

1
2 **The Effects of Speech Masking on Neural Tracking of Acoustic and Semantic Features of**
3 **Natural Speech**

4
5 Sonia Yasmin¹, Vanessa C. Irsik^{1,*}, Ingrid S. Johnsrude^{1,2}, and Björn Herrmann^{3,4}
6

7
8
9 ¹ Department of Psychology & The Brain and Mind Institute,
10 The University of Western Ontario, London, ON, Canada N6A 3K7

11 ² School of Communication and Speech Disorders,
12 The University of Western Ontario, London, ON, N6A 5B7, Canada

13 ³ Rotman Research Institute,
14 Baycrest, M6A 2E1, Toronto, ON, Canada

15 ⁴ Department of Psychology,
16 University of Toronto, M5S 1A1, Toronto, ON, Canada

17
18 * Correspondence concerning this article should be addressed to Sonia Yasmin, The Brain and Mind
19 Institute, The University of Western Ontario, London, Ontario, N6A 5B7, Canada. E-mail:
20 svarma6@uwo.ca
21

22
23 **Acknowledgements:**

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

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

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.

50

Introduction

51 Many sound environments in everyday life contain background sounds, such as ambient music or speech,
52 that can mask the target speech signal, resulting in communication challenges (Mattys, 2012; Song et al.,
53 2011; Meyer et al., 2013). Segregation of speech from background sound is facilitated by a host of
54 acoustic features such as onset times and harmonicity (Carroll et al., 2011; Flaherty et al., 2021; Kong et
55 al., 2012; Darwin & Carlyon, 1995; Darwin, 2008). For example, speech signals fluctuate in amplitude
56 at the semi-regular rate at which syllables, and words are uttered, typically below 10 Hz (Rosen, 1992).
57 Because the amplitude fluctuations in speech and background sound typically differ, sensitive tracking
58 of the amplitude fluctuations of speech provides a means to segregate speech from background sound.
59 Semantic information also facilitates speech-in-noise perception. The semantic context of what has been
60 heard can be used to predict upcoming words and, in turn, improve speech intelligibility in challenging
61 listening conditions (Holt & Bent, 2017; Holiday et al., 2008; Shi, 2014; Zekveld et al., 2011; Davis &
62 Johnsrude, 2007; Miller et al., 1951, Ganong, 1980; Pitt & Samuel, 1993; Norris et al., 2003). This is
63 especially important for individuals with hearing impairments, who experience disproportionate
64 challenges in settings with noisy backgrounds (Henry & Heinz, 2012; Monaghan et al., 2020; Bacon et
65 al., 1998; Alain et al., 2014). Understanding how neural encoding of acoustic and semantic information
66 occurs in different individuals and contexts is an important step towards clinical interventions for hearing
67 loss, which are critically needed. The current study is concerned with how neural encoding of the acoustic
68 amplitude fluctuations and the semantic context (measured based on semantic dissimilarity between
69 words) of speech is affected by different degrees of background masking noise. This is accomplished by
70 measuring neural tracking-responses between naturalistic stimulus properties and associated
71 electrophysiological activity.

72 Much of the research into the neural processing of acoustic and semantic features of speech has
73 relied on brief, disconnected sentences presented in a repetitive event-related design (Uhrig et al., 2020;
74 Kasparian et al., 2016; Handy, 2005; Luck, 2014; Salmelin, 2007; Picton, 2013; Pratarelli et al., 1995;
75 Lovrich et al., 1988; Connolly et al., 1994). However, speech in everyday life is typically more
76 continuous (Schiffman et al., 1984; Jefferson et al., 1978; Ochs et al., 1992; Pasupathi et al., 2002; Ochs
77 et al., 1992), requires the integration of words into a larger semantic context and topical thread (Ehrlich
78 and Rayner, 1981; Hale, 2001; Frank, 2013; Smith & Levy, 2013), and may be intrinsically motivating
79 for a listener to comprehend. Listeners may thus engage differently with continuous speech compared to
80 disconnected sentences, and the recruited neural mechanisms may thus also differ.

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

90 The neural processing of continuous speech is often measured by calculated a linear mapping
91 between features of a continuous speech stimulus and the electro- or magnetoencephalographic
92 (EEG/MEG) signals recorded while participants listen to the speech (Crosse et al., 2016; Das et al., 2020;
93 Iotzov, & Parra, 2019; Synigal, et al., 2020). The result of such stimulus-to-neural-response mapping is
94 the temporal response function (TRF; Crosse et al., 2016; Broderick et al., 2018; Crosse et al., 2021).
95 TRF deflections can be interpreted similarly to components of the event-related potential for discrete
96 speech tokens such as words (Broderick et al., 2018; Crosse & Lalor, 2014; Luck, 2012; Luck, 2014).
97 The TRF approach has most frequently been used to investigate how acoustic properties of speech, such
98 as the amplitude envelope, are encoded in the brain, and how this differs as a function of task demands
99 (Das et al., 2020; Das et al. 2018; Verschueren et al., 2021; Fuglsang et al., 2017; Akram et al., 2016;
100 Teoh et al., 2019; Drennan & Lalor, 2019). For example, the magnitude of the TRF calculated for the
101 amplitude envelope of speech is larger for speech that is attended compared to speech that is ignored in
102 two-talker listening contexts (Verschueren et al., 2021; Fuglsang et al., 2017; Fiedler et al., 2019;
103 Puvvada & Simon., 2017; Brodbeck et al., 2020). The degree to which neural activity tracks the acoustic
104 speech envelope has also been linked to speech comprehension (Verschueren et al., 2021; Decruey et al.,
105 2019; Decruey et al., 2020).

106 Previous studies have revealed that the N100 response (bearing resemblance to the acoustic TRF)
107 to noise-vocoded speech is correlated with comprehension scores (Obleser & Kotz., 2011). Similarly,
108 acoustic envelope tracking also shows a positive relationship with intelligibility (Decruey et al., 2019;
109 Decruey et al., 2020). However, the relationship between neural tracking of acoustic speech features and
110 speech intelligibility may not be linear. When speech is parametrically degraded using noise-vocoding,
111 envelope tracking indexed by the TRF exhibits a U shape: amplitude is greatest for moderate levels of

112 degradation, and smaller both for intact and for highly degraded (1-channel vocoded) speech (Hauswald
113 et al., 2022).

114 Noise-vocoding differs substantially from speech masked by babble noise. The latter resembles
115 more closely situations that most individuals experience in everyday life, and that are reported by older
116 individuals to be challenging and effortful (Frisina et al., 1997; Gordon-Salant, 2006). Here, we
117 investigate whether, when speech is masked by a 12-talker background babble noise at different signal
118 to noise ratios, envelope tracking measured as the TRF exhibits a similar inverted U-shape to that
119 observed by Hausfeld et al (2022).

120 TRFs have also been used to investigate whether semantic features during continuous speech
121 listening are encoded in the brain (Broderick et al., 2018; Gillis et al., 2021; Devaraju et al., 2021). In
122 such studies, each word in a spoken story is represented by a high-dimensional numerical vector that
123 captures semantic information. Words for which the corresponding vectors correlate highly are more
124 semantically similar than words for which the vectors correlate less (Pennington et al., 2014; Mikolov et
125 al., 2013). By calculating correlations for consecutive words within a sentence or a story, a dissimilarity
126 score can be calculated for each word, reflecting the degree to which a word is incongruent with the
127 preceding semantic context (Broderick et al., 2018; Broderick et al., 2020; Broderick et al., 2021). These
128 dissimilarity scores are then used to calculate a “semantic” TRF, reflecting this incongruity, or
129 “surprisal”; which is thought to reflect the representation of contextual information in the brain (Gillis
130 et al, 2021; Broderick et al., 2018; Broderick et al., 2020; Broderick et al., 2021).

131 Similar to the acoustic TRF (Hauswald et al., 2022; Das et al., 2020; Das et al. 2018; Verschueren
132 et al., 2021; Fuglsang et al., 2017; Akram et al., 2016; Teoh et al., 2019; Drennan & Lalor, 2019), the
133 magnitude of the TRF calculated for the semantic dissimilarity is larger for attended compared to ignored
134 speech (Broderick et al., 2018). However, the degree to which neural encoding of semantic dissimilarity
135 is affected by speech masking is not clear. In previous studies, speech was masked by a single talker at
136 one signal-to-noise ratio (SNR), and the magnitude of the semantic TRF was reduced for the unattended
137 speaker (Broderick et al., 2018; Brodbeck et al., 2018). However, single-talker masking differs
138 substantially from multi-talker masking (Jones & Macken, 1995; Zaglauer et al., 2017; Macken et al.,
139 2003). A single-talker masker may not overlap spectrally very much with the target (depending on the
140 pitch difference between the target and masker voices), it will have a highly variable envelope that will
141 differ from that of the target, and so physical interference between target and masker will be minimal.
142 Nevertheless, a single talker masker is potentially confusable with the target, and might be distracting
143 (Summers & Roberts., 2020) enhancing masking efficacy. Twelve-talker babble, in contrast, is more

144 spectrally dense, and has a flatter envelope, and thus physically interferes with (i.e., energetically masks)
145 a single-talker target more than a single-talker masker. Furthermore, 12-talker babble does not contain
146 intelligible word-level information. Thus, results from research using single-talker masking probably will
147 not generalize to a situation in which multiple competing talkers are present. Indeed, recent studies have
148 found that intelligible single-talker maskers reduce acoustic tracking of the target speech when compared
149 to babble maskers (Song et al., 2019; Song et al., 2020).

150 How semantic context, or dissimilarity, encoding is affected by multi-talker background noise at
151 different SNRs is unknown. In the current study, we use 12-talker babble noise at different SNRs to
152 investigate how SNR affects the encoding of acoustic and semantic features of speech. Given that
153 individuals appear highly engaged in story listening even in the presence of moderate background noise
154 (Herrmann & Johnsrude, 2020; Irsik et al., 2022a), we expect that semantic processing, indexed by
155 semantic dissimilarity tracking, also remains high at moderate background noise, and will only decrease
156 for highly masked speech. Moreover, the relationship between semantic tracking and intelligibility has
157 been scarcely explored. A few studies have investigated how the N400 response, potentially similar to
158 the semantic TRF, is related to intelligibility. These studies have revealed a positive relationship between
159 the N400 response and intelligibility (Broderick et al., 2018; Strauß et al., 2013; Jamison et al., 2016),
160 suggesting that we will observe a positive relationship between semantic dissimilarity tracking and
161 behaviourally measured intelligibility.

162 Neural tracking of continuous speech is often investigated using audiobook narrations (Broderick
163 et al., 2018; Broderick et al., 2020; Broderick et al., 2021). Such materials are typically well articulated,
164 sentences build systematically on each other, and there is a clear and well-understood grammatical
165 framework in place (Thanh, 2015; Carter & Mncarthy, 1995). Speech in everyday life is subject to more
166 disfluencies than audiobook narrations as speakers often use slang, filler-words, sentence fragments,
167 corrections, unintentional pauses, and more flexible grammar (Bortfeld et al., 2001; Tree et al., 1995). It
168 is possible that these discrepancies between naturalistic speech and audiobook narrations may affect
169 intelligibility, effort, and/or neural processing (Arnold et al., 2003; Brennan et al., 2001). Because we are
170 interested primarily in naturalistic listening, we use engaging, spoken stories from the story-telling
171 podcast The Moth (<https://themoth.org>; Regev et al., 2019; Simony et al., 2016; Irsik et al., 2022a) which
172 may mirror speech in everyday life more closely than do audiobooks (Ochs & Capps, 1996; Ervin-Tripp
173 & Küntay, 1997).

174 In the current study, we use spoken stories to investigate how neural tracking of the acoustic
175 amplitude fluctuations (envelopes) and semantic context of engaging, naturalistic speech are affected by

176 background babble noise, and relate this to speech intelligibility of the same materials. We construct
177 TRFs by linearly mapping acoustic and semantic features of speech onto corresponding EEG activity
178 (Crosse et al., 2016; Crosse et al., 2021).

179 **Methods**

180 We re-analyzed EEG and behavioural data from a previous study (Irsik et al., 2022a). With a few minor
181 exceptions indicated explicitly below, the analyses, results, and conclusions are novel and do not overlap
182 with those reported previously (Irsik et al., 2022a). We provide the relevant information about stimuli,
183 procedures, and methods, and also refer to the details provided previously (Irsik et al., 2022a).

184 **Participants**

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

192 **Acoustic stimulation and procedure**

193 Each of the 39 participants listened to four spoken stories from The Moth podcast (<https://themoth.org>):
194 *Reach for the Stars One Small Step at a Time* (by Richard Garriott, ~13 min), *The Bounds of Comedy*
195 (by Colm O'Regan, ~10 min), *Nacho Challenge* (by Omar Qureshi, ~11 min), and *Discussing Family*
196 *Trees in School Can Be Dangerous* (by Paul Nurse, ~10 min). Twelve-talker babble noise, taken from
197 the revised Speech in Noise (R-SPIN) test (Bilger, 1984), was added to the stories at five different signal-
198 to-noise ratios (SNRs): clear, +12, +7, +2, -3 dB. The SNR changed every 30-33 seconds to one of the
199 five levels without repeating the same level twice in direct succession. When mixing stories with
200 maskers, both the level of the story and the babble masker were adjusted in order to ensure sound level
201 remained constant throughout each story, and was consistent across the stories. Stories were played via
202 headphones (Sennheiser HD 25 Light) in a single-walled sound-attenuating booth (Eckel Industries) and
203 participants were instructed to listen carefully to each story. After each story, participants answered ten
204 comprehension questions about the story to ensure they were paying attention.

205 Speech intelligibility for each story, measured as words reported from target phrases, across
206 different signal to noise ratios, was assessed in a separate group of 82 participants using online platforms
207 for experiment hosting (Pavlovia) and recruitment (MTurk, CloudResearch interface). Each participant
208 listened to the same materials described above; specifically one of four stories for which SNR changed
209 about every 30-33 seconds to one of five levels (clear, +12, +7, +2, -3 dB). For each story, 80 or 100
210 phrases/sentences (4-8 words) were selected for intelligibility testing (4 phrase/sentences per 30-33s
211 segment). During the experiment, one of the four spoken stories was played to a participant. The story
212 paused occasionally (about every 16 s), and the participant was asked to type the last phrase/sentence
213 uttered before the story paused into a text box. Just before the target utterance was heard, a fixation cross
214 on the screen changed colour to tell participants that they had to remember verbatim what they were
215 about to hear, and then changed colour again for the duration of the phrase/sentence. That target
216 phrase/sentence was then reported during the pause that immediately followed (for details see Irsik et al.,
217 2022a). The story then resumed from the beginning of the target utterance. Intelligibility was calculated
218 as the proportion of correctly reported words, separately for each SNR condition.

219 **EEG recording and preprocessing**

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

225 The data were pre-processed offline using custom MATLAB scripts and the Fieldtrip toolbox
226 (Oostenveld et al., 2011). Data were re-referenced by subtracting the average across both mastoids from
227 each channel. Line noise was suppressed using a 60-Hz notch filter. The data were high-pass filtered (0.5
228 Hz, 3429 points, Hann window) and low-pass filtered (22 Hz, 211 points, Kaiser window). Continuous
229 EEG data were segmented into separate time series time-locked to story onset and downsampled to 256
230 Hz. Independent components analysis was used to remove signal components reflecting blinks, eye
231 movement, and muscle activity (Makeig et al., 1996). Additional artifacts were removed after the
232 independent components analysis by setting the voltage for segments in which the EEG amplitude varied
233 more than 80 μ V within a 0.2-s period in any channel to 0 μ V (cf. Dmochowski et al., 2012, 2014; Cohen
234 and Parra, 2016). As a last step prior to TRF analyses, data were low-pass filtered at 10 Hz (141 points,
235 Kaiser window), because neural signals in the low-frequency range are most sensitive to acoustic and
236 semantic features (Zuk et al., 2021; Di Liberto et al., 2015).

237 **Speech transcription and identification of word-onset times**

238 Transcription for stories were done manually for each story. Non-words and incomprehensible mumbles
239 were ignored for the analysis of EEG. The onset time for each word in each story was obtained using the
240 Clarin's forced alignment software (Schiel, 1999). Onset times were manually verified, and incorrect
241 estimations were manually corrected.

242 **Acoustic and semantic temporal response functions**

243 We used a forward model based on the linear temporal response function (TRF; Crosse et al., 2016;
244 Crosse et al., 2021) to separately model the relationship between features of the auditory stimulus and
245 EEG activity (see Figure 1). The TRF model uses linear regression with ridge regularization (Crosse et
246 al., 2016; Crosse et al., 2021; Hoerl & Kennard 1970a; Hoerl & Kennard 1970b). Based on previous
247 work, the ridge regularization parameter λ was set to 10 (Fielder et al., 2019; Fielder et l., 2017).

248 The current TRF analyses focused on two representations of the auditory stimulus: the cochlear
249 envelope (i.e., envelope of a modelled cochleogram) and semantic dissimilarity. To estimate the acoustic
250 representation for each story, we modelled the cochleogram for the acoustic waveform of each story
251 using Lyon's Passive Ear model (Slaney, 1988a) as implemented in the Auditory Toolbox Version 2
252 (Slaney, 1998b). The toolbox calculates the auditory nerve responses using the probability of firing along
253 the auditory nerve given the acoustic properties of an input sound (Figure 1). We then averaged across
254 all auditory filters of the cochleogram. The analytic Hilbert transform of the amplitude envelope was
255 calculated. We low-pass filtered the envelope using a 40-Hz filter (Butterworth filter), calculated the first
256 derivative, and set all negative values to zero in order to obtain the amplitude-onset envelope (Fiedler et
257 al., 2017; Fiedler et al., 2010; Hertrich et al., 2012). This amplitude-onset envelope was used as a
258 regressor for the TRF analysis (Figure 1).

259 Representations for semantic dissimilarity were obtained using previously described methods
260 (Broderick et al., 2018; Broderick et al., 2018). We utilized pretrained vectors from the Global Vectors
261 for Word Representation (GloVe) project to obtain a semantic representation for each word in form of
262 numerical vectors (i.e., word embeddings; 300 dimensions; Pennington et al., 2014;
263 <https://nlp.stanford.edu/projects/glove/>). GloVe is an unsupervised learning model that maps words into
264 vector space based on their semantic relationships. The numerical vectors of words that are semantically
265 more similar are more correlated (e.g., frog vs toad) compared to the vectors of words that are
266 semantically less similar (e.g., frog vs shoe). The GloVe corpus consists of 400,000 vocabulary entries
267 and their corresponding numerical vectors. For each word of the story transcripts, we obtained the
268 corresponding word vector from GloVe, if it existed in the corpus. On average across the four stories,

269 11% of words were not available in the GloVe corpus and they were thus not considered for calculating
270 semantic dissimilarity regressors for the EEG TRF analysis. Using the word vectors, a semantic
271 dissimilarity value was calculated for each word of each story based on the local sentence context
272 (Broderick et al., 2018). Specifically, the Pearson correlation between the vector of the current word and
273 the averaged vectors across all preceding words of the sentence was calculated. Each correlation value
274 was subtracted from 1 to calculate the dissimilarity value (Figure 1). A regressor for the TRF analysis
275 was then created by placing each word's dissimilarity value at its respective word-onset time (while
276 values at all other time points were zero). This regressor was created at the sampling frequency of the
277 EEG data.

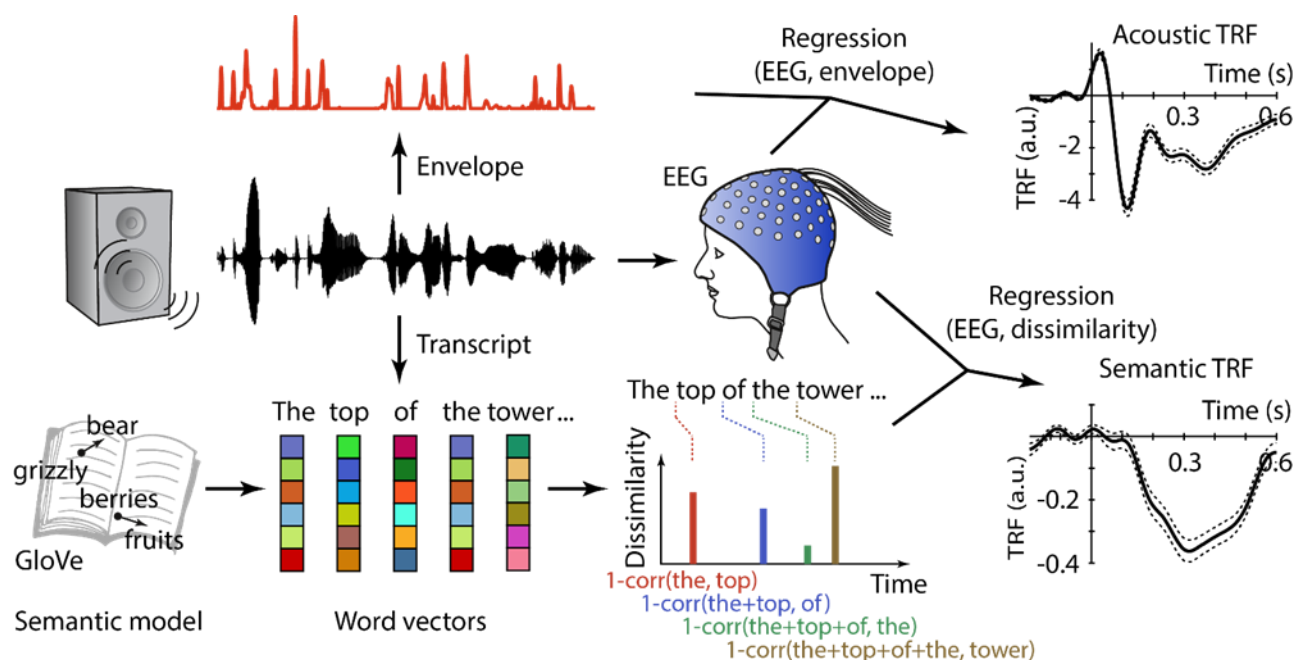


Figure 1. Schematic of the procedure for obtaining acoustic and semantic temporal response functions (TRFs). The middle row schematically shows stimulation, acoustic waveform, and EEG recording. Top row represents the calculation of the onset-amplitude envelope from the acoustic waveform. The amplitude envelope is regressed against the EEG data to obtain an acoustic TRF. The bottom row schematically displays the calculation of the semantic TRF. Global Vectors for Word Representation (GloVe) pretrained vectors, representing word meaning, were identified 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.

278 Because the dissimilarity regressor contains impulses at word onsets (with values being otherwise
279 zero), it is sensitive to brain responses associated with the acoustic onset of words. In order to mitigate
280 the influence of acoustic properties on the semantic TRF, we also calculated a 'static' TRF, where the
281 regressor is calculated using the median dissimilarity value across words for all word onsets (all other

282 samples remain zero). Hence, this regressor also contained impulses at word onsets, but the impulses
283 were all of similar magnitude and no semantic dissimilarity variations were represented.

284 For each participant, EEG channel, and ~30 s data segment corresponding to different SNR levels
285 within stories, a TRF was calculated using time windows of -0.3 s to 0.7 s between the input time series
286 of stimulus features (acoustic, semantic) and the corresponding EEG time courses, measured from word
287 onset. TRFs were averaged across segments (approximately 30 s), separately for each SNR level. To
288 obtain the final semantic TRF, we subtracted the ‘static’ TRF from the TRF derived using the
289 dissimilarity vector. The result of these TRF calculations was one acoustic TRF and one semantic TRF
290 for each SNR level, EEG channel, and participant.

291 **Analysis of the relation between SNR levels and TRF amplitude and latency**

292 For the analysis of amplitude and latency of specific deflections in the TRF, we averaged across a subset
293 of fronto-centro-parietal channels (FC1, FC2, FCz, FC3, FC4, C1, C2, Cz, C3, C4, CP1, CP2, CPz, CP3,
294 CP4, P1, P2, Pz, P3, P4) known to be sensitive to responses elicited by acoustic and semantic
295 manipulations (Broderick et al., 2018; Connolly et al., 1992; Connolly et al., 1994; Martin et al., 1999;
296 Finke et al., 2016; Martin et al., 2005). We used custom MATLAB scripts to automatically identify
297 response peaks within selected time ranges. For the acoustic TRF, we estimated the peak latency for the
298 negative deflection within 100-250 ms for each participant and SNR level. We call this negative
299 deflection the “acoustic tracking response”. Although there is obvious resemblance to the typical
300 N1/N100 component of event-related potentials (Crosse & Lalor, 2014), we want to avoid the assumption
301 that what we observe here is indeed the N1/N100. The amplitude for the acoustic tracking response was
302 calculated as the mean amplitude across 10 ms centered on a participant’s individual peak latency. Our
303 investigations for the acoustic TRF are restricted to the negativity at 100-250 ms, as visual inspection of
304 the time course in Figure 2a demonstrates this peak to be most susceptible to SNR-related changes.

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

311 We evaluate the degree to which acoustic and semantic tracking changes linearly or quadratically
312 over SNRs. To this end, a quadratic function was fitted separately to the latency and amplitude data as a
313 function of SNR levels (coded: [-2 -1 0 1 2]), separately for each participant. Quadratic fits, appropriate

314 to test whether the data conform to a U-shape, as predicted, were calculated separately for the acoustic
315 TRF (acoustic tracking response) and the semantic TRF (semantic tracking response), and separately for
316 both amplitude and latency data. The resulting linear and quadratic coefficients were tested against zero
317 using a one-sample t-test to identify significant nonzero linear and quadratic trends of TRF
318 amplitude/latency as a function of SNR.

319 We also conducted one-way repeated measures ANOVAs (rmANOVAs) to quantify effects of
320 SNR on acoustic and semantic TRF amplitudes and latencies. We performed posthoc pairwise
321 comparisons using independent samples t-tests, with false discovery rate correction (FDR; Benjamini
322 and Hochberg, 2016), between neighboring SNR levels to evaluate differences. FDR corrected p-values
323 are referred to as p_{FDR} .

324 **Relationship between acoustic and semantic TRFs, and speech intelligibility**

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

337

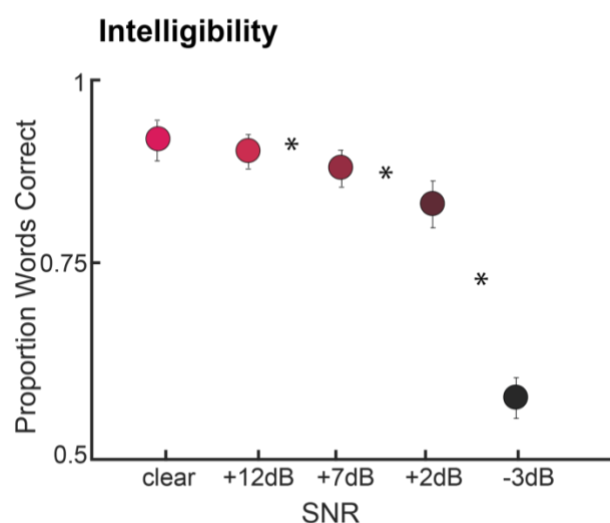


Figure 2. Effects of SNR on Intelligibility. Mean proportion of correctly reported words plotted as a function of SNR (clear, +12, +7, +2, -3 dB SNR). Asterisks indicate that the intelligibility of the two flanking SNRs differ significantly. Error bars reflect the standard error of the mean. * $p < 0.05$.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

338

339 **Effect size**

340 Effect sizes are reported as partial eta squared for ANOVAs (η^2_p ; Kennedy, 1970) and Cohen's d for t -
341 tests (d ; Cohen., 1988).

342

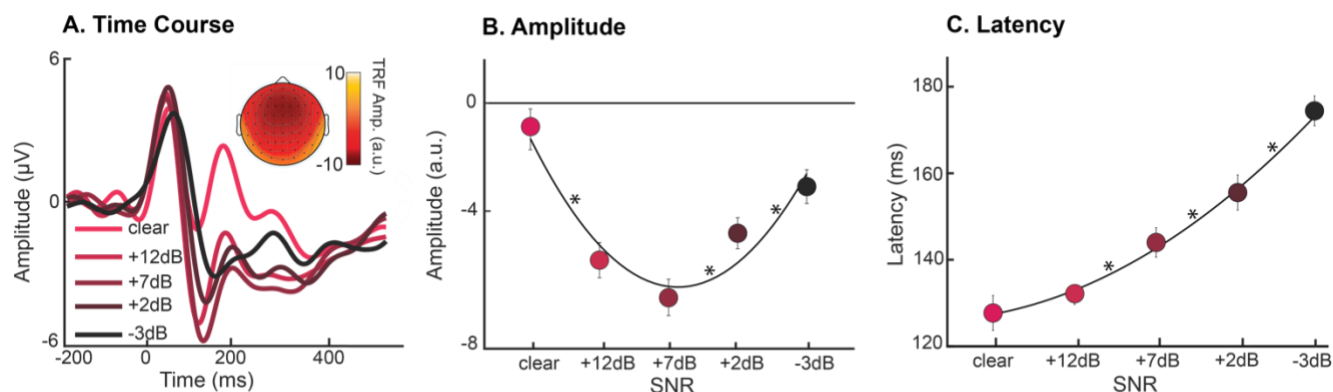
Results

343 **Amplitude and latency of acoustic TRFs are modulated by the degree of background masking**

344 We found that the amplitude of the acoustic tracking response was quadratically modulated by SNR (t_{38}
345 = 9.225, $p = 3.06 \times 10^{-11}$, $d = 1.477$). There was no significant linear modulation of acoustic tracking
346 response amplitude by SNR ($t_{38} = -1.556$, $p = 0.1281$, $d = 0.249$). To further explore the quadratic effect,
347 we conducted a rmANOVA ($F_{4,152} = 19.537$, $p = 5.44 \times 10^{-13}$, $\eta^2_p = 0.3396$), followed by pair-wise
348 comparisons between SNR levels. After false discovery rate (FDR) correction, we observed significant
349 differences for all neighboring SNR levels except between the +12 dB to +7dB conditions (clear smaller
350 than +12 dB: $t_{38} = 6.194$, $p_{FDR} = 3.1 \times 10^{-7}$, $d = 1.099$; +12 vs +7 dB: $t_{38} = 1.866$, $p_{FDR} = 0.069$, $d = 0.333$;
351 +7 greater than +2 dB: $t_{38} = 3.262$, $p_{FDR} = 0.0023$, $d = 0.607$; +2 greater than -3 dB: $t_{38} = 2.423$, $p_{FDR} =$
352 0.0203, $d = 0.460$). These results indicate a U-shape, as predicted: the acoustic tracking response
353 amplitude increased for minimal to moderate background noise relative to clear speech, and then
354 decreased again for speech that is highly masked (Figure 3B).

355 Acoustic tracking response latency increased linearly with decreasing SNR ($t_{38} = 10.979$, $p = 2.39$
356 $\times 10^{-13}$, $d = 1.758$). There was also a quadratic relationship between SNR and acoustic tracking response
357 latency ($t_{38} = 2.452$, $p = 0.0189$, $d = 0.393$). We followed up on the linear and quadratic effects with an

358 rmANOVA ($F_{4,152} = 43.61$, $p = 2.5 \times 10^{-10}$, $\eta^2_p = 0.534$) and pair-wise comparisons between neighboring
359 SNR levels. After FDR correction, all neighboring SNR levels differed significantly except the clear and
360 +12 dB conditions (clear vs +12 dB: $t_{38} = -1.254$, $p_{FDR} = 0.218$, $d = 0.209$; +12 vs +7 dB: $t_{38} = 3.880$,
361 $p_{FDR} = 0.0004$, $d = 0.667$; +7 vs +2 dB: $t_{38} = 3.355$, $p_{FDR} = 0.0018$, $d = 0.493$; +2 vs -3 dB: $t_{38} = 4.183$,
362 $p_{FDR} = 0.0002$, $d = 0.802$).



363

364 **Figure 3. Effects of SNR on acoustic TRFs.** **A.** TRF time courses (averaged across fronto-central-parietal
365 electrode cluster) for each SNR condition and scalp topography for the acoustic tracking response (negative
366 deflection at around 150 ms). **B.** The mean acoustic tracking response amplitude across participants, displayed for
367 each SNR condition. Significant differences in response magnitude exist between clear and +12 dB SNR, +7 dB
368 and +2 dB SNR, and +2 dB and -3 dB SNR. **C.** The mean acoustic tracking response latency across participants,
369 displayed for each SNR. Significant differences in response latency exist between +12 dB and +7 dB, +7 dB and
370 +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.
371 Error bars reflect the standard error of the mean. *neighbouring SNRs differ at $p < 0.05$.

372

373 **Amplitude and latency of semantic TRFs are modulated by the degree of background masking**

374 We evaluated the relationship between the degree of background masking of speech and the neural
375 responses to semantic encoding of the story (i.e., semantic dissimilarity; Figure 4). We observed that the
376 semantic tracking response amplitude was quadratically modulated by SNR ($t_{38} = 2.731$, $p = 0.0095$, $d =$
377 0.437), whereas the linear modulation was not significant ($t_{38} = 0.872$, $p = 0.389$, $d = 0.1397$). We
378 followed up on this result using an rmANOVA ($F_{4,152} = 2.706$, $p = 0.032$, $\eta^2_p = 0.0665$) and pair-wise
379 comparisons between neighboring SNR levels. After FDR correction, the semantic tracking response
380 amplitude was lower at the least favourable SNR condition compared to its neighbour (-3 dB and +2 dB
381 SNR; $t_{38} = 3.399$, $p_{FDR} = 0.0016$, $d = 0.542$), whereas tracking did not differ between any other pairs (for
382 all $p_{FDR} > 0.05$).

383 As for the acoustic tracking response, the semantic tracking response latency increased linearly
384 with decreasing SNR ($t_{38} = 2.834$, $p = 0.0073$, $d = 0.454$), and no quadratic trend was found ($t_{38} = 1.211$,

385 $p = 0.233$, $d = 0.194$). The rmANOVA revealed a significant effect of SNR ($F_{4,152} = 3.043$, $p = 0.019$,
386 $\eta^2_p = 0.074$), although no two SNR levels differed after FDR correction (neighboring and not).

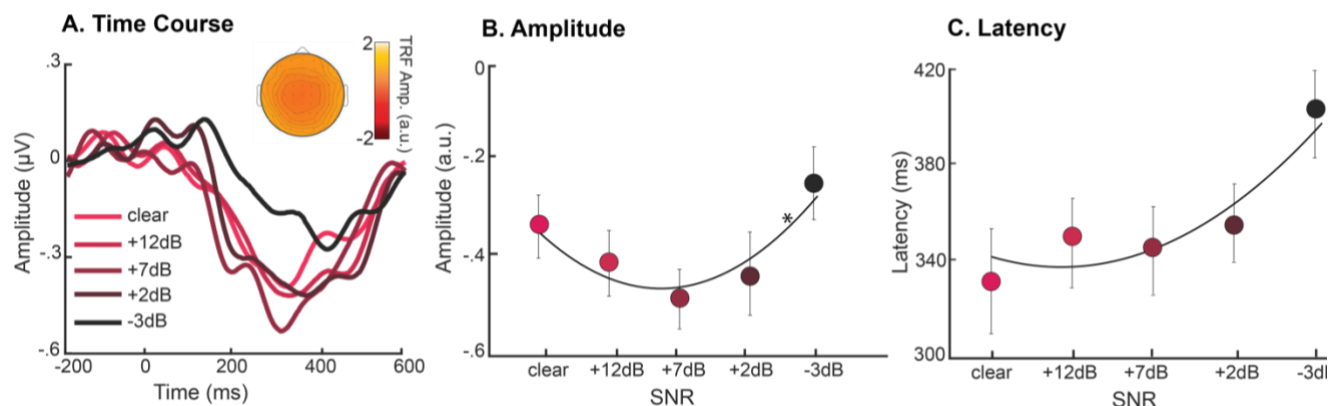


Figure 4. Effects of SNR on semantic TRFs. TRF time courses (averaged across fronto-central-parietal electrode cluster) for each SNR condition and scalp topography for the semantic tracking response (negative deflection at around 350 ms). **B.** The mean semantic tracking response amplitude across participants, 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. * $p < 0.05$.

387

388 Comparison of semantic and acoustic TRFs and their relation to speech intelligibility

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

395 The acoustic tracking response amplitude showed a stronger linear relationship with SNR
396 (positive relationship) than the semantic tracking response amplitude (negative relationship) ($t_{38} = 2.723$,
397 $p = 0.0096$, $d = 0.610$ (Figure 5). The acoustic tracking response amplitude was also more quadratically
398 related to SNR than the semantic tracking response amplitude ($t_{38} = 4.214$, $p = 1.5 \times 10^{-4}$, $d = 0.962$;
399 Figure 5). This is consistent with the observation that the semantic tracking response amplitude only
400 dropped at the lowest SNR level (-3 dB SNR; Figure 3B). As can be seen in Figure 5A, SNR had a much
401 bigger impact on the acoustic than on the semantic tracking response amplitude.

402 The acoustic tracking response latency was more strongly linearly related to SNR than the
 403 semantic tracking response latency ($t_{38} = 4.764$, $p = 2.77 \times 10^{-5}$, $d = 1.0821$). Figure 5B shows that the
 404 acoustic tracking response latency strongly increases with decreasing SNR, whereas the semantic
 405 tracking response latency is less affected; this can also be seen in Figures 3 and 4. There was no difference
 406 between acoustic and semantic tracking in the degree of quadratic fit ($t_{38} = -0.193$, $p = 0.848$, $d = 0.464$).

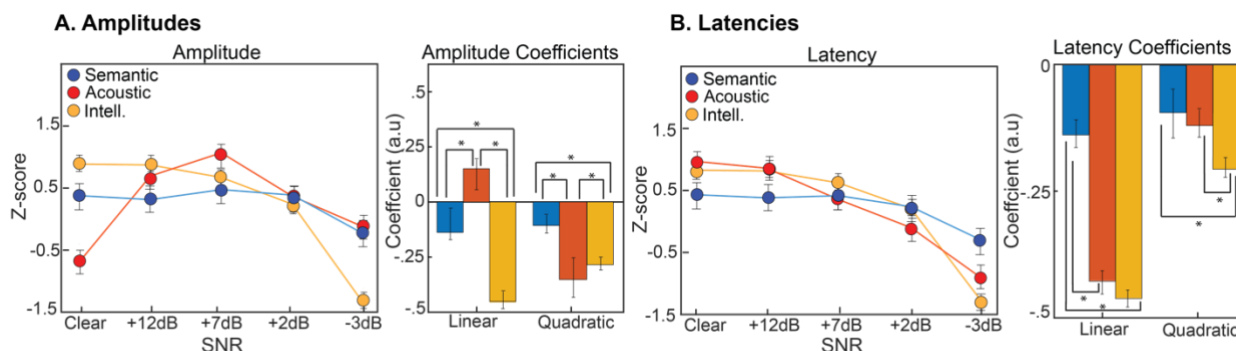


Figure 5. Normalized acoustic, semantic, and intelligibility data. **A left:** The mean z-scored amplitude for the acoustic tracking response and semantic tracking response (sign-inverted such that larger values mean larger responses), as well as intelligibility data are shown as a function of SNR. **Right:** 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. * $p < 0.05$

407 In order to compare how SNR affects speech intelligibility and neural responses, we compared
 408 coefficients obtained from linear and quadratic fits (Figure 5). We found that speech intelligibility
 409 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
 412 semantic tracking response amplitude ($t_{38} = -3.433$, $p_{FDR} = 8.2 \times 10^{-4}$, $d = 0.668$), but less quadratically
 413 modulated than the acoustic tracking response amplitude ($t_{38} = -3.823$, $p_{FDR} = 2.1 \times 10^{-4}$, $d = 0.744$). This
 414 is probably because the acoustic TRF magnitude increased significantly for intermediate SNRs, whereas
 415 intelligibility did not, and intelligibility appears to drop more precipitously at the lowest SNR (-3 dB)
 416 than does semantic tracking. These results indicate that the relationship between SNR and speech
 417 intelligibility is not entirely reflected either in the relationship between SNR and acoustically driven TRF
 418 amplitudes, or in the relationship between SNR and semantically driven TRF amplitudes.

419 There was no difference in linear coefficients between the acoustic tracking response latency (t_{38}
420 $= 0.439$, $p_{FDR} = 0.661$, $d = 0.0855$) as a function of SNR, and speech intelligibility as a function of SNR.
421 This suggests that with decreasing SNR, the linear decrease in speech intelligibility was similar in degree
422 to the linear latency increase of the acoustic TRF. However, speech intelligibility was more quadratically
423 modulated by SNR than was the acoustic tracking response latency ($t_{38} = -4.297$, $p_{FDR} = 3.6 \times 10^{-5}$, $d =$
424 0.835), likely as a consequence of a substantial drop in intelligibility for the most difficult SNR (-3 dB)
425 that was absent for the acoustic tracking response latency. Compared to the semantic tracking response
426 latency, speech intelligibility declined more linearly with decreasing SNR ($t_{38} = 7.386$, $p_{FDR} = 2.3 \times 10^{-$
427 11 , $d = 1.437$) and was more quadratically modulated by SNR ($t_{38} = 3.879$, $p_{FDR} = 1.7 \times 10^{-4}$, $d = 0.755$).

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

434

Discussion

435 In the current study, we investigated how the neural encoding of the acoustic envelopes and semantics
436 of engaging, spoken stories is affected by different degrees of masking with multi-talker babble. We
437 further examined how the effects of masker level on neural tracking relates its effects on intelligibility
438 of the same materials. We looked for particular components, or deflections, of the acoustic and semantic
439 tracking functions, thought to reflect acoustic and semantic processing. We observed these characteristic
440 deflections around 100 ms for the acoustic envelope, and around 350 ms, for semantic dissimilarity,
441 consistent with previous reports. We found that the neural tracking of the acoustic and semantic features
442 of speech are modulated by background noise in different ways. Specifically, the amplitude of acoustic
443 envelope tracking followed a U-shape with decreasing SNR, similar to what has been observed before
444 (Hauswald et al., 2022). In contrast, semantic TRF amplitude was relatively stable across SNRs, dropping
445 only at the least favourable SNR. Latencies increased linearly with decreasing SNR. The distinction
446 between linear and quadratic relationships in these responses highlights the importance of examining a
447 broad range of SNRs. Decreases in speech intelligibility with decreasing SNR appear to most closely
448 resemble acoustic TRF latencies, but the profile of intelligibility across SNR otherwise did not seem to

449 entirely reflect either acoustic or semantic processing. The current data suggest complex relationships
450 between neural encoding of acoustic and semantic features of speech and speech intelligibility.

451 **Acoustic TRF is modulated by the degree of background masking**

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

462 Other recent work suggests a U-shaped relationship between neural tracking of the speech
463 envelope and the degree of speech degradation (Hauswald et al., 2022), similar to the current study.
464 Hauswald et al (2022) observed that the magnitude of the acoustically derived TRF was quadratically
465 modulated such that acoustic tracking was largest for moderate levels of noise-vocoded speech, but
466 smaller for both clear and highly degraded noise-vocoded speech (1-channel). Hauswald et al (2022)
467 suggests that this quadratic relation may be explained by increased attention/cognitive control associated
468 with listening effort for moderate degradation levels (cf. Pichora-Fuller et al 2016; Herrmann &
469 Johnsrude 2020; Yerkes et al., 1908; Brehm & Self, 1989; Eckert et al., 2016; Kuchinsky et al., 2016),
470 whereas less attention/cognitive control is deployed for highly intelligible speech and speech for which
471 comprehension is too difficult (Hauswald et al., 2022). The fact the response amplitude elicited by simple
472 sounds, such as tones, linearly decreases with increasing masking level (Michalewski et al., 2009; Martin
473 et al., 1999; Martin et al., 2005) suggests that the quadratic relation observed for speech may be related
474 to factors beyond pure acoustic processing, possibly attention/cognitive control. Indeed, neural tracking
475 of the amplitude envelope of speech is larger for attended speech compared to ignored speech in two-
476 talker listening contexts (Verschueren et al., 2021; Fuglsang et al., 2017). The U-shaped modulation of
477 the acoustic TRF amplitude may thus reflect increased attention or cognitive control for moderately
478 masked, still intelligible, speech relative to clear speech, whereas neural tracking is reduced when
479 masking reduces speech intelligibility beyond some point, and the listener essentially ‘gives up’ (Picou

480 & Ricketts, 2018; Pichora-Fuller et al. 2016). The response latency of the acoustic TRF, which increased
481 linearly with increasing masker level, may reflect the acoustic impact of speech masking on envelope
482 tracking more directly.

483 **Semantic TRF is modulated by masker level**

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

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

506 **Changes in speech intelligibility most closely resemble changes in acoustic response latency**

507 Previous studies have demonstrated that the N100 (acoustic) response to noise-vocoded speech correlates
508 with comprehension scores (Obleser & Kotz., 2011). Acoustic envelope tracking has also been shown to
509 increase with speech understanding (Decruy et al., 2019; Decruy et al., 2020). Surprisingly, envelope

510 tracking is larger in older compared to younger adults (Presacco et al., 2016; Presacco et al., 2019),
511 despite the fact that older adults typically comprehend speech less well. Semantic processing, as captured
512 by the N400 response, is also sensitive to whether or not speech was understood (Broderick et al., 2018;
513 Strauß et al., 2013; Jamison et al., 2016). We investigated whether speech intelligibility is reflected in
514 responses either to the acoustic or the semantic features of speech by examining function fits to
515 intelligibility data, and to acoustic and semantic tracking amplitudes and latencies, as a function of SNR
516 (Figure 5).

517 The U-shape of the acoustic TRF amplitudes over SNRs did not resemble the intelligibility data.
518 The increase in acoustic TRF latency over SNRs was a closer match to the intelligibility data, but
519 intelligibility appeared to decline less steeply than acoustic latency increased from clear speech to -3dB
520 SNR (Figure 5). In contrast to the decline in intelligibility from clear speech to -3 dB SNR, the semantic
521 TRF was robust across moderate masking levels (up to and including the penultimate masker level, +2
522 dB).

523 The current intelligibility data reflect the proportion of correctly reported words (Figure 2). Word
524 report is an artificial task and is not identical to speech comprehension. It is possible that, if we had
525 measured comprehension as gist report (“do you understand the utterance, yes/no”), instead of word
526 report, we may have seen a closer correspondence to the amplitudes of the semantic tracking response.
527 We examined the effect of a broad range of SNRs on neural tracking responses to acoustic and semantic
528 properties of natural speech, which has previously not been explored fully. Our data suggest a complex
529 relationship between intelligibility measured using word report, and neural tracking of different features
530 of speech, over a range of masking levels. We see key differences in the way acoustics and semantics are
531 tracked as a function of noise level; specifically, we observed that neural tracking of semantic
532 dissimilarity, and thus context, is more resilient, when compared to acoustics and intelligibility, to
533 challenging listening conditions, at least in healthy young adults.

534

Conclusion

535 In the current study, we investigated how the EEG signal tracks the amplitude envelope and the semantic
536 content of engaging, continuous speech, and how neural tracking is affected by different degrees of multi-
537 talker masking. We also investigated how the effect of masking level on neural tracking related to the
538 effect of masking level on intelligibility, measured as word report for the same story materials. The
539 amplitude of the acoustic response was substantially greater at moderate masking levels compared either
540 to clear speech, or to the lowest SNR, perhaps due to increased attention/increased cognitive control

541 when speech comprehension was challenging, but manageable. In contrast, neural tracking of the
542 semantic information was stable and robust to noise, declining only at the least favourable SNR.
543 Response latencies increased linearly with increasing masking, more for acoustic envelope tracking than
544 for semantic tracking. Changes in speech intelligibility with increased speech masking mirrored most
545 closely the changes in the response latency to the acoustic envelope of speech, but were also somewhat
546 robust to changes in SNR, averaging between 80 and 90% words reported correctly up to the least
547 favourable SNR, where word report dropped to 50%. This stability with an abrupt decline at the lowest
548 SNR resembles the magnitude of the neural tracking response to semantic information. Our data
549 demonstrate how different features of the same speech signal are reflected in different aspects of the
550 neural tracking response, measured using EEG, and point to a complex relationship between speech
551 intelligibility and neural speech encoding.

552

References

- 553 Akram, S., Simon, J. Z., & Babadi, B. (2016). Dynamic estimation of the auditory temporal response
554 function from MEG in competing-speaker environments. *IEEE Transactions on Biomedical*
555 *Engineering*, 64(8), 1896-1905.
- 556
- 557 Alain, C., Roye, A., & Salloum, C. (2014). Effects of age-related hearing loss and background noise on
558 neuromagnetic activity from auditory cortex. *Frontiers in systems neuroscience*, 8, 8.
- 559
- 560 Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new
561 information. *Journal of psycholinguistic research*, 32(1), 25-36.
- 562
- 563 Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on
564 the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and*
565 *Hearing Research*, 41(3), 549-563.
- 566
- 567 Bentin, S., Kutas, M., & Hillyard, S. A. (1995). Semantic processing and memory for attended and
568 unattended words in dichotic listening: behavioral and electrophysiological evidence. *Journal of*
569 *Experimental Psychology: Human Perception and Performance*, 21(1), 54.
- 570
- 571 Bilger, R. C. (1984). Manual for the clinical use of the Revised SPIN test. University of Illinois Press

572

573 Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in
574 conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, 44(2), 123-147.

575

576 Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual review of psychology*, 40(1),
577 109-131.

578

579 Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous
580 speech. *Journal of Memory and Language*, 44(2), 274-296.

581

582 Brodbeck, C., Jiao, A., Hong, L. E., & Simon, J. Z. (2020). Neural speech restoration at the cocktail
583 party: Auditory cortex recovers masked speech of both attended and ignored speakers. *PLoS*
584 *biology*, 18(10), e3000883.

585

586 Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).
587 Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative
588 speech. *Current Biology*, 28(5), 803-809.

589

590 Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic context enhances the early auditory
591 encoding of natural speech. *Journal of Neuroscience*, 39(38), 7564-7575.

592

593 Broderick, M. P., Di Liberto, G. M., Anderson, A. J., Rofes, A., & Lalor, E. C. (2021). Dissociable
594 electrophysiological measures of natural language processing reveal differences in speech
595 comprehension strategy in healthy ageing. *Scientific reports*, 11(1), 1-12.

596

597 Broderick, M. P., Zuk, N. J., Anderson, A. J., & Lalor, E. C. (2020). More than Words:
598 Neurophysiological Correlates of Semantic Dissimilarity Depend on Comprehension of the Speech
599 Narrative. *bioRxiv*.

600

601 Carroll, J., Tiaden, S., & Zeng, F. G. (2011). Fundamental frequency is critical to speech perception in
602 noise in combined acoustic and electric hearing. *The Journal of the Acoustical Society of*
603 *America*, 130(4), 2054-2062.

604

605 Carter, R., & Mncarthy, M. (1995). Grammar and the spoken language. *Applied linguistics*, *16*(2), 141-
606 158.

607

608 Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd edn.). Hillsdale.

609

610 Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and
611 semantic processing of the terminal word of spoken sentences. *Journal of cognitive neuroscience*, *6*(3),
612 256-266.

613

614 Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity
615 to acoustic and semantic properties of terminal words in sentences. *Brain and language*, *43*(1), 1-18.

616

617 Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment
618 to continuous auditory speech in noise-free conditions. *Journal of Neuroscience*, *35*(42), 14195-14204.

619

620 Crosse, M. J., & Lalor, E. C. (2014). The cortical representation of the speech envelope is earlier for
621 audiovisual speech than audio speech. *Journal of neurophysiology*, *111*(7), 1400-1408.

622

623 Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response
624 function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous
625 stimuli. *Frontiers in human neuroscience*, *10*, 604.

626

627 Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear
628 modeling of neurophysiological responses to speech and other continuous stimuli: methodological
629 considerations for applied research. *Frontiers in Neuroscience*, *15*.

630

631 Das, N., Bertrand, A., & Francart, T. (2018). EEG-based auditory attention detection: boundary
632 conditions for background noise and speaker positions. *Journal of neural engineering*, *15*(6), 066017.

633

- 634 Das, N., Vanthornhout, J., Francart, T., & Bertrand, A. (2020). Stimulus-aware spatial filtering for single-
635 trial neural response and temporal response function estimation in high-density EEG with applications
636 in auditory research. *Neuroimage*, *204*, 116211.
- 637
- 638 Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of*
639 *the Royal Society B: Biological Sciences*, *363*(1493), 1011-1021.
- 640
- 641 Darwin, C. J., Carlyon, R. P., & Moore, B. (1995). Hearing. *Handbook of Perception and Cognition*.
- 642
- 643 Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface
644 between audition and speech perception. *Hearing research*, *229*(1-2), 132-147.
- 645
- 646 Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech
647 understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *Journal of*
648 *cognitive neuroscience*, *23*(12), 3914-3932.
- 649
- 650 *Meetings on Acoustics 161ASA* (Vol. 12, No. 1, p. 060001). Acoustical Society of America.
- 651
- 652 Deacon, D., Mehta, A., Tinsley, C., & Noursak, J. M. (1995). Variation in the latencies and amplitudes
653 of N400 and NA as a function of semantic priming. *Psychophysiology*, *32*(6), 560-570.
- 654
- 655 Decruy, L., Vanthornhout, J., & Francart, T. (2019). Evidence for enhanced neural tracking of the speech
656 envelope underlying age-related speech-in-noise difficulties. *Journal of neurophysiology*, *122*(2), 601-
657 615.
- 658
- 659 Decruy, L., Vanthornhout, J., & Francart, T. (2020). Hearing impairment is associated with enhanced
660 neural tracking of the speech envelope. *Hearing Research*, *393*, 107961.
- 661
- 662 Denes, P. B., Denes, P., & Pinson, E. (1993). *The speech chain*. Macmillan.
- 663
- 664 Devaraju, D. S., Kemp, A., Eddins, D. A., Shrivastav, R., Chandrasekaran, B., & Hampton Wray, A.
665 (2021). Effects of Task Demands on Neural Correlates of Acoustic and Semantic Processing in

- 666 Challenging Listening Conditions. *Journal of Speech, Language, and Hearing Research*, 64(9), 3697-
667 3706.
- 668
- 669 Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech
670 reflects phoneme-level processing. *Current Biology*, 25(19), 2457-2465.
- 671
- 672 Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to
673 competing speakers. *Proceedings of the National Academy of Sciences*, 109(29), 11854-11859.
- 674
- 675 Drennan, D. P., & Lalor, E. C. (2019). Cortical tracking of complex sound envelopes: modeling the
676 changes in response with intensity. *eneuro*, 6(3).
- 677
- 678 Dudley, H. (1939). Remaking speech. *The Journal of the Acoustical Society of America*, 11(2), 169-177.
- 679
- 680 Eckert, M. A., Teubner-Rhodes, S., & Vaden Jr, K. I. (2016). Is listening in noise worth it? The
681 neurobiology of speech recognition in challenging listening conditions. *Ear and hearing*, 37(Suppl 1),
682 101S.
- 683
- 684 Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during
685 reading. *Journal of verbal learning and verbal behavior*, 20(6), 641-655.
- 686
- 687 Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band
688 differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, 39(29),
689 5750-5759.
- 690
- 691 Ervin-Tripp, S. M., & Küntay, A. (1997). The occasioning and structure of conversational
692 stories. *Typological Studies in Language*, 34, 133-166.
- 693
- 694 Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-
695 channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed
696 speech. *Journal of neural engineering*, 14(3), 036020.
- 697

- 698 Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech
699 facilitates neural selectivity in acoustically challenging conditions. *Neuroimage*, *186*, 33-42.
700
- 701 Finke, M., Büchner, A., Ruigendijk, E., Meyer, M., & Sandmann, P. (2016). On the relationship between
702 auditory cognition and speech intelligibility in cochlear implant users: An ERP
703 study. *Neuropsychologia*, *87*, 169-181.
704
- 705 Flaherty, M. M., Browning, J., Buss, E., & Leibold, L. J. (2021). Effects of Hearing Loss on School-
706 Aged Children's Ability to Benefit From F0 Differences Between Target and Masker Speech. *Ear and*
707 *hearing*, *42*(4), 1084-1096.
708
- 709 Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence
710 comprehension. *Topics in cognitive science*, *5*(3), 475-494.
711
- 712 Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible
713 neural mechanisms. *Hearing research*, *106*(1-2), 95-104.
714
- 715 Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in
716 real-world acoustic scenes. *Neuroimage*, *156*, 435-444.
717
- 718 Gaskell, M. G., & Mirkovic, J. (Eds.). (2016). *Speech perception and spoken word recognition*.
719 Psychology Press.
720
- 721 Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of experimental*
722 *psychology: Human perception and performance*, *6*(1), 110.
723
- 724 Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T., & Brodbeck, C. (2021). Neural markers of speech
725 comprehension: measuring EEG tracking of linguistic speech representations, controlling the speech
726 acoustics. *Journal of Neuroscience*, *41*(50), 10316-10329.
727

- 728 Golumbic, E. Z., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective
729 speech envelope tracking in auditory cortex at a “cocktail party”. *Journal of Neuroscience*, 33(4), 1417-
730 1426.
- 731
- 732 Gordon-Salant, S. (2006, November). Speech perception and auditory temporal processing performance
733 by older listeners: implications for real-world communication. In *Seminars in Hearing* (Vol. 27, No. 04,
734 pp. 264-268). Copyright© 2006 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York,
735 NY 10001, USA..
- 736
- 737 Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Second meeting of the north*
738 *american chapter of the association for computational linguistics*.
- 739
- 740 Hauswald, A., Keitel, A., Chen, Y. P., Rösch, S., & Weisz, N. (2020). Degradation levels of continuous
741 speech affect neural speech tracking and alpha power differently. *European Journal of Neuroscience*.
- 742
- 743 Handy, T. C. (Ed.). (2005). *Event-related potentials: A methods handbook*. MIT press.
- 744
- 745 Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., & Ackermann, H. (2012). Magnetic brain activity phase-
746 locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech
747 signal. *Psychophysiology*, 49(3), 322-334.
- 748
- 749 Henry, K. S., & Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss
750 emerges in background noise. *Nature neuroscience*, 15(10), 1362-1364.
- 751
- 752 Herrmann, B., Henry, M. J., & Obleser, J. (2013). Frequency-specific adaptation in human auditory
753 cortex depends on the spectral variance in the acoustic stimulation. *Journal of neurophysiology*, 109(8),
754 2086-2096
- 755
- 756 Herrmann, B., & Johnsrude, I. S. (2020). Absorption and enjoyment during listening to acoustically
757 masked stories. *Trends in hearing*, 24, 2331216520967850.
- 758

- 759 Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal
760 problems. *Technometrics*, *12*(1), 55-67.
- 761
- 762 Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal
763 problems. *Technometrics*, *12*(1), 69-82.
- 764
- 765 Holliday, R. E., Reyna, V. F., & Brainerd, C. J. (2008). Recall of details never experienced: Effects of
766 age, repetition, and semantic cues. *Cognitive Development*, *23*(1), 67-78.
- 767
- 768 Holmes, E., Folkeard, P., Johnsrude, I. S., & Scollie, S. (2018). Semantic context improves speech
769 intelligibility and reduces listening effort for listeners with hearing impairment. *International Journal of*
770 *Audiology*, *57*(7), 483-492.
- 771
- 772 Holt, R. F., & Bent, T. (2017). Children's use of semantic context in perception of foreign-accented
773 speech. *Journal of Speech, Language, and Hearing Research*, *60*(1), 223-230.
- 774
- 775 Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated
776 with sustained speech processing demands. *Ear and hearing*, *34*(5), 523-534.
- 777
- 778 Iotzov, I., & Parra, L. C. (2019). EEG can predict speech intelligibility. *Journal of Neural*
779 *Engineering*, *16*(3), 036008.
- 780
- 781 Irsik, V. C., Johnsrude, I. S., & Herrmann, B. (2022a). Neural activity during story listening is
782 synchronized across individuals despite acoustic masking. *Journal of cognitive neuroscience*, *34*(6), 933-
783 950.
- 784
- 785 Irsik, V. C., Johnsrude, I. S., & Herrmann, B. (2022b). Age-related deficits in dip-listening evident for
786 isolated sentences but not for spoken stories. *Scientific reports*, *12*(1), 1-17.
- 787
- 788 Jamison, C., Aiken, S. J., Kiefte, M., Newman, A. J., Bance, M., & Sculthorpe-Petley, L. (2016).
789 Preliminary investigation of the passively evoked N400 as a tool for estimating speech-in-noise
790 thresholds. *American Journal of Audiology*, *25*(4), 344-358.

791

792 Jefferson, G. Sequential aspects of storytelling in conversation. *Stud. Org. Convers. Interact.* **1**, 219–248
793 (1978).

794

795 Jones, D. M., & Macken, W. J. (1995). Auditory babble and cognitive efficiency: Role of number of
796 voices and their location. *Journal of Experimental Psychology: Applied*, *1*(3), 216.

797

798 Kasparian, K., & Steinhauer, K. (2016). Confusing similar words: ERP correlates of lexical-semantic
799 processing in first language attrition and late second language acquisition. *Neuropsychologia*, *93*, 200-
800 217.

801

802 Kennedy, J. J. (1970). The eta coefficient in complex ANOVA designs. *Educational and Psychological*
803 *Measurement*, *30*(4), 885-889.

804

805 Kong, Y. Y., Mullangi, A., & Marozeau, J. (2012). Timbre and speech perception in bimodal and bilateral
806 cochlear-implant listeners. *Ear and hearing*, *33*(5), 645.

807

808 Kuchinsky, S. E., Dubno, J. R., & Eckert, M. A. (2016). Advances in quantifying listening effort: growth
809 curve analyses of pupillometry data. *The Journal of the Acoustical Society of America*, *139*(4), 2101-
810 2101.

811

812 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400
813 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

814

815 Kyong, J. S., Kwak, C., Han, W., Suh, M. W., & Kim, J. (2020). Effect of Speech Degradation and
816 Listening Effort in Reverberating and Noisy Environments Given N400 Responses. *Journal of Audiology*
817 *& Otology*, *24*(3), 119.

818

819 Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with
820 precise temporal resolution. *European journal of neuroscience*, *31*(1), 189-193.

821

- 822 Lovrich, D., Novick, B., & Vaughan Jr, H. G. (1988). Topographic analysis of auditory event-related
823 potentials associated with acoustic and semantic processing. *Electroencephalography and Clinical*
824 *Neurophysiology/Evoked Potentials Section*, 71(1), 40-54.
- 825
- 826 Luck, S. J. (2012). Event-related potentials.
- 827
- 828 Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT
829 press.
- 830
- 831 Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory
832 streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of*
833 *Experimental Psychology: Human Perception and Performance*, 29(1), 43.
- 834
- 835 Marmarelis, V. Z. (2004). *Nonlinear dynamic modeling of physiological systems* (Vol. 10). John Wiley
836 & Sons.
- 837
- 838 Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by
839 high-pass noise masking on N1 and the mismatch negativity to speech sounds/ba/and/da. *Journal of*
840 *Speech, Language, and Hearing Research*, 42(2), 271-286.
- 841
- 842 Martin, B. A., & Stapells, D. R. (2005). Effects of low-pass noise masking on auditory event-related
843 potentials to speech. *Ear and hearing*, 26(2), 195-213.
- 844
- 845 McCarthy, G., & Nobre, A. C. (1993). Modulation of semantic processing by spatial selective
846 attention. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 88(3), 210-
847 219.
- 848
- 849 Meyer, J., Dentel, L., & Meunier, F. (2013). Speech recognition in natural background noise. *PloS*
850 *one*, 8(11), e79279.
- 851

- 852 Michalewski, H. J., Starr, A., Zeng, F. G., & Dimitrijevic, A. (2009). N100 cortical potentials
853 accompanying disrupted auditory nerve activity in auditory neuropathy (AN): Effects of signal intensity
854 and continuous noise. *Clinical Neurophysiology*, *120*(7), 1352-1363.
- 855
- 856 Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context
857 of the test materials. *Journal of experimental psychology*, *41*(5), 329.
- 858
- 859 Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in
860 vector space. *arXiv preprint arXiv:1301.3781*.
- 861
- 862 Monaghan, J. J., Garcia-Lazaro, J. A., McAlpine, D., & Schaette, R. (2020). Hidden hearing loss impacts
863 the neural representation of speech in background noise. *Current Biology*, *30*(23), 4710-4721.
- 864
- 865 Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and
866 words. *Journal of cognitive neuroscience*, *4*(1), 15-22.
- 867
- 868 Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive*
869 *psychology*, *47*(2), 204-238.
- 870
- 871 Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of
872 degraded speech. *Neuroimage*, *55*(2), 713-723.
- 873
- 874 Ochs, E., & Capps, L. (1996). Narrating the self. *Annual review of anthropology*, *25*(1), 19-43.
- 875
- 876 Pasupathi, M., Lucas, S. & Coombs, A. Conversational functions of autobiographical remembering:
877 Long-married couples talk about conflicts and pleasant topics. *Discourse Process*. **34**, 163–192 (2002).
- 878
- 879 Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected
880 in brain and behavior. *Ear and hearing*, *39*(2), 204.
- 881

- 882 Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word
883 representation. In *Proceedings of the 2014 conference on empirical methods in natural language*
884 *processing (EMNLP)* (pp. 1532-1543).
885
- 886 Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... &
887 Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding
888 effortful listening (FUEL). *Ear and hearing, 37*, 5S-27S.
889
- 890 Picton, T. (2013). Hearing in time: evoked potential studies of temporal processing. *Ear and*
891 *hearing, 34*(4), 385-401.
892
- 893 Picou, E. M., & Ricketts, T. A. (2018). The relationship between speech recognition, behavioural
894 listening effort, and subjective ratings. *International journal of audiology, 57*(6), 457-467.
895
- 896 Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme
897 identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*(4),
898 699.
899
- 900 Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail
901 party? A late locus of selective attention to natural speech. *European Journal of Neuroscience, 35*(9),
902 1497-1503.
903
- 904 Pratarelli, M. E. (1995). Modulation of semantic processing using word length and complexity: an ERP
905 study. *International journal of psychophysiology, 19*(3), 233-246.
906
- 907 Presacco, A., Simon, J. Z., & Anderson, S. (2016). Evidence of degraded representation of speech in
908 noise, in the aging midbrain and cortex. *Journal of neurophysiology, 116*(5), 2346-2355.
909
- 910 Presacco, A., Simon, J. Z., & Anderson, S. (2019). Speech-in-noise representation in the aging midbrain
911 and cortex: Effects of hearing loss. *PloS one, 14*(3), e0213899.
912

- 913 Puvvada, K. C., & Simon, J. Z. (2017). Cortical representations of speech in a multitalker auditory
914 scene. *Journal of Neuroscience*, *37*(38), 9189-9196.
915
- 916 Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is
917 differentially modulated by attention and language experience. *Brain and Language*, *213*, 104891.
918
- 919 Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (2019). Propagation of information
920 along the cortical hierarchy as a function of attention while reading and listening to stories. *Cerebral*
921 *Cortex*, *29*(10), 4017-4034.
922
- 923 Ritter, W., Rotkin, L., & Vaughan Jr, H. G. (1980). The modality specificity of the slow negative
924 wave. *Psychophysiology*, *17*(3), 222-227.
925
- 926 Salmelin, R. (2007). Clinical neurophysiology of language: the MEG approach. *Clinical*
927 *Neurophysiology*, *118*(2), 237-254.
928
- 929 Schiffrin, D. How a story says what it means and does. *Text Interdiscip. J. Study Discourse* **4**, 313–346
930 (1984).
931
- 932 Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with
933 primarily temporal cues. *Science*, *270*(5234), 303-304.
934
- 935 Shi, L. F. (2014). Measuring effectiveness of semantic cues in degraded English sentences in non-native
936 listeners. *International journal of audiology*, *53*(1), 30-39.
937
- 938 Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016).
939 Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature*
940 *communications*, *7*(1), 1-13.
941
- 942 Simson, R., Vaughan Jr, H. G., & Ritter, W. (1977). The scalp topography of potentials in auditory and
943 visual Go/NoGo tasks. *Electroencephalography and clinical neurophysiology*, *43*(6), 864-875.
944

- 945 Slaney, M. (1988a). *Lyon's cochlear model* (Vol. 13). Apple Computer, Advanced Technology Group.
946
- 947 Slaney, M. (1998b). Auditory toolbox. *Interval Research Corporation, Tech. Rep, 10*(1998), 1194.
948
- 949 Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is
950 logarithmic. *Cognition, 128*(3), 302-319.
951
- 952 Song, J., Martin, L., & Iverson, P. (2020). Auditory neural tracking and lexical processing of speech in
953 noise: Masker type, spatial location, and language experience. *The Journal of the Acoustical Society of*
954 *America, 148*(1), 253-264.
955
- 956 Song, J., Martin, L., & Iverson, P. (2019). Native and non-native speech recognition in noise: Neural
957 measures of auditory and lexical processing. In *International Congress of Phonetic Sciences* (pp. 5-9).
958
- 959 Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: neural
960 correlates. *Journal of cognitive neuroscience, 23*(9), 2268-2279.
961
- 962 Srinivasan, N., & Zahorik, P. (2011). The effect of semantic context on speech intelligibility in
963 reverberant rooms. In *Proceedings of Meetings on Acoustics 161ASA* (Vol. 12, No. 1, p. 060001).
964 Acoustical Society of America.
965
- 966 Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: revisiting
967 the N400. *Journal of Cognitive Neuroscience, 25*(8), 1383-1395.
968
- 969 Summers, R. J., & Roberts, B. (2020). Informational masking of speech by acoustically similar
970 intelligible and unintelligible interferers. *The Journal of the Acoustical Society of America, 147*(2), 1113-
971 1125.
972
- 973 Synigal, S. R., Teoh, E. S., & Lalor, E. C. (2020). Including measures of high gamma power can improve
974 the decoding of natural speech from EEG. *Frontiers in human neuroscience, 130*.
975

- 976 Teoh, E. S., & Lalor, E. C. (2020). Attention differentially affects acoustic and phonetic feature encoding
977 in a multispeaker environment. *bioRxiv*.
978
- 979 Thanh, N. C. (2015). The Differences between Spoken and Written Grammar in English, in Comparison
980 with Vietnamese (Las Diferencias entre la Gramática Oral y Escrita del Idioma Inglés en Comparación
981 con el Idioma Vietnamita). *Gist Education and Learning Research Journal*, *11*, 138-153.
982
- 983 Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in
984 spontaneous speech. *Journal of memory and language*, *34*(6), 709-738.
985
- 986 Uhmann, S. (1992). Contextualizing relevance: On some forms and functions of speech rate changes in
987 everyday conversation. *The contextualization of language*, 297-336.
988
- 989 Uhrig, S., Perkis, A., & Behne, D. M. (2020). Effects of speech transmission quality on sensory
990 processing indicated by the cortical auditory evoked potential. *Journal of Neural Engineering*, *17*(4),
991 046021
992
- 993 Verschueren, E., Vanthornhout, J., & Francart, T. (2021). The effect of stimulus intensity on neural
994 envelope tracking. *Hearing Research*, *403*, 108175.
995
- 996 Ward, C. M., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2016). Effects of age, acoustic challenge,
997 and verbal working memory on recall of narrative speech. *Experimental aging research*, *42*(1), 97-111.
998
- 999 Wang, L., Wu, E. X., & Chen, F. (2020). Contribution of RMS-Level-Based Speech Segments to Target
1000 Speech Decoding Under Noisy Conditions. In *Interspeech* (pp. 121-124).
1001
- 1002 West, R. L., & Cohen, S. L. (1985). The systematic use of semantic and acoustic processing by younger
1003 and older adults. *Experimental aging research*, *11*(2), 81-86.
1004
- 1005 Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., & Johnsrude, I. S. (2012). Effortful
1006 listening: the processing of degraded speech depends critically on attention. *Journal of*
1007 *Neuroscience*, *32*(40), 14010-14021.

1008

1009 Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and*
1010 *hearing*, 37, 35S-43S.

1011

1012 Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation.

1013

1014 Zaglauer, M., Drotleff, H., & Liebl, A. (2017). Background babble in open-plan offices: A natural masker
1015 of disruptive speech?. *Applied Acoustics*, 118, 1-7.

1016

1017 Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnerberg, J. (2013). The effects of working memory
1018 capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society*
1019 *of America*, 134(3), 2225-2234.

1020

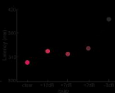
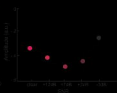
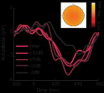
1021 Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., & Rönnerberg, J. (2011). The
1022 influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear*
1023 *and hearing*, 32(6), e16-e25.

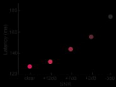
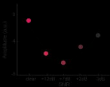
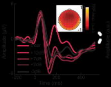
1024

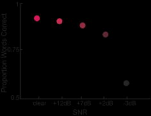
1025 Zuk, N. J., Murphy, J. W., Reilly, R. B., & Lalor, E. C. (2021). Envelope reconstruction of speech and
1026 music highlights stronger tracking of speech at low frequencies. *PLoS computational biology*, 17(9),
1027 e1009358.

1028











1=can(the, top)

1=can(the+top, of)

1=can(the+top+of, the)

1=can(the+top+of+the, away)