enjoy spoken stories, even when they experience effort and miss occasional words as a result of moderate background masking (Herrmann and Johnsrude, 2020). Engagement measured neurally through across-participant synchronization of neural activity also appears to be little affected by moderate background masking (Irsik et al., 2022a). Moreover, older adults appear to benefit from speech glimpses in background noise for intelligibility more when listening to spoken stories than when listening to disconnected sentences (Irsik et al., 2022b). This suggests that something about stories – perhaps the degree to which they pique interest and motivate listening – results in qualitatively different listening behaviour in older people compared to disconnected sentences.

The neural processing of continuous speech is often measured by calculated a linear mapping between features of a continuous speech stimulus and the electro- or magnetoencephalographic (EEG/MEG) signals recorded while participants listen to the speech (Crosse et al., 2016; Das et al., 2020; Iotzov and Parra, 2019; Synigal et al., 2020). The result of such stimulus-to-neural-response mapping is the temporal response function (TRF; Crosse et al., 2016; Broderick et al., 2018; Crosse et al., 2021). TRF deflections can be interpreted similarly to components of the event-related potential for discrete speech tokens such as words (Broderick et al., 2018; Crosse and Lalor, 2014; Luck, 2012, 2014). The TRF approach has most frequently been used to investigate how acoustic properties of speech, such as the amplitude envelope, are encoded in the brain, and how this differs as a function of task demands (Das et al., 2020; Das et al., 2018; Verschueren et al., 2021; Fuglsang et al., 2017; Akram et al., 2016; Teoh and Lalor, 2020; Drennan and Lalor, 2019). For example, the magnitude of the TRF calculated for the amplitude envelope of speech is larger for speech that is attended compared to speech that is ignored in two-talker listening contexts (Verschueren et al., 2021; Fuglsang et al., 2017; Fiedler et al., 2019; Puvvada and Simon., 2017; Brodbeck et al., 2020).

Greater neural tracking of the acoustic speech envelope has also been associated with better speech comprehension (Verschueren et al., 2021; Decruy et al., 2019, 2020). However, the relationship between neural tracking of acoustic features and speech intelligibility may not be linear. When speech is parametrically degraded using noise-vocoding, envelope tracking exhibits a U-shape: TRF amplitude is greatest for moderate levels of degradation, and smaller both for intact and for highly degraded (1-channel vocoded) speech (Hauswald et al., 2022). Critically, noise-vocoding differs substantially from speech masked by babble noise. The latter resembles more closely situations that most individuals experience in everyday life, and that are reported by older individuals to be challenging and effortful (Gordon-Salant, 2006). Here, we investigate whether envelope tracking exhibits a similar inverted U-shape (to that observed by Hauswald et al., 2022) when speech is masked by a 12-talker background babble noise at different signal to noise ratios.

TRFs have also been used to investigate whether semantic features during continuous speech listening are encoded in the brain (Broderick et al., 2018; Gillis et al., 2021; Devaraju et al., 2021). In such studies, each word in a spoken story is represented by a high-dimensional numerical vector that captures semantic information. Words for which the corresponding vectors correlate highly are more semantically similar than words for which the vectors correlate less (Pennington et al., 2014; Mikolov et al., 2013). By calculating correlations for consecutive words within a sentence or a story, a dissimilarity score can be calculated for each word, reflecting the degree to which a word is incongruent with the preceding semantic context (Broderick et al., 2018, 2020, 2021). These dissimilarity scores are then used to calculate a "semantic" TRF, for which a larger TRF deflection reflects stronger responses to semantic word incongruency, given the speech context. Hence, the TRF deflection indicates the degree to which semantic context is encoded in the brain by exploiting the tracking of word incongruency (Gillis et al., 2021; Broderick et al., 2018, 2020, 2021).

Similar to the acoustic TRF (Hauswald et al., 2022; Das et al., 2020; Das et al., 2018; Verschueren et al., 2021; Fuglsang et al., 2017; Akram et al., 2016; Teoh and Lalor, 2020; Drennan and Lalor, 2019), the magnitude of the TRF calculated for the semantic dissimilarity is larger for attended compared to ignored speech (Broderick et al., 2018; Broderick et al., 2019). However, the degree to which neural encoding of semantic dissimilarity is affected by speech masking is not clear. In previous studies, speech was masked by a single talker at one signal-to-noise ratio (SNR), and the magnitude of the semantic TRF was reduced for the unattended speaker (Broderick et al., 2018; Brodbeck et al., 2020). However, single-talker masking differs substantially from multi-talker masking (Jones and Macken, 1995; Zaglauer et al., 2017; Macken et al., 2003). A single-talker masker may not overlap spectrally very much with the target (depending on the pitch difference between the target and masker voices), it will have a highly variable envelope that will differ from that of the target, and so physical interference between target and masker will be minimal. Nevertheless, a single talker masker is potentially confusable with the target, and might be distracting (Summers and Roberts., 2020) enhancing masking efficacy. Twelve-talker babble, in contrast, is more spectrally dense, and has a flatter envelope, and thus physically interferes with (i.e., energetically masks) a single-talker target more than a single-talker masker. Furthermore, 12-talker babble does not contain intelligible word-level information and results from research using single-talker masking will thus unlikely generalize to a situation in which multiple competing talkers are present. Indeed, recent studies have found that intelligible single-talker maskers reduce acoustic tracking of the target speech when compared to babble maskers (Song et al., 2019, 2020), but how semantic context, or dissimilarity, encoding is affected by multi-talker background noise at different SNRs is unknown. Given that individuals appear highly engaged in story listening even in the presence of moderate background noise (Herrmann and Johnsrude, 2020; Irsik et al., 2022a), we expect that semantic-context encoding also remains high at moderate background noise, and will only decrease for highly masked speech. Moreover, the relationship between semantic tracking and intelligibility has been scarcely explored. Speech intelligibility as measured through word reports declines with increasing masking level (Irsik et al., 2022a; Herrmann, 2023), but it will likely decrease more steeply than neural representations of semantic context.

Finally, neural tracking of continuous speech is often investigated using audiobook narrations (Broderick et al., 2018, 2020, 2021). Such materials are typically well articulated, sentences build systematically on each other, and there is a clear and well-understood grammatical framework in place (Thanh, 2015; Carter and Mncarthy, 1995). Speech in everyday life is subject to more disfluencies than audiobook narrations as speakers often use slang, filler-words, sentence fragments, corrections, unintentional pauses, and more flexible grammar (Bortfeld et al., 2001; Tree, 1995). It is possible that these discrepancies between naturalistic speech and audiobook narrations may affect intelligibility, effort, and/or neural processing (Arnold et al., 2003; Brennan and

Schober, 2001). Because we are interested primarily in naturalistic listening, we use engaging, spoken stories from the story-telling podcast The Moth (https://themoth.org; Regev et al., 2019; Simony et al., 2016; Irsik et al., 2022a) which may mirror speech in everyday life more closely than do audiobooks (Ochs and Capps, 1996; Ervin-Tripp and Küntay, 1997).

In the current study, we investigate how neural tracking of the acoustic amplitude fluctuations (envelopes) and semantic context of engaging, naturalistic stories are affected by background babble noise. We construct TRFs by linearly mapping acoustic and semantic features of speech onto corresponding EEG activity (Crosse et al., 2016, 2021). We further relate the acoustic and semantic TRFs to speech intelligibility assessed through word report (Irsik et al., 2022a, 2022b; Wendt et al., 2017; Winn and Teece, 2021). Intelligibility is distinct from comprehension in that intelligibility represents low level feature decoding of sounds, these sounds do not need to be words, or even meaningful (Hustad, 2008; Fontan et al., 2015). In contrast, comprehension is higher level and incorporates semantic information (Kutas and Federmeier, 2000). We specifically focus on word-report intelligibility to examine the extent to which neural encoding of semantic context is affected by the reduction in the number of individual words that can be understood.

2. Methods

We re-analyzed EEG and behavioural data from a previous study (Irsik et al., 2022a). With a few minor exceptions indicated explicitly below, the analyses, results, and conclusions are novel and do not overlap with those reported previously (Irsik et al., 2022a).

2.1. Participants

Thirty-nine EEG datasets (mean age of participants: 20.3 years; age-range: 18–32 years; 19 males 20 females) and 82 behavioural data sets (mean age of participants: 28.8 years; age-range: 18–36 years; 51 males 31 females) were available for analysis. All participants provided informed written consent and were without self-reported hearing loss, neurological issues, or psychiatric disorders. The study was conducted in accordance with the Declaration of Helsinki, the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (TCPS2-2014), and approved by the local Health Sciences Research Ethics Board of the University of Western Ontario (REB #112015; REB#112574).

2.2. Acoustic stimulation and procedure

Each of the 39 participants listened to four spoken stories from The Moth podcast (https://themoth.org): *Reach for the Stars One Small Step at a Time* (by Richard Garriott, ~13 min), *The Bounds of Comedy* (by Colm O'Regan, ~10 min), *Nacho Challenge* (by Omar Qureshi, ~11 min), and *Discussing Family Trees in School Can Be Dangerous* (by Paul Nurse, ~10 min). Twelve-talker babble noise, taken from the revised Speech in Noise (R–SPIN) test (Bilger, 1984), was added to the stories at five different signal-to-noise ratios (SNRs): clear, +12, +7, +2, −3 dB. SNR is defined as the difference in dB between the root-mean-square of the speech signal and the root-mean-square of the background babble. The SNR changed pseudo-randomly every 30–33 s to one of the five levels without repeating the same level twice in direct succession. When mixing stories with maskers to achieve a specific SNR, both the root-mean-square level of the story and the babble masker were adjusted in order to ensure sound level remained constant throughout each story, and was consistent across the stories. Three versions of SNR randomization were generated for each story and counterbalanced across participants such that no specific segment of the story was confounded by a specific SNR. Stories were played via headphones (Sennheiser HD 25 Light) in a single-walled sound-attenuating booth (Eckel Industries) and participants were instructed to listen carefully to each story. After each story, participants answered ten comprehension questions about the story to ensure they were paying attention.

Speech intelligibility for each story, measured as words reported from target phrases, across different signal to noise ratios, was assessed in a separate group of 82 participants using online platforms for experiment hosting (Pavlovia) and recruitment (MTurk, CloudResearch interface). Each participant listened to one of the four stories described above, and the SNR changed about every 30–33 s to one of five levels (clear, +12, +7, +2, −3 dB). For each story, 80 or 100 target phrases/sentences (4–8 words) were selected for intelligibility testing (4 phrase/sentences per 30–33s segment). The story paused occasionally (about every 16 s), and the participant was asked to type the last phrase/sentence (target) uttered before the story paused into a text box. Just before the target utterance was heard, a fixation cross on the screen changed colour to tell participants that they had to remember verbatim what they were about to hear, and then changed colour again for the duration of the phrase/sentence. That target phrase/sentence was then reported during the pause that immediately followed (for details see Irsik et al., 2022a). The story then resumed from the beginning of the target utterance. Intelligibility was calculated as the proportion of correctly reported words, separately for each SNR condition.

2.3. EEG recording and preprocessing

EEG was recorded from 64 active electrodes (Ag/AgCl) placed on the scalp using an electrode cap according to the 10/20 system (Biosemi ActiveTwo system) and both mastoids. A feedback loop between the common mode sense (CMS) active electrode and a driven passive electrode (see www.biosemi.com/faq/cms&drl.htm) was used as a reference for all other electrodes. EEG was recorded at a sampling frequency of 1024 Hz (208-Hz low-pass filter).

The data were pre-processed offline using custom MATLAB scripts and the Fieldtrip toolbox (Oostenveld et al., 2011). Data were re-referenced by subtracting the average across both mastoids from each channel. Line noise was suppressed using a 60-Hz notch filter. The data were high-pass filtered (0.5 Hz, 3429 points, Hann window) and low-pass filtered (22 Hz, 211 points, Kaiser window). Continuous EEG data were segmented into separate time series time-locked to story onset and downsampled to 256 Hz. Independent components analysis was used to remove signal components reflecting blinks, eye movement, and muscle activity (Makeig et al., 1995). Additional artifacts were removed after the independent components analysis by setting the voltage for segments in which the EEG amplitude varied more than 80 μV within a 0.2-s period in any channel to 0 μV (Cohen and Parra, 2016). As a last step prior to TRF analyses, data were low-pass filtered at 10 Hz (141 points, Kaiser window), because neural signals in the low-frequency range are most sensitive to acoustic and semantic features (Zuk et al., 2021; Di Liberto et al., 2015).

2.4. Speech transcription and identification of word-onset times

Transcription for stories were done manually for each story. Non-words and incomprehensible mumbles were ignored for the analysis of EEG. The onset time for each word in each story was obtained using the Clarin's forced alignment software (Yuan and Liberman, 2008). Onset times were manually verified, and incorrect estimations were manually corrected.

2.5. Acoustic and semantic temporal response functions

We used a forward model based on the linear temporal response function (TRF; Crosse et al., 2016; Crosse et al., 2021) to separately estimate the relationship between features of the auditory stimulus and EEG activity (see Fig. 1). The TRF model uses linear regression with ridge regularization (Crosse et al., 2016; Crosse et al., 2021; Hoerl and

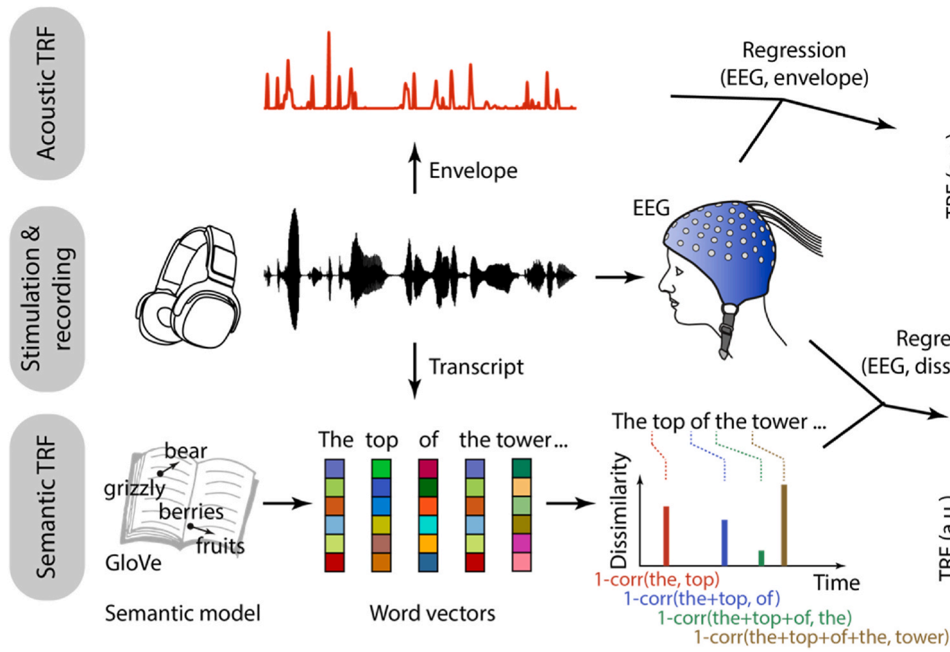

Fig. 1. Schematic of the procedure for obtaining acoustic and semantic temporal response functions (TRFs). Middle row: Schematic depiction of sound stimulation, acoustic waveform, and EEG recording. **Top row:** Schematic display of the calculation of the onset-amplitude envelope from the acoustic waveform. The amplitude envelope is regressed against the EEG data to obtain an acoustic TRF. **Bottom row:** Schematic depiction of the calculation of the semantic TRF. Global Vectors for Word Representation (GloVe) was used to obtain numerical word vectors, representing word meaning, for each word of each story. Colors in the vectors schematically represent different magnitudes. A semantic dissimilarity value was calculated for each word as 1 minus the correlation between the current word's vector and the averaged vectors across all preceding words of a sentence. A dissimilarity regressor was created by placing each word's dissimilarity value at its respective word-onset time (while values at all other time points were zero). The dissimilarity vector is regressed against the EEG data to obtain a semantic TRF. The acoustic and semantic TRFs displayed on the right reflect the mean TRFs across signal-to-noise ratio levels in the current study.

Kennard, 1970a; Hoerl and Kennard, 1970b) to map a continuous stimulus feature onto continuous neural activity (Crosse et al., 2016, 2021). Based on previous work, the ridge regularization parameter λ, which prevents overfitting, was set to 10 (Fiedler et al., 2019; Fiedler et al., 2017).

The current TRF analyses focused on two representations of the auditory stimulus: the acoustic (amplitude) envelope and semantic dissimilarity. For the acoustic envelope, we calculated the cochleogram for the acoustic waveform of each story using Lyon's Passive Ear model (Slaney, 1988) as implemented in the Auditory Toolbox Version 2 (Slaney, 1998). The cochleogram reflects the auditory nerve responses over time for different frequencies along the cochlear. We then averaged across frequencies (i.e., auditory filters) of the cochleogram to obtain the cochlear amplitude envelope, separately for each story. The analytic Hilbert transform of the acoustic envelope was calculated. We low-pass filtered the envelope using a 40-Hz filter (Butterworth filter), calculated the first derivative, and set all negative values to zero in order to obtain the amplitude-onset envelope (Fiedler et al., 2017; Hertrich et al., 2012). This amplitude-onset envelope was used as a regressor for the TRF analysis (Fig. 1).

Semantic dissimilarity was calculated as follows (Broderick et al., 2018; Broderick et al., 2018). We used Global Vectors for Word Representation (GloVe) to obtain a semantic representation for each word in form of numerical vectors (i.e., word embeddings; 300 dimensions; Pennington et al., 2014; https://nlp.stanford.edu/projects/glove/). GloVe is an unsupervised learning model that maps words into vector space based on their semantic relationships. The numerical vectors of words that are semantically more similar are more correlated (e.g., frog vs toad) compared to the vectors of words that are semantically less similar (e.g., frog vs shoe). The GloVe corpus consists of 400,000 vocabulary entries and their corresponding numerical vectors. For each word of the story transcripts, we obtained the corresponding word vector from GloVe, if it existed in the corpus. On average across the four stories, 11% of words were not available in the GloVe corpus and they were thus not considered for calculating semantic dissimilarity for the TRF analysis. Using the word vectors, a semantic dissimilarity value was calculated for each word of each story based on the local sentence context (see Broderick et al., 2018). Specifically, the Pearson correlation between the vector of the current word and the averaged vectors across all preceding words of the sentence was calculated. Each correlation value was subtracted from 1 to calculate the dissimilarity value (Fig. 1). A regressor for the TRF analysis was then created by placing each word's dissimilarity value at its respective word-onset time (while values at all other time points were zero). This regressor was created at the sampling frequency of the EEG data.

Because the dissimilarity regressor contains impulses at word onsets (with values being otherwise zero), it is sensitive to brain responses associated with the acoustic onset of words. In order to mitigate the influence of acoustic properties on the semantic TRF, we also calculated a 'static' TRF, where the regressor is calculated using the median dissimilarity value across words for all word onsets (all other samples remain zero). Hence, this regressor also contained impulses at word onsets, but the impulses were all of similar magnitude and no semantic dissimilarity variations were represented.

Separately for each participant, EEG electrode, and ~30 s data segment corresponding to different SNR levels within stories, a TRF was calculated using time windows of −0.3 s–0.7 s between the input time series of stimulus features (acoustic, semantic) and the corresponding EEG time courses, measured from word onset. TRFs were averaged across the data segments, separately for each SNR level. To obtain the final semantic TRF, we subtracted the 'static' TRF from the TRF derived using the dissimilarity vector. The result of these TRF calculations was one acoustic TRF and one semantic TRF for each SNR level, EEG channel, and participant.

2.6. Analysis of the relation between SNR levels and TRF amplitude and latency

For the analysis of amplitude and latency of specific deflections in the TRF, we averaged the TRFs across a fronto-centro-parietal electrode cluster (FC1, FC2, FCz, FC3, FC4, C1, C2, Cz, C3, C4, CP1, CP2, CPz, CP3, CP4, P1, P2, Pz, P3, P4) known to be sensitive to responses elicited by acoustic and semantic manipulations (Broderick et al., 2018; Connolly et al., 1992; Connolly et al., 1992; Martin et al., 1999; Finke et al., 2016; Martin et al., 1999; Martin and Stapells, 2005). We used custom MATLAB scripts to automatically identify response peaks within selected time ranges. For the acoustic TRF, we estimated the peak latency for the negative deflection within 100–250 ms for each participant and SNR level. We call this negative deflection the "acoustic tracking response". Although there is obvious resemblance to the typical N1/N100 component of event-related potentials (Crosse and Lalor, 2014), we want to avoid the assumption that what we observe here is indeed the N1/N100. The amplitude for the acoustic tracking response was calculated as the mean amplitude across 10 ms centered on a participant's individual peak latency. Our investigations for the acoustic TRF are restricted to the negativity at 100–250 ms, as visual inspection of the time course in Fig. 2a demonstrates this peak to be most susceptible to SNR-related changes.

For the semantic TRF, we estimated the peak latency for the negative deflection within 300–450 ms for each participant and SNR level. We call this negative deflection the "semantic tracking response". This deflection resembles the typical N400 component of event-related potentials, which has been associated with semantic incongruency (Kutas and Federmeier, 2011; Broderick et al., 2018), but, again, we do not assume that what we observe here is indeed the N400. The amplitude for the semantic tracking response was calculated as the mean across 100 ms centered on a participant's individual peak latency.

We evaluate the degree to which acoustic and semantic tracking changes linearly or quadratically over SNRs. To this end, a quadratic function was fitted separately to the latency and amplitude data as a function of SNR levels (coded: [−2 -1 0 1 2]), separately for each participant. Quadratic fits, appropriate to test whether the data conform to a U-shape, as predicted, were calculated separately for the acoustic TRF (acoustic tracking response) and the semantic TRF (semantic tracking response), and separately for both amplitude and latency data. The resulting linear and quadratic coefficients were tested against zero using a one-sample *t*-test to identify significant nonzero linear and quadratic trends of TRF amplitude/latency as a function of SNR.

We also conducted one-way repeated measures ANOVAs (rmANOVAs) to quantify effects of SNR on acoustic and semantic TRF amplitudes and latencies. We performed posthoc pairwise comparisons using independent samples t-tests, with false discovery rate correction (Benjamini and Hochberg, 1995), between neighboring SNR levels to evaluate differences. FDR corrected p-values are referred to as p_{FDR}.

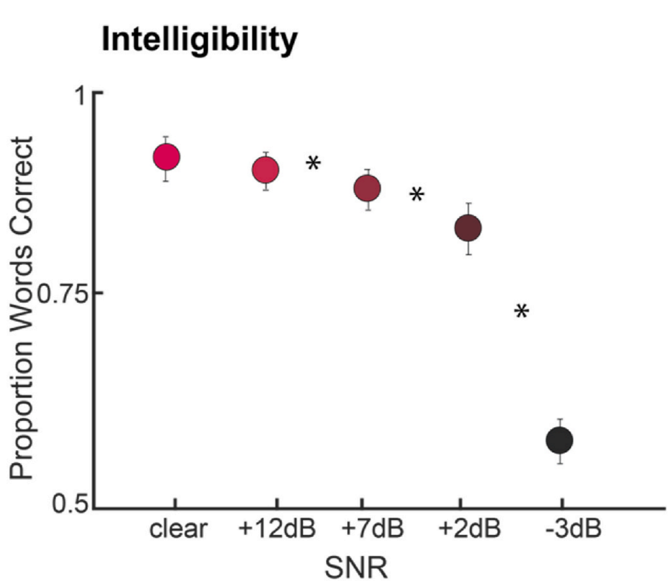

Fig. 2. Effects of SNR on Intelligibility. Mean proportion of correctly reported words plotted as a function of SNR (clear, +12, +7, +2, −3 dB SNR). Asterisks indicate that the intelligibility of the two flanking SNRs differ significantly. Error bars reflect the standard error of the mean. *p < 0.05. For more detailed information see Irsik et al. (2022a).

2.7. Relationship between acoustic and semantic TRFs, and speech intelligibility

Amplitudes and latencies of acoustic and semantic TRFs as well as speech intelligibility (from online testing; Fig. 2) have different units and magnitudes. In order to compare them directly, we calculated z-scores for each participant. That is, separately for each individual and dependent measure, we took the value at each SNR, subtracted the average across the five SNRs, and then divided by the standard deviation of that measure across SNRs. Z-score normalized acoustic TRF amplitude and latency, and z-normalized semantic TRF amplitude and latency were also sign inverted by multiplying the data by −1, to ensure that larger values indicate larger amplitudes and shorter latencies, enabling comparison with speech intelligibility data (for which a larger value means better intelligibility). In order to compare these responses, we again fit quadratic functions separately to the acoustic TRF amplitude, semantic TRF amplitude, acoustic TRF latency, and semantic TRF latency, and to the speech intelligibility data, across SNRs. We used t-tests on the resulting coefficients to examine whether changes across SNR in speech intelligibility were more similar to the acoustic TRF, the semantic TRF, or neither.

The neural tracking data are captured well by the quadratic function (including linear and quadratic coefficients; see below). Intelligibility data are often fitted with a sigmoidal function (Irsik et al., 2022b; Herrmann, 2023), but the quadratic fit captured the relevant variance as well and enables us to compare the linear and quadratic coefficients between neural and behavioural data. The neural data would be very poorly fit with a sigmoidal function. As such, our approach of comparing linear and quadratic coefficients reflects a compromise that enables use to quantify differences among acoustic, semantic, and intelligibility data. Specifically, our aim is to quantify differences among these data, as opposed to quantifying the best mathematical model for these responses. We are interested, primarily, in the distinction between linear versus nonlinear trends in these tracking data, as nonlinearities indicate that these tracking responses do not slavishly follow SNR. Rather, there are cognitive mechanisms that modulate the relationship between neural tracking and SNR.

2.8. Effect size

Effect sizes are reported as partial eta squared for ANOVAs ($\eta^2_{p;}$ Kennedy, 1970) and Cohen's d for *t*-tests (d; Cohen, 1992).

3. Results

3.1. Amplitude and latency of acoustic TRFs are modulated by the degree of background masking

We found that the amplitude of the acoustic tracking response was quadratically modulated by SNR ($t_{38} = 9.225$, $p = 3.06 \times 10^{-11}$, $d = 1.477$). There was no significant linear modulation of acoustic tracking response amplitude by SNR ($t_{38} = -1.556$, $p = 0.1281$, $d = 0.249$). To further explore the quadratic effect, we conducted a rmANOVA ($F_{4,152} = 19.537$, $p = 5.44 \times 10^{-13}$, $\eta^2_p = 0.3396$), followed by pair-wise comparisons between SNR levels. After false discovery rate (FDR) correction, we observed significant differences for all neighboring SNR levels except between the +12 dB to +7 dB conditions (clear smaller than +12 dB: $t_{38} = 6.194$, $p_{FDR} = 3.1 \times 10^{-7}$, $d = 1.099$; +12 vs + 7 dB: $t_{38} = 1.866$, $p_{FDR} = 0.069$, $d = 0.333$; +7 greater than +2 dB: $t_{38} = 3.262$, $p_{FDR} = 0.0023$, $d = 0.607$; +2 greater than −3 dB: $t_{38} = 2.423$, $p_{FDR} = 0.0203$, $d = 0.460$). These results indicate a U-shape: the acoustic tracking response amplitude increased for minimal to moderate background noise relative to clear speech, and then decreased again for speech that is highly masked (Fig. 3B).

Acoustic tracking response latency increased linearly with decreasing SNR ($t_{38} = 10.979$, $p = 2.39 \times 10^{-13}$, $d = 1.758$). There was also a quadratic relationship between SNR and acoustic tracking response latency ($t_{38} = 2.452$, $p = 0.0189$, $d = 0.393$). We followed up on the linear and quadratic effects with a rmANOVA ($F_{4,152} = 43.61$, $p = 2.5 \times 10^{-10}$, $\eta^2_{p=}0.534$) and pair-wise comparisons between neighboring SNR levels. After FDR correction, all neighboring SNR levels differed significantly except the clear and +12 dB conditions (clear vs +12 dB: $t_{38} = -1.254$, $p_{FDR} = 0.218$, $d = 0.209$; +12 vs + 7 dB: $t_{38} = 3.880$, $p_{FDR} = 0.0004$, $d = 0.667$; +7 vs + 2 dB: $t_{38} = 3.355$, $p_{FDR} = 0.0018$, $d = 0.493$; +2 vs – 3 dB: $t_{38} = 4.183$, $p_{FDR} = 0.0002$, $d = 0.802$).

3.2. Amplitude and latency of semantic TRFs are modulated by the degree of background masking

We evaluated the relationship between the degree of background masking of speech and the neural responses to semantic encoding of the story (i.e., semantic dissimilarity; Fig. 4). We observed that the semantic tracking response amplitude was quadratically modulated by SNR ($t_{38} = 2.731$, $p = 0.0095$, $d = 0.437$), whereas the linear modulation was not significant ($t_{38} = 0.872$, $p = 0.389$, $d = 0.1397$). We followed up on this result using a rmANOVA ($F_{4,152} = 2.706$, $p = 0.032$, $\eta^2_p = 0.0665$) and pair-wise comparisons between neighboring SNR levels. After FDR correction, the semantic tracking response amplitude was lower at the least favourable SNR condition compared to its neighbour (−3 dB and +2 dB SNR; $t_{38} = 3.399$, $p_{FDR} = 0.0016$, $d = 0.542$), whereas tracking did not differ between any other pairs (for all $p_{FDR} > 0.05$).

As for the acoustic tracking response, the semantic tracking response latency increased linearly with decreasing SNR ($t_{38} = 2.834$, $p = 0.0073$, $d = 0.454$), and no quadratic trend was found ($t_{38} = 1.211$, $p = 0.233$, $d = 0.194$). The rmANOVA revealed a significant effect of SNR ($F_{4,152} = 3.043$, $p = 0.019$, $\eta^2_p = 0.074$), although no two SNR levels differed after FDR correction (neighboring and not).

3.3. Comparison of semantic and acoustic TRFs and their relation to speech intelligibility

In order to investigate differences in how SNR affected neural acoustic and semantic tracking, and to examine whether the change in intelligibility over SNR related to the acoustically driven responses or the semantically driven responses, quadratic functions were fit to z-scored data and the resulting linear and quadratic coefficients were compared between measures. We first contrasted coefficients between the acoustic and the semantic tracking responses, before comparing each of these to coefficients from fits to intelligibility data.

Fig. 5A shows the amplitude of the TRFs and speech intelligibility. The amplitude of the acoustic tracking response showed a stronger linear relationship with SNR (positive relationship) than the amplitude of the semantic tracking response (negative relationship) ($t_{38} = 2.723$, $p = 0.0096$, $d = 0.610$). The acoustic tracking response amplitude was also more quadratically related to SNR than the semantic tracking response amplitude ($t_{38} = 4.214$, $p = 1.5 \times 10^{-4}$, $d = 0.962$). This is consistent with the observation that the semantic tracking response amplitude only dropped at the lowest SNR level (−3 dB SNR; Fig. 3B). Fig. 5B shows the latency of the TRFs and speech intelligibility. The acoustic tracking response latency was more strongly linearly related to SNR than the semantic tracking response latency ($t_{38} = 4.764$, $p = 2.77 \times 10^{-5}$, $d = 1.0821$), showing that the acoustic tracking response latency increased more with decreasing SNR than the semantic tracking response latency (see also Figs. 3 and 4). There was no difference between acoustic and semantic tracking in for the quadratic relation to SNR ($t_{38} = -0.193$, $p = 0.848$, $d = 0.464$).

In order to compare how SNR affects speech intelligibility and neural responses, we compared coefficients obtained from linear and quadratic fits (Fig. 5). We found that speech intelligibility declined more linearly with decreasing SNR than did either the acoustic tracking response amplitude ($t_{38} = 14.052$, $p_{FDR} = 4.89 \times 10^{-27}$, $d = 2.733$) or the semantic tracking response amplitude ($t_{38} = 8.845$, $p_{FDR} = 1 \times 10^{-14}$, $d = 1.721$). Speech intelligibility was also more quadratically modulated by SNR than the semantic tracking response amplitude ($t_{38} = -3.433$, $p_{FDR} = 8.2 \times 10^{-4}$, $d = 0.668$), but less quadratically modulated than the acoustic tracking response amplitude ($t_{38} = -3.823$, $p_{FDR} = 2.1 \times 10^{-4}$, $d = 0.744$). This is probably because the acoustic TRF magnitude increased significantly for intermediate SNRs, whereas intelligibility did not, and intelligibility appears to drop more precipitously at the lowest SNR (−3 dB) than does semantic tracking. These results indicate that the relationship between SNR and speech intelligibility is not entirely reflected either in the relationship between SNR and acoustically driven TRF amplitudes, or in the relationship between SNR and semantically driven TRF amplitudes.

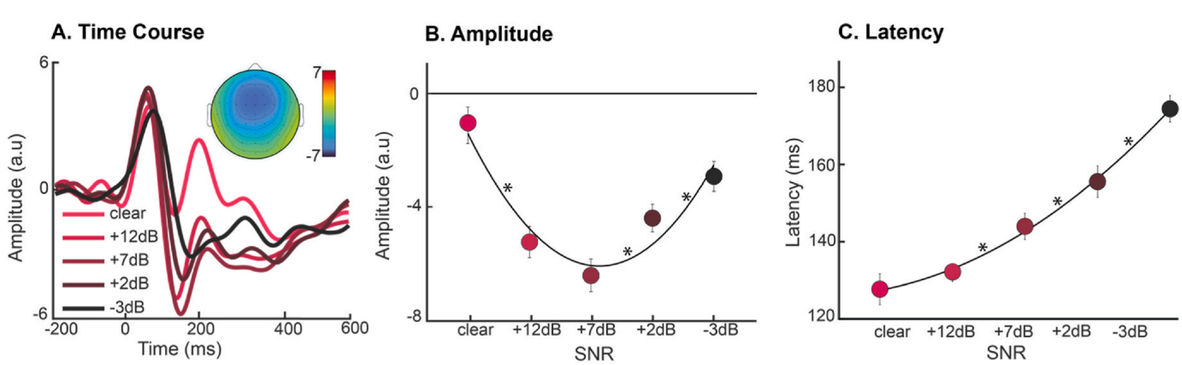

Fig. 3. Effects of SNR on acoustic TRFs. A. TRF time courses (averaged across fronto-central-parietal electrode cluster) for each SNR condition and scalp topography for the acoustic tracking response (negative deflection at around 150 ms). **B.** The mean acoustic tracking response amplitude across participants, displayed for each SNR condition. Significant differences in response magnitude exist between clear and +12 dB SNR, +7 dB and +2 dB SNR, and +2 dB and −3 dB SNR. **C.** The mean acoustic tracking response latency across participants, displayed for each SNR. Significant differences in response latency exist between +12 dB and +7 dB, +7 dB and +2 dB, and +2 dB and −3 dB. The black lines in panels B and C indicate the best fitting line from a quadratic fit. Error bars reflect the standard error of the mean. *neighboring SNRs differ at p < 0.05.

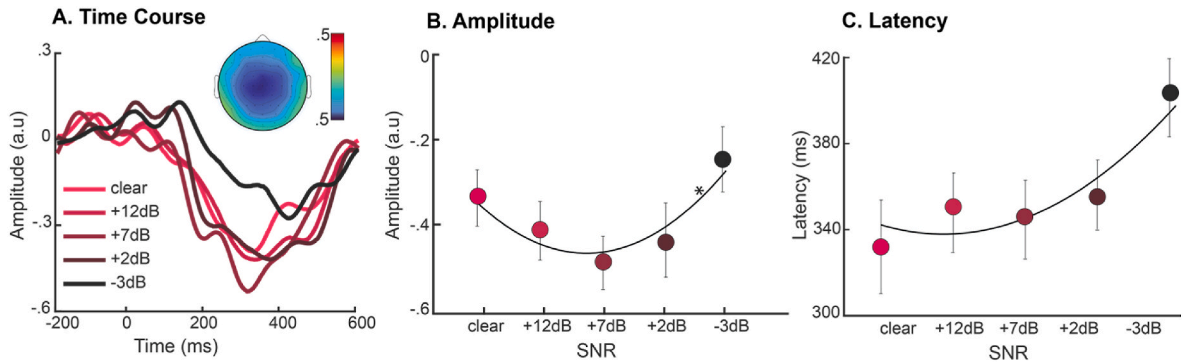

Fig. 4. Effects of SNR on semantic TRFs. TRF time courses (averaged across fronto-central-parietal electrode cluster) for each SNR condition and scalp topography for the semantic tracking response (negative deflection at around 350 ms). **B.** The mean semantic tracking response amplitude across participants, displayed for each SNR condition. Significant differences in response amplitude exist between the +2 dB and −3 dB conditions, only. **C.** The mean semantic tracking response latency across participants, displayed for each SNR. No significant differences in response latency exist between neighboring conditions. The black lines in panels B and C indicate the best fitting line from a quadratic fit. Error bars reflect the standard error of the mean. *$p < 0.05$.

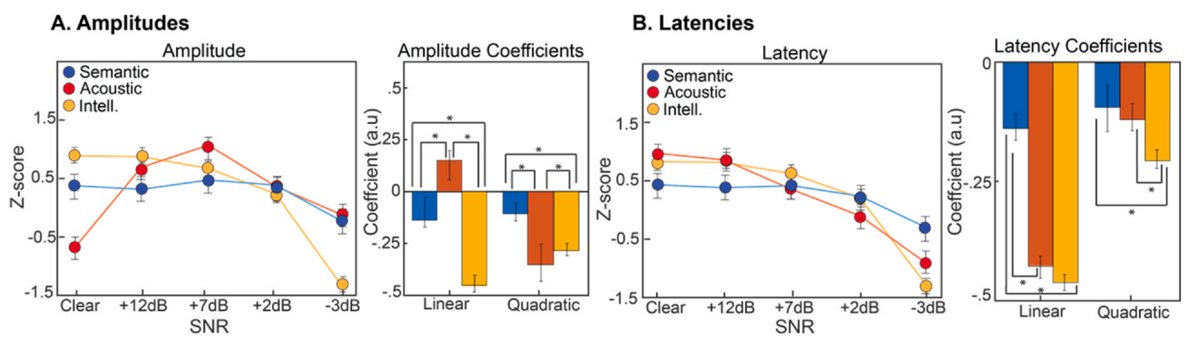

Fig. 5. Normalized acoustic, semantic, and intelligibility data. A left: The mean z-scored amplitude for the acoustic tracking response and semantic tracking response (sign-inverted such that larger values mean larger responses), as well as intelligibility data are shown as a function of SNR. **Right:** The quadratic and linear coefficients obtained by fitting 2nd order polynomial functions to the amplitude and intelligibility data. **B left:** The mean z-scored latency for the acoustic tracking response and semantic tracking response (sign-inverted such that larger values mean shorter latencies), as well as intelligibility data are shown as a function of SNR. **Right:** The quadratic and linear coefficients obtained by fitting 2nd order polynomial functions to the amplitude and intelligibility data. Note that the behavioural intelligibility data and coefficients (in yellow) are identical between panels A and B. Error bars reflect the standard error of the mean. *$p < 0.05$.

There was no difference in linear coefficients between the acoustic tracking response latency ($t_{38} = 0.439$, $p_{FDR} = 0.661$, $d = 0.0855$) as a function of SNR, and speech intelligibility as a function of SNR. This suggests that with decreasing SNR, the linear decrease in speech intelligibility was similar in degree to the linear latency increase of the acoustic TRF. However, speech intelligibility was more quadratically modulated by SNR than was the acoustic tracking response latency ($t_{38} = -4.297$, $p_{FDR} = 3.6 \times 10^{-5}$, $d = 0.835$), likely as a consequence of a substantial drop in intelligibility for the most difficult SNR (−3 dB) that was absent for the acoustic tracking response latency. Compared to the semantic tracking response latency, speech intelligibility declined more linearly with decreasing SNR ($t_{38} = 7.386$, $p_{FDR} = 2.3 \times 10^{-11}$, $d = 1.437$) and was more quadratically modulated by SNR ($t_{38} = 3.879$, $p_{FDR} = 1.7 \times 10^{-4}$, $d = 0.755$).

The comparisons described in this section suggest that speech intelligibility is affected differently by SNR compared to acoustic and semantic TRFs. The acoustic TRF latency somewhat resembled the speech intelligibility data, although the decline in intelligibility for the least favourable SNR (−3 dB) was not matched by a corresponding latency increase in the acoustic TRF. Changes in SNR did not appear to influence the semantic TRF amplitude and latency, except at the least favourable SNR. This pattern is different to that for speech intelligibility.

4. Discussion

In the current study, we investigated how the neural encoding of the acoustic envelopes and semantics of engaging, spoken stories is affected by different degrees of masking with multi-talker babble. We further examined how the effects of masker level on neural tracking relates its effects on intelligibility. We observed that the neural tracking of the acoustic and semantic features of speech are modulated by background noise in different ways. Specifically, the amplitude of acoustic envelope tracking followed a U-shape with decreasing SNR. In contrast, semantic TRF amplitude was relatively stable across SNRs, dropping only at the least favourable SNR. Latencies increased linearly with decreasing SNR. Decreases in speech intelligibility with decreasing SNR appear to most closely resemble acoustic TRF latencies, but the profile of intelligibility across SNR otherwise did not seem to entirely reflect either acoustic or semantic processing. The current data suggest complex relationships between neural encoding of acoustic and semantic features of speech and speech intelligibility under varying degrees of speech masking.

4.1. Acoustic TRF is modulated by the degree of background masking

In the current study, we observed that amplitude of the neural tracking of the speech envelope was larger at moderate SNRs than for clear speech or for less favourable SNRs (Fig. 3B; 5 A). In contrast, the latency for the acoustic tracking response increased linearly with masking level (Fig. 3C; 5 B). Previous investigations using simple speech stimuli, such as “ba” and “da” sounds, masked by broadband noise, have generally observed linear reductions in response amplitude (Martin et al., 1999; Martin and Stapells, 2005) and linear increases in response latencies with decreasing SNR (Martin et al., 1999; Finke et al., 2016; Martin et al., 1999). The latter we also observed here. Mirroring the observations for simple sounds, a few works using more complex speech stimuli have shown a larger magnitude of the acoustic TRF (Wang et al., 2020) and an increase in response latencies in the presence of competing speech, when compared to unmasked speech (Brodbeck et al., 2020).

Other recent work suggests a U-shaped relationship between neural tracking of the speech envelope and the degree of speech degradation (Hauswald et al., 2022), similar to the current study. Hauswald et al. (2022) observed that the magnitude of the acoustically derived TRF was quadratically modulated such that acoustic tracking was largest for moderate levels of noise-vocoded speech, but smaller for both clear and highly degraded noise-vocoded speech (1-channel). The authors suggest that this quadratic relation may be explained by increased attention/cognitive control associated with listening effort for moderate degradation levels (cf. Pichora-Fuller et al., 2016; Herrmann and Johnsrude, 2020; Yerkes and Dodson, 1908; Brehm and Self, 1989; Eckert et al., 2016; Kuchinsky et al., 2016), whereas less attention/cognitive control is deployed for highly intelligible speech and speech for which comprehension is too difficult (Hauswald et al., 2022). The fact that the response amplitude elicited by simple sounds, such as tones, linearly decreases with increasing masking level (Michalewski et al., 2009; Martin et al., 1999; Martin et al., 1999) suggests that the quadratic relation observed for speech may be related to factors beyond pure acoustic processing, possibly attention/cognitive control. Indeed, neural tracking of the amplitude envelope of speech is larger for attended speech compared to ignored speech in two-talker listening contexts (Verschueren et al., 2021; Fuglsang et al., 2017). The U-shaped modulation of the acoustic TRF amplitude may thus reflect increased attention or cognitive control for moderately masked, still intelligible, speech relative to clear speech, whereas neural tracking is reduced when masking reduces speech intelligibility beyond some point, and the listener essentially 'gives up' (Picou and Ricketts, 2018; Pichora-Fuller et al., 2016). The response latency of the acoustic TRF, which increased linearly with increasing masker level, may reflect the acoustic impact of speech masking on envelope tracking more directly.

4.2. Semantic TRF is modulated by masker level

We observed a negative deflection at around 300–450 ms after word onset that was associated with variations in how well a word was predicted based on semantic dissimilarity (Fig. 4). This is consistent with the original work using TRFs to investigate neural processing of semantic context in continuous speech (Broderick et al., 2018; Broderick et al., 2020; Broderick et al., 2021). This negative deflection in the TRF is also consistent with the N400 component of the event-related potential elicited by semantically incongruent words in simple sentences (Ritter et al., 1980;Nigam et al., 1992; Deacon et al., 1995; Strauβ et al., 2013).

The magnitude of the semantic tracking response was similar to that for clear speech across increasing levels of speech masking, although it declined abruptly for the least favourable −3 dB SNR condition, at which speech intelligibility was at around 55% (Fig. 2). This pattern of stable responding with an abrupt decline is reflected in the fit of a quadratic but not linear function to the data. We also observed a trend towards increasing response latency with decreasing SNR, although this effect was weak. Previous work has demonstrated that the semantic TRF response is larger for attended compared to ignored speech when it is masked by a competing talker (Broderick et al., 2018). Noise vocoding is known to influence the magnitude and latency of the N400 response (Strauβ et al., 2013), and others have demonstrated that the latency of the N400 increases when speech is masked with a babble noise (Connolly et al., 1992). Our work suggests that the semantic TRF response is relatively robust to changes in babble-noise level as long as something over 50%, but under 80%, of words are intelligible during story listening (the 5-dB resolution between SNR levels in our work does not allow a more fine-grained conclusion). It thus appears that the brain tracks semantic context well even in the presence of moderate background noise, potentially explaining why behavioural (Herrmann and Johnsrude, 2020) and neural (Irsik et al., 2022a) engagement with stories is relatively unaffected by background noise.

4.3. Changes in speech intelligibility most closely resemble changes in acoustic response latency

Previous studies have demonstrated that the N100 (acoustic) response to noise-vocoded speech correlates with comprehension scores (Obleser and Kotz., 2011). Acoustic envelope tracking has also been shown to increase with speech understanding (Decruy et al., 2019, 2020). Surprisingly, envelope tracking is larger in older compared to younger adults (Presacco et al., 2016, 2019), despite the fact that older adults typically comprehend speech less well. Semantic processing, as captured by the N400 response, is also sensitive to whether or not speech was understood (Broderick et al., 2018; Strauβ et al., 2013; Jamison et al., 2016). We investigated whether speech intelligibility, as measured by word report, is reflected in responses either to the acoustic or the semantic features of speech by examining function fits to intelligibility data, and to acoustic and semantic tracking amplitudes and latencies, as a function of SNR (Fig. 5). The U-shape of the acoustic TRF amplitudes over SNRs did not resemble the intelligibility data. The increase in acoustic TRF latency over SNRs was a closer match to the intelligibility data, but intelligibility appeared to decline less steeply than acoustic latency increased from clear speech to -3dB SNR (Fig. 5). In contrast to the decline in intelligibility from clear speech to −3 dB SNR, the semantic TRF was robust across moderate masking levels (up to and including the penultimate masker level, +2 dB).

The current intelligibility data reflect the proportion of correctly reported words (Fig. 2). Word report is not identical to comprehension of a sentence or, more generally, speech. The current study cannot speak to whether a comprehension measure, such as gist report or answers to story comprehension questions, would have resulted in a closer correspondence to the amplitudes of the semantic tracking response. Notably, word report still encompasses elements of comprehension evidenced by improved word report scores for real versus nonsense words (Kimura and Seal, 2003; Saint-Aubin and Poirier, 2000). Semantic tracking may reflect higher-level speech comprehension, that does not appear to be captured well by speech intelligibility; this may explain the differences between the word report scores and semantic tracking, when examined as a function of SNR (Fig. 4).

We examined the effect of a broad range of SNRs on neural tracking responses to acoustic and semantic properties of natural speech, which has previously not been explored fully. Our data suggest a complex relationship between intelligibility measured using word report, and neural tracking of different features of speech, over a range of masking levels. We see key differences in the way acoustics and semantics are tracked as a function of noise level; specifically, we observed that neural tracking of semantic dissimilarity, and thus context, is more resilient, when compared to acoustics and intelligibility, to challenging listening conditions, at least in healthy young adults.

5. Conclusion

In the current study, we investigated how the EEG signal tracks the amplitude envelope and the semantic content of engaging, continuous speech, and how neural tracking is affected by different degrees of multi-talker masking. We also investigated how the effect of masking level on neural tracking related to the effect of masking level on intelligibility, measured as word report for the same story materials. The amplitude of the acoustic response was greater at moderate masking levels compared either to clear speech, or to the lowest SNR, perhaps due to increased attention/increased cognitive control when speech comprehension was challenging, but manageable. In contrast, neural tracking of the semantic information was stable and robust to noise, declining only at the least favourable SNR. Response latencies increased linearly with increasing masking, more for acoustic envelope tracking than for semantic tracking. Changes in speech intelligibility with increased speech masking mirrored most closely the changes in the response latency to the acoustic envelope of speech, but were also somewhat robust to changes in SNR, averaging between 80 and 90% words reported correctly up to the least favourable SNR, where word report dropped to 50%. This stability with an abrupt decline at the lowest SNR resembles the magnitude of the neural tracking response to semantic information. Our data demonstrate that different aspects of the neural tracking response are differentially affected by noise. Furthermore, the effects of noise on tracking differ for acoustic and semantic information. These findings lend support to the idea that intelligibility, reflected in the acoustic tracking, and comprehension, reflected in the semantic tracking response, reflect on different underlying cognitive processes.

Author statement

Sonia Yasmin: Data analysis, Statistical analysis, Writing- Original draft preparation, Vanessa Irsik: Data collection, Ingrid Johnsrude: Conceptualization, Supervision, Writing- Draft revisions, Björn Herrmann: Conceptualization, Supervision, Data analysis, Writing- Draft revisions.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the Canadian Institutes of Health Research (MOP133450 to I.S. Johnsrude). BH was supported by the Canada Research Chair program (CRC-2019-00156, 232733). SY was supported by the Natural Sciences and Engineering Council of Canada CGS-D scholarship and the Vector Institute Post Graduate Affiliates Program.

References

Akram, S., Simon, J.Z., Babadi, B., 2016. Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 64 (8), 1896–1905.
Alain, C., Roye, A., Salloum, C., 2014. Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. Front. Syst. Neurosci. 8, 8.
Arnold, J.E., Fagnano, M., Tanenhaus, M.K., 2003. Disfluencies signal theee, um, new information. J. Psycholinguist. Res. 32 (1), 25–36.
Bacon, S.P., Opie, J.M., Montoya, D.Y., 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. J. Speech Lang. Hear. Res. 41 (3), 549–563.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57 (1), 289–300.
Bilger, R.C., 1984. Manual for the Clinical Use of the Revised SPIN Test. University of Illinois Press.
Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., Brennan, S.E., 2001. Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. Lang. Speech 44 (2), 123–147.
Brehm, J.W., Self, E.A., 1989. The intensity of motivation. Annu. Rev. Psychol. 40 (1), 109–131.
Brennan, S.E., Schober, M.F., 2001. How listeners compensate for disfluencies in spontaneous speech. J. Mem. Lang. 44 (2), 274–296.
Brodbeck, C., Jiao, A., Hong, L.E., Simon, J.Z., 2020. Neural speech restoration at the cocktail party: auditory cortex recovers masked speech of both attended and ignored speakers. PLoS Biol. 18 (10), e3000883.
Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. 28 (5), 803–809.
Broderick, M.P., Anderson, A.J., Lalor, E.C., 2019. Semantic context enhances the early auditory encoding of natural speech. J. Neurosci. 39 (38), 7564–7575.
Broderick, M.P., Zuk, N.J., Anderson, A.J., Lalor, E.C., 2020. More than Words: Neurophysiological Correlates of Semantic Dissimilarity Depend on Comprehension of the Speech Narrative. bioRxiv.
Broderick, M.P., Di Liberto, G.M., Anderson, A.J., Rofes, A., Lalor, E.C., 2021. Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. Sci. Rep. 11 (1), 1–12.
Carroll, J., Tiaden, S., Zeng, F.G., 2011. Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. J. Acoust. Soc. Am. 130 (4), 2054–2062.
Carter, R., Mncarthy, M., 1995. Grammar and the spoken language. Appl. Linguist. 16 (2), 141–158.
Cohen, J., 1992. Statistical power analysis. Current directions in psychological science 1 (3), 98–101.
Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *ENeuro*.
Connolly, J.F., Phillips, N.A., 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. J. Cognit. Neurosci. 6 (3), 256–266.
Connolly, J.F., Phillips, N.A., Stewart, S.H., Brake, W.G., 1992. Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. Brain Lang. 43 (1), 1–18.
Crosse, M.J., Lalor, E.C., 2014. The cortical representation of the speech envelope is earlier for audiovisual speech than audio speech. J. Neurophysiol. 111 (7), 1400–1408.
Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10, 604.
Crosse, M.J., Zuk, N.J., Di Liberto, G.M., Nidiffer, A.R., Molholm, S., Lalor, E.C., 2021. Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research. Front. Neurosci. 15.
Darwin, C.J., 2008. Listening to speech in the presence of other sounds. Phil. Trans. Biol. Sci. 363 (1493), 1011–1021.
Darwin, C.J., Carlyon, R.P., Moore, B., 1995. Hearing. Handbook of Perception and Cognition, pp. 387–424.
Das, N., Bertrand, A., Francart, T., 2018. EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. J. Neural. Eng. 15 (6), 066017.
Das, N., Vanthornhout, J., Francart, T., Bertrand, A., 2020. Stimulus-aware spatial filtering for single-trial neural response and temporal response function estimation in high-density EEG with applications in auditory research. Neuroimage 204, 116211.
Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear. Res. 229 (1–2), 132–147.
Deacon, D., Mehta, A., Tinsley, C., Nousak, J.M., 1995. Variation in the latencies and amplitudes of N400 and NA as a function of semantic priming. Psychophysiology 32 (6), 560–570.
Decruy, L., Vanthornhout, J., Francart, T., 2019. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. J. Neurophysiol. 122 (2), 601–615.
Decruy, L., Vanthornhout, J., Francart, T., 2020. Hearing impairment is associated with enhanced neural tracking of the speech envelope. Hear. Res. 393, 107961.
Devaraju, D.S., Kemp, A., Eddins, D.A., Shrivastav, R., Chandrasekaran, B., Hampton Wray, A., 2021. Effects of task demands on neural correlates of acoustic and semantic processing in challenging listening conditions. J. Speech Lang. Hear. Res. 64 (9), 3697–3706.
Di Liberto, G.M., O`Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. 25 (19), 2457–2465.
Drennan, D.P., Lalor, E.C., 2019. Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. eneuro 6 (3).
Eckert, M.A., Teubner-Rhodes, S., Vaden Jr., K.I., 2016. Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. Ear Hear. 37 (Suppl. 1), 101S.
Ehrlich, S.F., Rayner, K., 1981. Contextual effects on word perception and eye movements during reading. J. Verb. Learn. Verb. Behav. 20 (6), 641–655.
Ervin-Tripp, S.M., Küntay, A., 1997. The occasioning and structure of conversational stories. Typol. Stud. Lang. 34, 133–166.
Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., Obleser, J., 2017. Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. J. Neural. Eng. 14 (3), 036020.
Fiedler, L., Wöstmann, M., Herbst, S.K., Obleser, J., 2019. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. Neuroimage 186, 33–42.
Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., Sandmann, P., 2016. On the relationship between auditory cognition and speech intelligibility in cochlear implant users: an ERP study. Neuropsychologia 87, 169–181.
Flaherty, M.M., Browning, J., Buss, E., Leibold, L.J., 2021. Effects of hearing loss on school-aged children's ability to benefit from F0 differences between target and masker speech. Ear Hear. 42 (4), 1084–1096.
Fontan, L., Tardieu, J., Gaillard, P., Woisard, V., Ruiz, R., 2015. Relationship between speech intelligibility and speech comprehension in babble noise. J. Speech Lang. Hear. Res. 58 (3), 977–986.
Frank, S.L., 2013. Uncertainty reduction as a measure of cognitive load in sentence comprehension. Topics in cognitive science 5 (3), 475–494.
Fuglsang, S.A., Dau, T., Hjortkjær, J., 2017. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. Neuroimage 156, 435–444.
Ganong, W.F., 1980. Phonetic categorization in auditory word perception. J. Exp. Psychol. Hum. Percept. Perform. 6 (1), 110.
Gillis, M., Vanthornhout, J., Simon, J.Z., Francart, T., Brodbeck, C., 2021. Neural markers of speech comprehension: measuring EEG tracking of linguistic speech

representations, controlling the speech acoustics. J. Neurosci. 41 (50), 10316–10329.
Gordon-Salant, S., 2006. Speech perception and auditory temporal processing performance by older listeners: implications for real-world communication. N*o. 04*. In: Seminars in Hearing, vol. 27. Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA, pp. 264–268. Copyright© 2006 by.
Hale, J., 2001. A probabilistic Earley parser as a psycholinguistic model. Second meeting of the north american chapter of the association for computational linguistics.
Handy, T.C. (Ed.), 2005. Event-related Potentials: A Methods Handbook. MIT press.
Hauswald, A., Keitel, A., Chen, Y.P., Rösch, S., Weisz, N., 2022. Degradation levels of continuous speech affect neural speech tracking and alpha power differently. European Journal of Neuroscience 55 (11–12), 3288–3302.
Henry, K.S., Heinz, M.G., 2012. Diminished temporal coding with sensorineural hearing loss emerges in background noise. Nat. Neurosci. 15 (10), 1362–1364.
Herrmann, B., 2023. The perception of artificial-intelligence (AI) based synthesized speech in younger and older adults. Int. J. Speech Technol. 1–21.
Herrmann, B., Johnsrude, I.S., 2020. Absorption and enjoyment during listening to acoustically masked stories. Trends in hearing 24, 2331216520967850.
Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., Ackermann, H., 2012. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49 (3), 322–334.
Hoerl, A.E., Kennard, R.W., 1970a. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.
Hoerl, A.E., Kennard, R.W., 1970b. Ridge regression: applications to nonorthogonal problems. Technometrics 12 (1), 69–82.
Holliday, R.E., Reyna, V.F., Brainerd, C.J., 2008. Recall of details never experienced: effects of age, repetition, and semantic cues. Cognit. Dev. 23 (1), 67–78.
Holt, R.F., Bent, T., 2017. Children's use of semantic context in perception of foreign-accented speech. J. Speech Lang. Hear. Res. 60 (1), 223–230.
Hustad, K.C., 2008. The Relationship between Listener Comprehension and Intelligibility Scores for Speakers with Dysarthria. Journal of Speech, Language, and Hearing Research 51 (3), 562–573.
Iotzov, I., Parra, L.C., 2019. EEG can predict speech intelligibility. J. Neural. Eng. 16 (3), 036008.
Irsik, V.C., Johnsrude, I.S., Herrmann, B., 2022a. Neural activity during story listening is synchronized across individuals despite acoustic masking. J. Cognit. Neurosci. 34 (6), 933–950.
Irsik, V.C., Johnsrude, I.S., Herrmann, B., 2022b. Age-related deficits in dip-listening evident for isolated sentences but not for spoken stories. Sci. Rep. 12 (1), 1–17.
Jamison, C., Aiken, S.J., Kiefte, M., Newman, A.J., Bance, M., Sculthorpe-Petley, L., 2016. Preliminary investigation of the passively evoked N400 as a tool for estimating speech-in-noise thresholds. Am. J. Audiol. 25 (4), 344–358.
Jefferson, G., 1978. Sequential aspects of storytelling in conversation. Stud. Org. Convers. Interact. 1, 219–248.
Jones, D.M., Macken, W.J., 1995. Auditory babble and cognitive efficiency: role of number of voices and their location. J. Exp. Psychol. Appl. 1 (3), 216.
Kasparian, K., Steinhauer, K., 2016. Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second language acquisition. Neuropsychologia 93, 200–217.
Kennedy, J.J., 1970. The eta coefficient in complex ANOVA designs. Educ. Psychol. Meas. 30 (4), 885–889.
Kimura, D., Seal, B.N., 2003. Sex differences in recall of real or nonsense words. Psychol. Rep. 93 (1), 263–264.
Kong, Y.Y., Mullangi, A., Marozeau, J., 2012. Timbre and speech perception in bimodal and bilateral cochlear-implant listeners. Ear Hear. 33 (5), 645.
Kuchinsky, S.E., Dubno, J.R., Eckert, M.A., 2016. Advances in quantifying listening effort: growth curve analyses of pupillometry data. J. Acoust. Soc. Am. 139 (4), 2101, 2101.
Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cognit. Sci. 4 (12), 463–470.
Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.
Lovrich, D., Novick, B., Vaughan Jr., H.G., 1988. Topographic analysis of auditory event-related potentials associated with acoustic and semantic processing. Electroencephalogr. Clin. Neurophysiology Evoked Potentials Sect. 71 (1), 40–54.
Luck, S.J., 2012. Event-related potentials. *APA handbook of research methods in psychology, Vol 1: Foundations, planning, measures, and psychometrics*. American Psychological Association, pp. 523–546.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique. MIT press, Cambridge, MA.
Macken, W.J., Tremblay, S., Houghton, R.J., Nicholls, A.P., Jones, D.M., 2003. Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. J. Exp. Psychol. Hum. Percept. Perform. 29 (1), 43.
Makeig, S., Bell, A., Jung, T.P., Sejnowski, T.J., 1995. Independent component analysis of electroencephalographic data. Advances in neural information processing systems 8.
Martin, B.A., Stapells, D.R., 2005. Effects of low-pass noise masking on auditory event-related potentials to speech. Ear Hear. 26 (2), 195–213.
Martin, B.A., Kurtzberg, D., Stapells, D.R., 1999. The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds/ba/and/da. J. Speech Lang. Hear. Res. 42 (2), 271–286.
Meyer, J., Dentel, L., Meunier, F., 2013. Speech recognition in natural background noise. PLoS One 8 (11), e79279.
Michalewski, H.J., Starr, A., Zeng, F.G., Dimitrijevic, A., 2009. N100 cortical potentials accompanying disrupted auditory nerve activity in auditory neuropathy (AN): effects of signal intensity and continuous noise. Clin. Neurophysiol. 120 (7), 1352–1363.
Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
Miller, G.A., Heise, G.A., Lichten, W., 1951. The intelligibility of speech as a function of the context of the test materials. J. Exp. Psychol. 41 (5), 329.
Monaghan, J.J., Garcia-Lazaro, J.A., McAlpine, D., Schaette, R., 2020. Hidden hearing loss impacts the neural representation of speech in background noise. Curr. Biol. 30 (23), 4710–4721.
Nigam, A., Hoffman, J.E., Simons, R.F., 1992. N400 to semantically anomalous pictures and words. J. Cognit. Neurosci. 4 (1), 15–22.
Norris, D., McQueen, J.M., Cutler, A., 2003. Perceptual learning in speech. Cognit. Psychol. 47 (2), 204–238.
Obleser, J., Kotz, S.A., 2011. Multiple brain signatures of integration in the comprehension of degraded speech. Neuroimage 55 (2), 713–723.
Ochs, E., Capps, L., 1996. Narrating the self. Annu. Rev. Anthropol. 25 (1), 19–43.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1-9.
Pasupathi, M., Lucas, S., Coombs, A., 2002. Conversational functions of autobiographical remembering: long-married couples talk about conflicts and pleasant topics. Discourse Process 34, 163–192.
Pennington, J., Socher, R., Manning, C.D., 2014. Glove: Global vectors for word representation. In: Proceedings of the 2014 Confer*ence on Empirical Methods in Natural Language Processing*. EMNLP), pp. 1532–1543.
Pichora-Fuller, M.K., Kramer, S.E., Eckert, M.A., Edwards, B., Hornsby, B.W., Humes, L. E., et al., 2016. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). Ear Hear. 37, 5S–27S.
Picou, E.M., Ricketts, T.A., 2018. The relationship between speech recognition, behavioural listening effort, and subjective ratings. Int. J. Audiol. 57 (6), 457–467.
Picton, T., 2013. Hearing in time: evoked potential studies of temporal processing. Ear Hear. 34 (4), 385–401.
Pitt, M.A., Samuel, A.G., 1993. An empirical and meta-analytic evaluation of the phoneme identification task. J. Exp. Psychol. Hum. Percept. Perform. 19 (4), 699.
Pratarelli, M.E., 1995. Modulation of semantic processing using word length and complexity: an ERP study. Int. J. Psychophysiol. 19 (3), 233–246.
Presacco, A., Simon, J.Z., Anderson, S., 2016. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. J. Neurophysiol. 116 (5), 2346–2355.
Presacco, A., Simon, J.Z., Anderson, S., 2019. Speech-in-noise representation in the aging midbrain and cortex: effects of hearing loss. PLoS One 14 (3), e0213899.
Puvvada, K.C., Simon, J.Z., 2017. Cortical representations of speech in a multitalker auditory scene. J. Neurosci. 37 (38), 9189–9196.
Regev, M., Simony, E., Lee, K., Tan, K.M., Chen, J., Hasson, U., 2019. Propagation of information along the cortical hierarchy as a function of attention while reading and listening to stories. Cerebr. Cortex 29 (10), 4017–4034.
Ritter, W., Rotkin, L., Vaughan Jr., H.G., 1980. The modality specificity of the slow negative wave. Psychophysiology 17 (3), 222–227.
Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 336 (1278), 367–373.
Saint-Aubin, J., Poirier, M., 2000. Immediate serial recall of words and nonwords: tests of the retrieval-based hypothesis. Psychonomic Bull. Rev. 7 (2), 332–340.
Salmelin, R., 2007. Clinical neurophysiology of language: the MEG approach. Clin. Neurophysiol. 118 (2), 237–254.
Shi, L.F., 2014. Measuring effectiveness of semantic cues in degraded English sentences in non-native listeners. Int. J. Audiol. 53 (1), 30–39.
Simony, E., Honey, C.J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., Hasson, U., 2016. Dynamic reconfiguration of the default mode network during narrative comprehension. Nat. Commun. 7 (1), 1–13.
Slaney, M., 1988. Lyon's Cochlear Model, vol. 13. Apple Computer, Advanced Technology Group.
Slaney, M., 1998. Auditory toolbox. Interval Research Corporation, Tech. Rep 10 (1998), 1194.
Smith, N.J., Levy, R., 2013. The effect of word predictability on reading time is logarithmic. Cognition 128 (3), 302–319.
Song, J.H., Skoe, E., Banai, K., Kraus, N., 2011. Perception of speech in noise: neural correlates. J. Cognit. Neurosci. 23 (9), 2268–2279.
Song, J., Martin, L., Iverson, P., 2019. Native and Non-native Speech Recognition in Noise: Neural Measures of Auditory and Lexical Processing. In*International Congress of Phonetic Sciences*, pp. 5–9.
Song, J., Martin, L., Iverson, P., 2020. Auditory neural tracking and lexical processing of speech in noise: masker type, spatial location, and language experience. J. Acoust. Soc. Am. 148 (1), 253–264.
Strauß, A., Kotz, S.A., Obleser, J., 2013. Narrowed expectancies under degraded speech: revisiting the N400. J. Cognit. Neurosci. 25 (8), 1383–1395.
Summers, R.J., Roberts, B., 2020. Informational masking of speech by acoustically similar intelligible and unintelligible interferers. J. Acoust. Soc. Am. 147 (2), 1113–1125.
Synigal, S.R., Teoh, E.S., Lalor, E.C., 2020. Including measures of high gamma power can improve the decoding of natural speech from EEG. Front. Hum. Neurosci. 130.
Teoh, E.S., Lalor, E.C., 2020. Attention Differentially Affects Acoustic and Phonetic Feature Encoding in a Multispeaker Environment. bioRxiv.
Thanh, N.C., 2015. The Differences between Spoken and Written Grammar in English, in Comparison with Vietnamese (Las Diferencias entre la Gramática Oral y Escrita del Idioma Inglés en Comparación con el Idioma Vietnamita). Gist Education and Learning Research Journal 11, 138–153.

Tree, J.E.F., 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. J. Mem. Lang. 34 (6), 709–738.

Uhrig, S., Perkis, A., Behne, D.M., 2020. Effects of speech transmission quality on sensory processing indicated by the cortical auditory evoked potential. J. Neural. Eng. 17 (4), 046021.

Verschueren, E., Vanthornhout, J., Francart, T., 2021. The effect of stimulus intensity on neural envelope tracking. Hear. Res. 403, 108175.

Wang, L., Wu, E.X., Chen, F., 2020. Contribution of RMS-level-based speech segments to target speech decoding under noisy conditions. In: Interspeech, pp. 121–124.

Wendt, D., Hietkamp, R.K., Lunner, T., 2017. Impact of noise and noise reduction on processing effort: a pupillometry study. Ear Hear. 38 (6), 690–700.

Winn, M.B., Teece, K.H., 2021. Listening effort is not the same as speech intelligibility score. Trends in Hearing 25, 23312165211027688.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology 18 (5), 459–482.

Yuan, J., Liberman, M., 2008. Speaker identification on the SCOTUS corpus. Journal of the Acoustical Society of America 123 (5), 3878.

Zaglauer, M., Drotleff, H., Liebl, A., 2017. Background babble in open-plan offices: a natural masker of disruptive speech? Appl. Acoust. 118, 1–7.

Zekveld, A.A., Rudner, M., Johnsrude, I.S., Festen, J.M., Van Beek, J.H., Rönnberg, J., 2011. The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. Ear Hear. 32 (6), e16–e25.

Zuk, N.J., Murphy, J.W., Reilly, R.B., Lalor, E.C., 2021. Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. PLoS Comput. Biol. 17 (9), e1009358.