1 **The effect of stimulus choice on an EEG-based objective measure of speech**
2 **intelligibility**

3 **Eline Verschueren ^{a,*}, Jonas Vanthornhout ^a and Tom Francart ^a**

4 ^a *Research Group Experimental Oto-rhino-laryngology (ExpORL), Department of*
5 *Neurosciences, KU Leuven - University of Leuven, 3000 Leuven, Belgium*

6 Correspondence*:

7 Eline Verschueren

8 Herestraat 49, bus 721, 3000 Leuven, Belgium

9 fax: +3216330478

10 Tel: +32163294521

11 eline.verschueren@kuleuven.be

12 **Conflicts of Interest and Source of Funding**

13 This project has received funding from the European Research Council (ERC) under the
14 European Union's Horizon 2020 research and innovation programme (grant agreement
15 No 637424 to Tom Francart). Further support came from KU Leuven Special Research
16 Fund under grant OT/14/119. Research of Jonas Vanthornhout (1S10416N) and Eline
17 Verschueren (1S86118N) is funded by a PhD grant of the Research Foundation
18 Flanders (FWO). The authors declare no conflict of interest.

19 ABSTRACT

20 **Objectives** Recently an objective measure of speech intelligibility, based on brain
21 responses derived from the electroencephalogram (EEG), has been developed using
22 isolated Matrix sentences as a stimulus. We investigated whether this objective
23 measure of speech intelligibility can also be used with natural speech as a stimulus, as
24 this would be beneficial for clinical applications.

25 **Design** We recorded the EEG in 19 normal-hearing participants while they listened to
26 two types of stimuli: Matrix sentences and a natural story. Each stimulus was presented
27 at different levels of speech intelligibility by adding speech weighted noise. Speech
28 intelligibility was assessed in two ways for both stimuli: (1) behaviorally and (2)
29 objectively by reconstructing the speech envelope from the EEG using a linear decoder
30 and correlating it with the acoustic envelope. We also calculated temporal response
31 functions (TRFs) to investigate the temporal characteristics of the brain responses in the
32 EEG channels covering different brain areas.

33 **Results** For both stimulus types the correlation between the speech envelope and the
34 reconstructed envelope increased with increasing speech intelligibility. In addition,
35 correlations were higher for the natural story than for the Matrix sentences. Similar to
36 the linear decoder analysis, TRF amplitudes increased with increasing speech
37 intelligibility for both stimuli. Remarkable is that although speech intelligibility remained
38 unchanged in the no noise and +2.5 dB SNR condition, neural speech processing was
39 affected by the addition of this small amount of noise: TRF amplitudes across the entire

41 scalp decreased between 0 to 150 ms, while amplitudes between 150 to 200 ms
42 increased in the presence of noise. TRF latency changes in function of speech
43 intelligibility appeared to be stimulus specific: The latency of the prominent negative
44 peak in the early responses (50-300 ms) increased with increasing speech intelligibility
45 for the Matrix sentences, but remained unchanged for the natural story.

46 **Conclusions** These results show (1) the feasibility of natural speech as a stimulus for
47 the objective measure of speech intelligibility, (2) that neural tracking of speech is
48 enhanced using a natural story compared to Matrix sentences and (3) that noise and
49 the stimulus type can change the temporal characteristics of the brain responses. These
50 results might reflect the integration of incoming acoustic features and top-down
51 information, suggesting that the choice of the stimulus has to be considered based on
52 the intended purpose of the measurement.

53 1 INTRODUCTION

54 In current clinical practice speech intelligibility is measured behaviorally by asking the
55 listeners to recall the words or sentences they heard. By doing so, not only the function
56 of the auditory periphery is measured (Do the speech sounds reach the brain?), but also
57 individual skills like working memory, language knowledge and cognition. When
58 measuring speech intelligibility to evaluate the function of a hearing aid, it is desirable to
59 evaluate the auditory periphery without these extra factors. In addition, the required
60 active participation of the participant can make these measurements challenging or
61 even impossible because of poor attention or motivation, especially in small children.

62 To overcome these challenges an objective measure of speech intelligibility, where no
63 input from the participant is required, would be of great benefit. Previous studies have
64 shown that the slowly varying speech envelope is essential for speech intelligibility
65 (Shannon et al., 1995), and that it can be reconstructed from brain responses using
66 electroencephalography (EEG) or magnetoencephalography (Luo and Poeppel, 2007;
67 Aiken and Picton, 2008; Ding and Simon, 2011). Correlating the reconstructed envelope
68 from the brain response with the real acoustic envelope, results in a measure of neural
69 envelope tracking, which is related to speech intelligibility in mostly the delta (Ding et al.,
70 2014; Molinaro and Lizarazu, 2017; Vanthronhout et al., 2018) and theta band (Luo and
71 Poeppel, 2007; Ding and Simon, 2013; Lesenfants et al., 2019a).

72 Vanthornhout et al. (2018) and Lesenfants et al. (2019a) demonstrated the application
73 of this measure of neural envelope tracking in an objective measure of speech

74 intelligibility using isolated Matrix sentences as a stimulus. Matrix sentences are 5-word
75 sentences containing a proper name, verb, numeral, adjective and object with 10
76 options per word category presented randomly (e.g., 'Sofie sees ten blue socks'). In
77 their studies the same Matrix sentences were used during a standardized behavioral
78 recall experiment and an EEG measurement, enabling direct comparison of speech
79 intelligibility to envelope tracking. However, for the purpose of clinical applications, the
80 use of isolated sentences may be sub-optimal. Sentences do not reflect everyday
81 communication where syllable, word and sentence rate are less controlled and more
82 semantic top-down processing is involved. Therefore, an objective measure of speech
83 intelligibility based on fully natural speech could (1) overcome subject-related factors
84 such as motivation and therefore possible drops in attention and (2) allow intelligibility
85 measurements of any speech fragment, which is impossible today using behavioral
86 measurements but may relate better to everyday communication.

87 In this study we investigate whether the objective measure of speech intelligibility by
88 Vanthornhout et al. (2018) using Matrix sentences at a wide range of Signal-to-Noise
89 Ratios (SNRs) can also be conducted with natural running speech, such as a narrated
90 story. We know stories can be used to measure neural envelope tracking (Ding and
91 Simon, 2012, 2013; Kong et al., 2014; Petersen et al., 2017) and Ding and Simon
92 (2013) were able to link neural envelope tracking to speech intelligibility at -3 dB SNR
93 using a story. By replicating the study of Vanthornhout et al. (2018), we want to
94 investigate the relation between neural envelope tracking and speech intelligibility at a
95 wide range of SNRs and compare the use of a natural story to individual sentences to

find the most optimal stimulus for clinical use. We hypothesize neural envelope tracking will not be similar for both stimuli as speech intelligibility relies on the active integration of two incoming information streams (Hickok and Poeppel, 2007; Anderson et al., 2018): (1) the bottom-up stream that processes the acoustic features through the auditory pathway until the auditory cortex and (2) the top-down stream originating in different brain regions. Brodbeck et al. (2018) review in their introduction how neural tracking of different speech features has been used to measure top-down and bottom-up processes. We hypothesize that if neural envelope tracking is mainly a feed-forward acoustic process, results for Matrix sentences will be enhanced compared to the story because of the rigid syllable, word and sentence rate reflected in the speech envelope of the Matrix sentences. If, on the other hand, neural envelope tracking captures the interaction between the incoming acoustic speech stream and top-down information, results for the story will be enhanced because of, e.g., increased semantic processing (Di Liberto et al., 2018; Broderick et al., 2018) and attention (Kerlin et al., 2010; Ding and Simon, 2012; Mesgarani and Chang, 2012; Vanthornhout et al., 2019).

2 MATERIAL AND METHODS

2.1 Participants

Nineteen participants aged between 18 and 28 years (3 men and 16 women) took part in the experiment after providing informed consent. Participants had Flemish as their mother tongue and were all normal-hearing, confirmed with pure tone audiometry (thresholds ≤ 25 dB HL at all octave frequencies from 125 Hz to 8 kHz). The study was

117 approved by the Medical Ethics Committee UZ Leuven / Research (KU Leuven) with
118 reference S57102. All participants were unpaid volunteers.

119 **2.2 Auditory stimuli**

120 During the experiment participants listened to three different stimuli: (1) isolated Matrix
121 sentences, (2) a natural story and (3) another story used to train the linear decoder on.

122 **2.2.1 Matrix sentences**

123 Flemish Matrix sentences contain 5 words spoken by a female speaker and have a
124 fixed syntactic structure of 'proper name-verb-numeral-adjective-object', for example,
125 'Sofie sees ten blue socks' with a speech rate of 4.1 syllables/second, 2.5 words/second
126 and 0.5 sentences/second. Each category of words has 10 alternatives and each
127 sentence consists of a random combination of these alternatives which induces a rigid
128 and artificial speech rate and reduces semantic context to a bare minimum. These
129 sentences are gathered into lists of 20 sentences. Speech was fixed at a level of 60
130 dBA and the noise level varied across trials. We used speech weighted noise (SWN)
131 which has the long-term-average spectrum of the stimulus and therefore results in
132 optimal energetic masking. Matrix sentences are a validated speech material to
133 measure speech intelligibility which allows us to directly compare EEG results with
134 speech intelligibility, similar to Vanthornhout et al. (2018) and Lesenfants et al. (2019a).
135 However, Matrix sentences have a rigid speech rate and lack semantic information,
136 resulting in an artificial speech stimulus not representative for everyday communication.

137 2.2.2 Natural story

138 The natural story we used is 'De Wilde Zwanen', written by Hans Christian Andersen
139 and narrated in Flemish by Katrien Devos (female speaker) with a speech rate of
140 approximately 3.5 syllables/second, 2.5 words/second and 0.2 sentences/second.
141 Speech was fixed at a level of 60 dBA and the noise level of the SWN varied across
142 trials. The main differences between the Matrix sentences (2.2.1) and fully natural
143 speech such as this narrated story are:

- 144 1. *Prosody*: Matrix sentences are part of a standardized speech material where every
145 word is spoken at the same intensity, while the story is naturally spoken with
146 intensity variations as a consequence.
- 147 2. *Speech rate*: Matrix sentences have a rigid syllable, word and sentence rate, while
148 the story has a naturally varying speech rate because of different word and
149 sentence lengths.
- 150 3. *Semantic context*: Matrix sentences are a random combination of words,
151 minimizing the use of semantic context. The story, on the other hand, is coherent
152 speech where the use of top-down processing is triggered, e.g., knowledge about
153 time, space and characters.
- 154 4. *Lexical prediction*: The permutations of the words are different in each Matrix
155 sentence, but the words themselves become more familiar to the participants
156 during the experiment, in contrast to the story.

157 2.2.3 Decoder story

158 A children's story, 'Milan', written and narrated in Flemish by Stijn Vranken (male
159 speaker), was presented to the participants with a speech rate of 3.7 syllables/second,
160 2.6 words/second and 0.3 sentences/second. This story is 14 minutes long and was
161 presented at a level of 60 dBA without noise. The purpose of this story was to have an
162 independent continuous stimulus without background noise to train a linear decoder on
163 (Vanthornhout et al., 2018) to reconstruct the speech envelope from the EEG. We
164 intentionally selected a story, because a decoder based on Matrix sentences would be
165 sensitive to sentence onsets, because of the artificially inserted silences, and not to the
166 envelope.

167 2.3 Behavioral experiment

168 Speech intelligibility was measured behaviorally in order to compare envelope tracking
169 results in terms of speech intelligibility. We need to measure speech intelligibility for
170 both stimuli separately because they differ in content and acoustic parameters (speaker,
171 speech rate, intonation). Adding a similar level of background noise will therefore not
172 result in a similar level of speech intelligibility (Decruij et al., 2018).

173 Before the EEG experiment we conducted a Matrix test. This test starts with 2 training
174 lists followed by 3 testing lists of 20 sentences each at different SNRs: -9.5; -6.5 and -
175 3.5 dB SNR. Speech was fixed at a level of 60 dB A and the noise level varied in
176 random order. Speech and noise were presented to the right ear. Participants had to
177 recall the sentence they heard. By counting the correctly recalled words, a percentage

178 correct per presented SNR was calculated. Next, a psychometric function was fitted on
179 the data points, similar to what is done in clinical practice.

180 To measure speech intelligibility for the natural story, we cannot ask the participants to
181 recall every word, instead we used a rating method during the EEG experiment.
182 Participants were asked to rate their speech intelligibility with the following question
183 appearing visually on a screen in front of them: 'Which percentage of the words did you
184 understand?' at the presented SNRs (-12.5; -9.5; -6.5; -3.5; -0.5 and 2.5 dB SNR). In
185 addition to the recall procedure for the Matrix sentences before the EEG experiment, we
186 also asked 9 of the 19 participants to rate their speech intelligibility for the Matrix
187 sentences during the EEG, similar to the natural story.

188 **2.4 EEG experiment**

189 Ten participants started the EEG experiment by listening to Matrix sentences followed
190 by the natural story. The remaining 9 participants did this in the reversed order. The
191 decoder story was presented in between. The natural story was cut in 7 equal parts of
192 approximately 4 minutes long, which we presented in chronological order. The first part
193 was always presented in silence to optimize comprehension of the storyline. The
194 following 6 parts were presented at 6 different SNRs in random order: -12.5; -9.5; -6.5; -
195 3.5; -0.5 and 2.5 dB SNR. The Matrix sentences were concatenated into 7 lists of 40
196 sentences with a silent gap between the sentences randomly varying between 0.8 and
197 1.2 seconds. Each 2-minute trial, containing 40 sentences at a particular SNR, was
198 presented twice to analyze test-retest reliability. The SNRs were the same SNRs as

199 used for the natural story, also in random order and including the condition without
200 noise. To maximize attention and keep the participants motivated, questions were
201 asked about each SNR trial, for example, 'What happened after sunset?' (natural story)
202 or 'Which colors of boats were mentioned?' (Matrix sentences). The answers were not
203 used for further analysis. After the question, the participants were asked to rate their
204 speech intelligibility with the following question: 'Which percentage of the words did you
205 understand?'.
Accepted Version

206 **2.5 Experimental setup**

207 Recordings were made in a soundproof and electromagnetically shielded room.
208 Speech was presented bilaterally at 60 dBA and the setup was calibrated using a 2cm3
209 coupler of the artificial ear (Brüel & Kjær 4152, Denmark) for each stimulus. The stimuli
210 were presented using APEX 3 (Francart et al., 2008), an RME Multiface II sound card
211 (Germany) and Etymotic ER-3A insert phones (Illinois, USA). A 64-channel BioSemi
212 ActiveTwo (the Netherlands) EEG recording system was used for the EEG recordings at
213 a sample rate of 8192 Hz. Participants sat in a comfortable chair and were asked to
214 move as little as possible during the recordings. There was no fixation point, but they
215 were instructed to keep their eyes open. The participants could see the experimenters
216 screen where questions appeared. They gave their responses by talking through a
217 microphone inside the EEG booth. We inserted a small break between the behavioral
218 and the EEG part and between the Matrix sentences and the natural story if necessary.

219 **2.6 Signal processing**

220 In this study we measured neural envelope tracking and linked this to speech
221 intelligibility and stimulus type (natural story versus isolated Matrix sentences). Neural
222 envelope tracking was calculated in two ways: We correlated the acoustic speech
223 envelope (2.6.1) with the speech envelope reconstructed from the EEG response
224 (2.6.2) with the help of a linear decoder. Secondly, we calculated temporal response
225 functions (TRFs) to investigate the temporal characteristics of the brain responses in the
226 EEG channels covering the scalp (2.6.3).

227 2.6.1 Acoustic envelope

228 The acoustic speech envelope was extracted from the stimulus according to Biesmans
229 et al. (2017), using a gammatone filterbank followed by a power law. We used a
230 filterbank containing 28 channels spaced by 1 equivalent rectangular bandwidth with
231 center frequencies from 50 Hz until 5000 Hz. The absolute value of each sample in
232 each channel was raised to the power of 0.6. All 28 channel envelopes were averaged
233 which resulted in one single envelope. As a next step, the acoustic speech envelope
234 was band-pass filtered, similar to the EEG signal, in the delta (0.5-4 Hz) or theta (4-8
235 Hz) frequency band with a Chebyshev filter with 80 dB attenuation at 10% outside the
236 passband. Only these low frequencies were further processed, because they contain
237 the information of interest of the slowly varying speech envelope.

238 2.6.2 Envelope reconstruction

239 After applying an anti-aliasing filter, the EEG data was downsampled from 8192 Hz to
240 256 Hz to reduce processing time and referenced to an average of the electrodes. Next,

241 EEG artefact rejection was done using a multi-channel Wiener filter (MWF) (Somers et
242 al., 2018). the MWF was calculated on the long decoder story without noise and applied
243 on the shorter Matrix and natural story SNR trials. After artefact rejection, the signal was
244 bandpass filtered, similar to the acoustic speech envelope and the sample rate was
245 further decreased from 256 Hz to 128 Hz. A schematic overview is shown in Figure 1.

246 To enable reconstruction of the speech envelope from the neural data as a measure
247 of neural envelope tracking, a linear decoder was created in the delta- (0.5-4 Hz) and
248 theta (4-8 Hz) band separately using the mTRF toolbox (Lalor et al., 2006, 2009) similar
249 to the decoder used in Vanthornhout et al. (2018). As speech elicits neural responses
250 with some delay, the decoder not only attributes weights to each EEG channel (spatial
251 filter), but it also takes the shifted neural responses of each channel into account
252 (temporal filter), resulting in a matrix R containing the shifted neural responses of each
253 channel. If g is the linear decoder and R the shifted neural data, the reconstruction of
254 the speech envelope $\hat{s}(t)$ was obtained by $\hat{s}(t) = \sum_n \sum_\tau g(n, \tau) R(t + \tau, n)$ with t the time
255 index, n ranging over the recording electrodes and τ ranging over the integration
256 window, i.e., the number of post-stimulus samples used to reconstruct the envelope.
257 The decoder was calculated by solving $g = (RR^T)^{-1}(Rs^T)$ with s the speech envelope
258 and applying ridge regression to prevent overfitting. We used an integration window of
259 250 ms post-stimulus resulting in the decoder matrix g of 64 (EEG channels) x 33 (time
260 delays within the integration window). The decoder was created using the Milan story
261 (14 minutes) without any noise.

262 As a last step the envelope was reconstructed by applying the decoder to both test

stimuli, the Matrix sentences and the natural story, at various noise levels. Each SNR trial consisted of 2 presentations of 80 seconds of speech after removing the artificially inserted silences between Matrix sentences varying between 0.8 and 1.2 seconds. To measure how similar this reconstructed envelope was to the acoustic envelope as a measure for neural envelope tracking, we calculated the bootstrapped Spearman correlation using Monte Carlo sampling (Pernet et al., 2012) after removing the silences in the stimulus and the corresponding part in the EEG. Removing the silences is necessary as the Matrix sentences contain quasi-regular silent gaps between the sentences which would be a confound. To check whether the obtained correlations were significant, we calculated the significance level of the correlation by correlating random permutations of the real and reconstructed envelope 1000 times and taking percentile 2.5 and 97.5 to obtain a 95% confidence interval. Additionally we calculated the chance levels of both stimuli to investigate whether the decoder has a preference. We hypothesized that because the decoder is a story and not a set of Matrix sentences, the decoder could be better suited to decode the natural story. To obtain the chance levels we reconstructed the envelope of the natural story similar to the standard analysis. Next we correlated the reconstructed envelope of each story trial with the acoustic envelope of all trials of both the natural story (except for the trial used) and the Matrix sentences and compared both.

2.6.3 Temporal response function estimation

The analysis above integrates all neural activity over channels and time lags, i.e. the post-stimulus samples used to create the decoder, and requires a decoder trained on a

285 separate story. To have a closer look at the spatiotemporal profile of the neural
286 responses and remove the assumption that neural processing is similar for the decoder
287 story and the test stimuli in different noise conditions, we calculated TRFs. A TRF is a
288 linear filter that describes how the acoustic speech envelope of the stimulus is
289 transformed into neural responses. It can be used to predict the EEG from the acoustic
290 envelope. This is the inverse approach of the previously mentioned envelope
291 reconstruction where the acoustic envelope is reconstructed from the EEG.

292 We calculated a TRF for every electrode channel in every participant per stimulus per
293 SNR condition. The first signal processing steps are identical to the envelope
294 reconstruction model starting with downsampling to 256 Hz, artefact rejection with
295 MWF, filtering (0.5-8 Hz) and further downsampling to 128 Hz. Next, TRFs were
296 calculated using the boosting algorithm (David et al., 2007; Brodbeck et al., 2018) with
297 an L2 error norm (using the Eelbrain source code (Brodbeck, 2017)). In summary,
298 boosting is an iterative algorithm starting from a TRF consisting of zeros. With each
299 iteration, the mean-squared error (MSE) is calculated for the prediction after changing
300 all TRF parameters separately by a small amount. The best resulting change after one
301 iteration (smallest MSE), is added to the TRF. This process is repeated until no further
302 relevant improvement is possible (David et al., 2007).

303 After calculation, the TRFs were convolved with a rotationally symmetric Gaussian
304 kernel of 5 samples long ($SD=2$) to smooth over time lags. To analyze the TRFs in the
305 time domain, we investigated the latency and amplitude of the negative and positive
306 peaks occurring within 0 and 500 ms after the stimulus (Ding and Simon, 2011; Obleser

307 and Kotz, 2011; Ding and Simon, 2012; Ding et al., 2014).

308 **2.7 Statistical Analysis**

309 Statistical analysis was performed using MATLAB (version R2016b) and R (version
310 3.3.2) software. The significance level was set at $\alpha=0.05$ unless otherwise stated.

311 For the behavioral tests and envelope reconstruction we compared dependent
312 samples (e.g. test- retest) using a nonparametric Wilcoxon signed-rank test. The
313 correlation between the acoustical envelope and the envelope reconstructed from the
314 EEG was correlated with SI for every filter band and every stimulus using Spearman's
315 rank correlation. Next, we assessed the relationship between speech intelligibility,
316 envelope reconstruction, filter band and stimulus by constructing a linear mixed effect
317 (LME) model with the following formula:

$$318 \text{ } corr \sim SI + stimulus + band + SI : band + SI : stimulus + SI : band : stimulus$$

319 where *corr* is defined as the Spearman correlation between the reconstructed and the
320 acoustic envelope, and fixed and interaction effects of *SI* (speech intelligibility), *stimulus*
321 (Matrix sentences or natural story) and *band* (the delta or theta filter band). An
322 additional random effect of intercept of the participants was included in the model to
323 allow neural tracking to be higher or lower per subject. To control for the different levels
324 of SNRs used for both stimuli to obtain a same level of SI, we constructed the exact
325 same model, but in function of SNR instead of SI.

326 To control if every chosen fixed and random effect benefited the model the Akaike

327 Information Criterion (AIC) was calculated which estimates a goodness-of-fit measure,
328 while correcting for model complexity. The model with the lowest AIC was selected and
329 its residual plot was analyzed to assess the normality assumption of the LME residuals.
330 Unstandardized regression coefficients (beta) with 95% confidence intervals and p-
331 value are reported in the results section.

332 To investigate which time samples of the TRF were significantly different from zero,
333 we conducted a cluster-based permutation test. To explore the potential significant
334 differences between the natural story and the Matrix sentences at the different SNRs,
335 we conducted a cluster-based analysis with a post hoc Bonferroni adjustment explained
336 in detail by Maris and Oostenveld (2007). Spearman's rank correlation was used to
337 investigate the possible change of amplitude and latency of the temporal-occipital peaks
338 over time.

339 **3 RESULTS**

340 **3.1 Behavioral speech intelligibility**

341 During the experiment we measured speech intelligibility behaviorally at different
342 SNRs for every participant. Figure 2 shows that the natural story (rating method) was
343 significantly more difficult than the Matrix sentences (recall method) ($p < 0.001$,
344 $CI(95\%) = [15.99; 23.34]$, $n=19$, Wilcoxon signed-rank test). This indicates that the
345 same SNR does not result in the same level of speech intelligibility for the different
346 stimuli. To be able to compare the natural story with the Matrix sentences, we need to

347 account for this.

348 To check whether the used method to measure speech intelligibility, rate (natural
349 story) versus recall (Matrix sentences), did not influence the results, we asked 9 of the
350 participants to rate their speech intelligibility for the Matrix sentences, similar to the
351 natural story, in addition to the standardized recall method. Comparing their rate and
352 recall scores for the same Matrix sentences at 3 SNRs did not reveal any significant
353 difference (-9.5 dB SNR: $p=0.19$, $CI(95\%)=[-11.50; 22.00]$; -6.5 dB SNR: $p=0.06$,
354 $CI(95\%)=[-29.50; 1.50]$; -3.5 dB SNR: $p=0.41$, $CI(95\%)=[-9.00; 2.75]$; $n=9$, Wilcoxon
355 signed-rank test).

356 **3.2 Envelope reconstruction**

357 To measure neural envelope tracking, we calculated the Spearman correlation
358 between the reconstructed envelope and the acoustic envelope. A test-retest analysis
359 showed no significant difference between test and retest correlations ($p=0.746$,
360 $CI(95\%) = [-0.004; 0.006]$, Wilcoxon signed-rank test), therefore we averaged the
361 correlation of the test and retest conditions resulting in one correlation per participant
362 per SNR per stimulus. Next, no significant difference was found between the chance
363 levels of the stimuli ($p=0.534$, $CI(95\%)=[-0.005; 0.003]$, Wilcoxon signed-rank test). The
364 95% confidence interval of this non significant difference was similar to the test-retest
365 variability ($CI(95\%)=[-0.005; 0.006]$), indicating that there is no important decoder
366 preference towards one of the stimuli.

367 We analyzed neural envelope tracking in the delta (0.5-4 Hz) and the theta (4-8 Hz)

band for the Matrix sentences and the natural story at various levels of speech intelligibility. Figure 3 shows that when speech intelligibility increases, the correlation between the acoustic and the reconstructed envelope, i.e. neural envelope tracking, increases for every filter band and every stimulus tested ($p < 0.001$, table 1, Spearman rank correlation).

To additionally investigate the influence of stimulus choice, we created an LME model as a function of speech intelligibility. The analysis shows that neural envelope tracking is enhanced for the natural story compared to the Matrix sentences (fixed effect stimulus, $p = 0.010$, LME, table 2). This enhancement does not significantly depend on the level of speech intelligibility or filter band (interaction effect SI:stimulus, $p = 0.155$; interaction effect SI:band:stimulus, $p = 0.912$; LME, table 2). Further, neural envelope tracking in the delta band (0.5-4 Hz) is higher than in the theta band (4-8 Hz) (fixed effect band, $p < 0.001$, LME, table 2) with a steeper slope in the delta band (0.5-4 Hz) (interaction effect SI:band, $p < 0.001$, LME, table 2).

When conducting the same analysis using SNR as a predictor for speech intelligibility, the same fixed and interaction effects were found to be significant as for the SI analysis (table 3). This shows that even at the same SNR neural envelope tracking for the natural story is enhanced compared to the Matrix sentences, making it impossible to disentangle between the effects of SNR and SI with the current data.

3.3 Temporal response function

The analysis above integrates all different time lags and channels to obtain an optimal

reconstruction of the envelope and requires a decoder trained on a separate story. In the following analysis we focus on how the neural responses follow the envelope in the time and spatial domain and remove the assumption that neural processing is similar for the decoder story and the test stimuli by investigating TRFs. TRFs were calculated on an individual level. This resulted in 868 TRFs per participant (64 channels x 2 stimuli x 7 SNRs). To visualize topographies, we averaged the TRFs per stimulus per SNR over participants. To investigate the time-course of the TRFs, we averaged TRFs for a temporal-occipital channel selection (Figure 4). This selection is based on the TRF results shown in Figure 5. A cluster-based permutation test (Maris and Oostenveld, 2007) shows the TRF samples significantly different from zero, highlighted in bold in Figure 6.

3.3.1 Effect of SNR on TRF

Figure 5 shows the spatiotemporal activation profile of respectively the Matrix sentences and the natural story. In the no-noise condition both stimuli show positive central and negative parieto-occipital amplitudes over time. When a small amount of noise is added and speech intelligibility remains almost unchanged from the no-noise condition (SNR=2.5 dB SNR; Matrix sentences: median SI=99.9%, sd=0.2; Natural story: median SI=99.0%, sd=4.7), the amplitudes across the entire scalp decrease between 0 to 150 ms, while amplitudes between 150 to 200 ms increase in both stimuli. Between 50 and 100 ms amplitudes even swap polarities from positive to negative in the centro-frontal channels (comparing the first 2 rows of the topographies in figure 5). When more noise is added (rows 3-7, figure 5) and speech intelligibility decreases

411 positive central and negative parieto-occipital activation decreases, especially in the 150
412 to 200 ms time lag. In the 50 to 100 ms time lag, on the other hand, the negative central
413 activation increases with decreasing speech intelligibility and reaches a maximum at
414 SNR=-3.5 dB SNR.

415 To zoom in on the amplitude changes over time, we visualized an average TRF for the
416 temporal-occipital channels per SNR in Figure 6. When speech intelligibility is very low
417 (SNR<-12.5 dB SNR) both stimuli have very low responses over time. With increasing
418 speech intelligibility, TRF amplitudes also increase gradually. Figure 7 shows the
419 latency and amplitude results of the negative peak that can be found around 100 ms on
420 a participant level over speech intelligibility. It was determined individually by selecting
421 the most negative amplitude of the TRF between 50 and 300 ms. With decreasing
422 speech intelligibility the amplitude of the negative peak per participant decreases for
423 both stimuli (Matrix sentences: Spearman rank correlation=0.49, $p<0.001$; Natural story:
424 Spearman rank correlation=0.26, $p=0.005$). A centro-frontal channel selection, which is
425 often used in a clinical setting, reveals similar peaks and significances although with
426 different polarity (Appendix A). This is in line with the patterns shown on the
427 topographies in figure 5: Both channel areas are related, the magnitudes vary in the
428 same way over SNR, but the polarities are swapped.

429 3.3.2 Effect of stimulus type on TRF

430 Besides the decreasing amplitude, latency also decreases for the Matrix sentences
431 with decreasing speech intelligibility (Spearman rank correlation=0.46, $p<0.001$). For the
432 natural story, on the other hand, latency is not significantly related to speech

433 intelligibility (Spearman rank correlation=0.02, $p=0.835$)).

434 Next to the difference between the Matrix sentences and the natural story concerning
435 latency changes, other stimulus dependent differences can be found. First, a cluster
436 analysis (Maris and Oostenveld, 2007) over all participants revealed significant
437 differences ($\alpha=0.025$) between both stimuli in the no-noise condition with larger
438 amplitudes for the Matrix sentences in the central and parieto-occipital channels,
439 highlighted in red in Figure 5. In contrast to this stimuli driven difference in the no-noise
440 condition, no significant differences between both stimuli could be found in the presence
441 of background noise. Second, in addition to the prominent negative peak between 100
442 and 200 ms, a positive significant peak arises around 300 ms for the Matrix sentences
443 at -9.5 dB SNR (Figure 6), while this is not the case for the natural story.

444 4 DISCUSSION

445 In this study we investigated whether the objective measure of speech intelligibility by
446 Vanthornhout et al. (2018) using Matrix sentences can also be conducted with natural
447 speech as this would be beneficial for clinical applications. To that end, we tested 19
448 normal-hearing participants. They listened to both the Matrix sentences and a natural
449 story at varying levels of speech intelligibility while their EEG was recorded. We found
450 that it is feasible to use natural speech as a stimulus for the objective measure of
451 speech intelligibility and that noise and the stimulus type can change the temporal
452 characteristics of the brain responses over the scalp.

4.1 The same SNR does not result in similar speech intelligibility for different stimuli

As a first step we measured speech intelligibility behaviorally for both stimuli at different noise levels. The results show that the same SNR does not result in similar speech intelligibility for the different stimuli. The natural story was found to be more difficult to understand than Matrix sentences. Although we controlled for the sex of the speaker and chose stimuli with similar speech rates and spectrum, the difference could still be due to different acoustic features such as for example prosody. The Matrix sentences namely are part of a standardized speech material where every word is spoken at the same intensity. The natural story, on the other hand, is narrated for children and has more variations. An additional reason to explain this difference is lexical prediction. Even though the permutations of the words are different in each Matrix sentence, the words themselves are all equally likely and familiar to the participants, in contrast to the natural story. Perhaps drawing from a larger pool of words for the Matrix sentences might have led to more similar intelligibility ratings between stimuli. Finally, speech intelligibility for both stimuli was measured in a different way: rating (natural story) versus recall (Matrix sentences). Similar to the very small and insignificant difference of 0.5 dB between rate and recall of Matrix sentences reported by Decruy et al. (2018), we did not find any statistical difference either between both measuring methods applied on the same Matrix sentences.

Besides the difference between the Matrix sentences and the natural story per participant, the variability between participants is also different per stimulus. Variability

is high for the natural story compared to the matrix sentences (figure 2). This difference could be explained by the fact that the Matrix sentences are a validated speech material created with the intention to have very low inter-subject variability to only reflect hearing performance independent of other individual skills. The natural story, on the other hand, is not controlled for this and is more dependent on individual skills such as for example linguistic knowledge, cognition and attention span.

4.2 Neural envelope tracking as an objective measure of speech intelligibility

We found that the correlation between the reconstructed and the acoustic envelope increased with speech intelligibility for both the Matrix sentences and the natural story. This supports the results of Luo and Poeppel (2007); Ding and Simon (2013); Ding et al. (2014); Molinaro and Lizarazu (2017); Vanthornhout et al. (2018) where an increase in speech intelligibility was also found to accompany an increase in envelope tracking and demonstrates that the objective measure of speech intelligibility using Matrix sentences by Vanthornhout et al. (2018) can be conducted with fully natural speech.

Next, the tracking results in the delta band were significantly higher than in the theta band while the significance levels remain the same, resulting in a steeper slope of envelope tracking as a function of speech intelligibility in the delta band. This difference in correlation magnitude between the frequency bands could be explained by the fact that the modulation spectrum of both stimuli has most energy in the delta band (Luo and Poeppel, 2007; Aiken and Picton, 2008).

When investigating the differences between both stimuli, we found that the use of

496 natural speech enhanced neural envelope tracking compared to Matrix sentences. This
497 suggests that neural envelope tracking might capture the interaction between the
498 incoming acoustic speech stream and top-down information (Hickok and Poeppel, 2007;
499 Gross et al., 2013) such as for example semantic processing (Di Liberto et al., 2018;
500 Broderick et al., 2018). A potential confound is that we used different SNRs for the two
501 stimulus types (to control for intelligibility). This means that the differences in envelope
502 tracking could be related simply to SNR rather than other stimulus properties. To
503 investigate this, we conducted the same analysis, but with SNR as predictor instead of
504 intelligibility, and again found significantly increased envelope tracking for the natural
505 story stimulus. This shows that SNR by itself does not account for the full difference
506 between the two stimulus types. A remarkable finding arising from this SNR analysis, is
507 that neural envelope tracking at a particular SNR is higher for the natural story
508 compared to the Matrix sentences, although SI is lower. This seems in contrast with the
509 hypothesis that neural envelope tracking relates to SI. We hypothesize that this
510 increase is present because the story stimulus elicits more brain activity (e.g., semantic
511 processing and working memory). This underlines the importance of always selecting 1
512 stimulus to investigate SI to eliminate inter-stimulus differences.

513 In addition, other confounding factors besides SNR could also be present. First,
514 although the acoustics of the stimuli were matched in terms of sex and speech rate of
515 the speaker and spectrum of the stimulus, acoustic differences like prosody are still
516 present, as discussed in section 4.1. Second, despite the questions asked to motivate
517 the participants, the reduced correlations for the Matrix sentences could be linked to

attention. Because listening to concatenated sentences can be boring, attention loss could occur which reduces neural envelope tracking (Ding and Simon, 2012; Kong et al., 2014; Petersen et al., 2017; Vanthornhout et al., 2019). For the natural story, on the other hand, attention could be less of an issue as attending this speech is entertaining possibly resulting in higher correlations.

4.3 The effect of noise and stimulus type on neural envelope tracking

In addition to envelope reconstruction to show the feasibility of natural speech as a stimulus for the objective measure (4.2), we conducted a TRF analysis. This analysis enables us to investigate the temporal characteristics of the brain responses over the entire scalp and removes the assumption that neural processing is similar for the decoder story and the test stimuli. The topographies in Figure 5 of both stimuli show a negative activation in the temporal-occipital channels and positive activation in the central channels. This is a typical topography of auditory evoked far-field potentials (Picton, 2011). The large negative peak within the 100 to 200 ms time lag (Figure 6) could be related to the N100, usually occurring at a latency between 70-150 ms (Picton, 2011).

4.3.1 Effect of SNR and speech intelligibility on TRFs

Generally we found, similar to envelope reconstruction, high TRF amplitudes over the entire scalp when speech intelligibility is high (SI=100%) and reduced amplitudes when speech intelligibility decreased for both stimuli, again showing feasibility of natural speech as a stimulus for the objective measure of speech intelligibility. Most remarkable

are the TRF amplitudes between 150 to 200 ms, which consistently decrease with decreasing speech intelligibility, perhaps indicating a time window sensitive to speech intelligibility. This amplitude decrease is similar to the behavior of the N1-P2 complex in function of SNR for tone- or syllable-induced event related potentials (ERPs) (Whiting et al., 1998; Billings et al., 2009). Another peculiarity are the noise induced topographic changes. When a small amount of noise is added and speech intelligibility remains almost unchanged from the no-noise condition (SNR=2.5 dB SNR; Matrix sentences: SI=99.9%; Natural story: SI=99.0%), TRF amplitudes across the entire scalp decrease between 0 to 150 ms, while amplitudes between 150 to 200 ms increase. Moreover, TRF amplitudes between 50 and 100 ms even switch polarities in the presence of noise. These increases in amplitudes could potentially be linked to enhanced top-down attention when listening to speech in noise (Fritz et al., 2007). Top-down attention is a selection process that steers neural focusing resources to the desired information stream (the clean speech in this case). This active top-down process causes changes in TRF amplitudes between 50 and 200 ms (Ding and Simon, 2012; Kong et al., 2014; Petersen et al., 2017; Vanthornhout et al., 2019). An additional remark about these noise induced TRF changes concerns the decoder used for envelope reconstruction. Although a decoder trained on clean speech is able to reconstruct the envelope of speech surrounded by noise reliably, it could be more optimal to train the decoder on speech in noise.

4.3.2 Effect of stimulus type on TRFs

Stimulus related differences can be found when comparing topography results

561 between both stimuli. TRF amplitudes are larger for the Matrix sentences in the central
562 and parieto-occipital channels compared to the natural story in the no-noise condition.
563 In the presence of background noise, even at a very high SNR, no significant difference
564 can be found anymore. A possible hypothesis could be the interaction between the
565 incoming acoustic speech stream and top-down information (Hickok and Poeppel, 2007;
566 Gross et al., 2013): In the no-noise condition Matrix sentences are mainly processed in
567 a feed-forward acoustical way. The enhanced TRF amplitudes could be caused by the
568 fixed syntactical 5-word structure of the Matrix sentences, resulting in a more rigid word
569 and sentence rate compared to the natural story. However, when noise is added, more
570 effort has to be paid to listen to the Matrix sentences (Wu et al., 2016; Houben et al.,
571 2013). This changes listening to the Matrix sentences from a bottom-up process to an
572 interactive bottom-up and top-down process similar to the natural story, diminishing the
573 differences between both stimuli.

574 Another stimulus related difference is the latency pattern over speech intelligibility.
575 The latency of the negative peak decreases with increasing speech intelligibility for the
576 Matrix sentences, while the latency remains unchanged for the natural story. A latency
577 decrease of N100 with increasing speech intelligibility, similar to the Matrix sentences,
578 has been reported in literature analyzing speech- (Petersen et al., 2017; Kong et al.,
579 2014), tone- (Billings et al., 2009) and syllable-induced ERPs (Whiting et al., 1998;
580 Kaplan-Neeman et al., 2006, but is not supported for TRFs for continuous speech by
581 Ding and Simon (2012). This different pattern between the Matrix sentences and the
582 natural story could be explained by two factors. (1) Top-down processing: This is

583 present for the natural story the entire time, for the Matrix sentences, on the other hand,
584 it increases with increasing noise level. Top-down processing requires more time, which
585 could result in delayed TRFs. (2) Attention: Listening to concatenated Matrix sentences
586 might be boring, especially when speech intelligibility decreases, which could result in
587 attention loss and less listening effort known to delay neural processing of speech (Ding
588 and Simon, 2012; Kong et al., 2014; Petersen et al., 2017; Vanthornhout et al., 2019).

589 A last result to point out is the positive peak around 300 ms for the Matrix sentences
590 at -9.5 dB SNR (SI=49%) (Figure 6). In ERPs a positive peak around this time lag is
591 known to occur when a participant tries to detect a target stimulus (Picton, 1992, 2011).
592 As the Matrix sentences do not contain semantic context, which makes content
593 questions not possible, counting questions were asked at every SNR trial, for example,
594 'Which colors of boats were mentioned?'. We hypothesize that the question type,
595 content questions for the natural story versus counting questions for the Matrix
596 sentences, accounts for this difference around 300ms. As a consequence, the type of
597 questions to ask is also an important factor to take into account for future research.

598 **4.4 Implications for applied research and potential clinical applications**

599 In this study we showed that the objective measure of speech intelligibility by
600 Vanthornhout et al. (2018) using Matrix sentences can also be conducted with natural
601 speech as a stimulus. Although neural tracking is enhanced using natural speech
602 instead of Matrix sentences, no significant differences in slope of neural tracking in
603 function of speech intelligibility are found. Therefore both stimuli are equally appropriate

604 to calculate the objective measure of speech intelligibility. The stimulus choice will have
605 to be considered based on the intended purpose of the measurement. For clinical
606 applications, for example, we should distinguish between applications in hearing aid
607 fitting and in general diagnostics of the auditory system. For hearing aid fitting, mainly
608 the peripheral processing of speech is of interest, so any type of natural speech is
609 appropriate. For diagnostics, a potential added benefit of a story is that it is closer to
610 everyday communication, and may relate better with subjective experience.

611 Before clinical application, however, several optimization and validation steps still
612 need to be undertaken: How can we reduce measuring time? How is test-retest
613 reliability within a subject over several test sessions? Also the optimal way of measuring
614 envelope tracking (decoder versus TRF) should be considered. Decoders are probably
615 better suited for clinical applications and TRFs for research. A decoder does not require
616 any channel selection or peak-picking, making it easy to use and interpret the results.
617 TRFs, on the other hand, reveal much more information which is useful for research
618 purposes, but less in the clinical field: Interpreting TRFs is time-consuming and can only
619 be done by highly trained personnel. A concern, however, when using decoders in the
620 clinic, is that subject-specific decoders require long measurement times. A solution
621 would be a generic decoder model which has already been shown to be successful (Di
622 Liberto and Lalor, 2017) and to not decrease performance (Lesenfants et al., 2019b).

623 **4.5 Limitations of this study**

Frequently recurring confounds throughout the discussion of this study are the attention paid by the listener and listening effort. We tried to control for this by asking the participants to focus and present them content questions after every trial. However, we cannot be sure they always paid attention. In future research it could be useful to also measure attention and listening effort using, e.g., pupillometry (Ohlenforst et al., 2017) or alpha power (Miles et al., 2017; Dimitrijevic et al., 2019). In addition, for the natural story, the order of presented SNRs could also have influenced the results. If, e.g., the higher noise levels were presented first, given that comprehension at the -12.5 SNR level is basically zero, the participant could lose the story line and have worse results in further trials compared to someone who listened to the easier noise levels in the beginning. To minimize this possible bias, we presented all SNRs in random order per participant and the first part of the natural story was always presented in quiet. A third limitation is that the presented Matrix sentences were repeated twice, while the natural story was not. However, we believe this was not a major confound as the Matrix sentences consist of a random combination of 5 word categories which make them very hard to remember. Finally, we rereferenced our data to a common average of the channels to not bias topography patterns relative to the chosen channel. A potential disadvantage of this approach is that the absolute TRF polarity is more difficult to interpret and to compare with previous studies using for example Cz- or mastoid rereferencing. This is, however, not a major concern in this study because we are interested in how TRF polarities and magnitudes of different conditions relate to each other and not to an absolute number.

4.6 Conclusion

647 We found increasing neural envelope tracking with increasing speech intelligibility for
648 both stimuli with an additional enhancement for natural speech compared to Matrix
649 sentences. These results show (1) the feasibility of natural speech as a stimulus for the
650 objective measure of speech intelligibility, (2) that neural envelope tracking is enhanced
651 using a story compared to Matrix sentences and (3) that noise and the stimulus type
652 can change the temporal characteristics of the brain responses.

653 **Acknowledgements**

654 The authors would like to thank Lien Decruy and Elien Vanluydt for their help in data
655 acquisition. All authors designed the experiment, contributed to the data analysis,
656 discussed the results and implications and commented on the manuscript at all stages.
657 E.V. performed the experiments and wrote the paper. This project has received funding
658 from the European Research Council (ERC) under the European Union's Horizon 2020
659 research and innovation programme (grant agreement No 637424 to Tom Francart).
660 Further support came from KU Leuven Special Research Fund under grant OT/14/119.
661 Research of Jonas Vanthornhout (1S10416N) and Eline Verschueren (1S86118N) is
662 funded by a PhD grant of the Research Foundation Flanders (FWO). The authors
663 declare no conflict of interest.

664 **REFERENCES**

- 665 Aiken, S. J. and Picton, T. W. (2008). Human Cortical Responses to the Speech
666 Envelope. *Ear & Hearing*, 29, 139–157.
- 667 Anderson, A. J., Broderick, M. P., and Lalor, E. C. (2018). Neuroscience: Great
668 Expectations at the Speech–Language Interface. *Current Biology*, 28, 1396–1398.
669 doi:10.1016/j.cub.2018.10.063
- 670 Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2017). Auditory-inspired speech
671 envelope extraction methods for improved EEG-based auditory attention detection in a
672 cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation*

673 *Engineering*, 25, 402–412. doi:10.1109/TNSRE.2016.2571900

674 Billings, C.J., Tremblay, K.L., Stecker, G.C. and Tolin, W.M. (2009). Human evoked
675 cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, 254,
676 15–24. doi:10.1016/j.heares.2009.04.002

677 Brodbeck, C. (2017). Eelbrain: 0.25. Zenodo. doi:10.5281/zenodo.1186450

678 Brodbeck, C., Presacco, A., and Simon, J. Z. (2018). Neural source dynamics of brain
679 responses to continuous stimuli: Speech processing from acoustics to comprehension.
680 *NeuroImage*, 172, 162–174. doi:10.1016/j.neuroimage.2018.01.042

681 Broderick, M. P., Anderson, A. J., Liberto, G. M. D., et al. (2018). Electrophysiological
682 correlates of semantic dissimilarity reflect the comprehension of natural , narrative
683 speech. *Current Biology*, 28, 803–809.

684 David, S. V., Mesgarani, N., and Shamma, S. A. (2007). Estimating sparse spectro-
685 temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*,
686 18, 191–212. doi:10.1080/09548980701609235

687 Decruey, L., Das, N., Verschueren, E., and Francart, T. (2018). The self-assessed Beksy
688 procedure: validation of a method to measure intelligibility of connected discourse.
689 *Trends in Hearing*, 22, 1–13. doi:10.1177/2331216518802702

690 Di Liberto, G. M. and Lalor, E. C. (2017). Indexing cortical entrainment to natural speech
691 at the phonemic level: Methodological considerations for applied research. *Hearing*

692 research, 348, 70-77. doi:10.1016/j.heares.2017.02.015

693 Di Liberto, G. M., Lalor, E. C., and Millman, R. E. (2018). Causal cortical dynamics of a
694 predictive enhancement of speech intelligibility. *NeuroImage*, 166, 247–258.
695 doi:10.1016/j.neuroimage.2017.10.066

696 Dimitrijevic, A., Smith, M.L., Kadis, D.S., & Moore, D.R. (2019) Neural indices of listening
697 effort in noisy environments. *Scientific Reports*, 9, 1–10.

698 Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the
699 speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, 88, 41–46.
700 doi:10.1016/j.neuroimage.2013.10.054

701 Ding, N. and Simon, J. Z. (2011). Neural coding of continuous speech in auditory cortex
702 during monaural and dichotic listening. *Journal of Neurophysiology*, 107, 78–89.
703 doi:10.1152/jn.00297.2011

704 Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while
705 listening to competing speakers. *Proceedings of the National Academy of Sciences of*
706 *the United States of America*, 109, 11854–9. doi:10.1073/pnas.1205381109

707 Ding, N. and Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-
708 Insensitive Cortical Representation of Speech. *Journal of Neuroscience*, 33, 5728–
709 5735. doi:10.1523/JNEUROSCI.5297-12.2013

710 Francart, T., van Wieringen, A., and Wouters, J. (2008). APEX 3: a multi-purpose test

711 platform for auditory psychophysical experiments. *Journal of Neuroscience Methods*,
712 172, 283–293. doi:10.1016/j.jneumeth.2008.04.020

713 Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007). Auditory attention—
714 focusing the searchlight on sound. *Curr. Opin. Neurobiol.*, 17, 437–455.
715 doi:10.1016/j.conb.2007.07.011

716 Gross, J., Hoogenboom, N., Thut, G., et al. (2013). Speech rhythms and multiplexed
717 oscillatory sensory coding in the human brain. *PLoS biology*, 11, e1001752.
718 doi:10.1371/journal.pbio.1001752

719 Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing.
720 *Nature reviews. Neuroscience*, 8, 393–402. doi:10.1038/nrn2113

721 Houben, R., van Doorn-Bierman, M. and Dreschler, W.A. (2013). Using response time to
722 speech as a measure for listening effort, *International Journal of Audiology*, 52:11, 753-
723 761. doi:10.3109/14992027.2013.832415

724 Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., Muchnik, C. (2006). Identification of
725 syllables in noise: electrophysiological and behavioral correlates. *J. Acoust. Soc. Am.*,
726 120 (2), 926–933.

727 Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing
728 cortical speech representations in a "cocktail party". *The Journal of neuroscience : the*
729 *official journal of the Society for Neuroscience*, 30, 620–8. doi:10.1523/JNEURO
730 SCI.3631-09.2010

731 Kong, Y.-Y., Mullangi, A., and Ding, N. (2014). Differential modulation of auditory
 732 responses to attended and unattended speech in different listening conditions. *Hearing*
 733 *Research*, 0, 73–81. doi:10.1002/ana.22528.Toll-like

734 Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., et al. (2006). The VESPA: A method for the
 735 rapid estimation of a visual evoked potential. *NeuroImage*, 32, 1549–1561.
 736 doi:10.1016/j.neuroimage.2006.05.054

737 Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving Precise
 738 Temporal Processing Properties of the Auditory System Using Continuous Stimuli.
 739 *Journal of Neurophysiology*, 102, 349–359. doi:10.1152/jn.90896.2008

740 Lesenfans, D., Vanthornhout, J., Verschueren, E., et al. (2019a). Predicting individual
 741 speech intelligibility from the neural tracking of acoustic- and phonetic-level speech
 742 representations. *Hearing Research*, 380, 1-9. doi:10.1016/j.heares.2019.05.006

743 Lesenfans, D., Vanthornhout, J., Verschueren, E. and Francart, T. (2019b). Data-driven
 744 spatial filtering for improved measurement of cortical tracking of multiple representations
 745 of speech. *Journal of neural engineering*, 16(6):066017. doi:10.1088/1741-2552/ab3c92

746 Luo, H. and Poeppel, D. (2007). Phase patterns of neuronal responses reliably
 747 discriminate speech in human auditory cortex. *Neuron*, 54, 1001–10.
 748 doi:10.1016/j.neuron.2007.06.004

749 Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-
 750 data. *Journal of Neuroscience Methods*, 164, 177–190.

751 doi:10.1016/j.jneumeth.2007.03.024

752 Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended
753 speaker in multi-talker speech perception. *Nature*, 485, 233–6.
754 doi:10.1038/nature11020

755 Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., Lissa, P. de, Graham, P., & Lyxell, B.
756 (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and
757 EEG. *Trends in hearing*, 21, 1–13.

758 Molinaro, N. and Lizarazu, M. (2017). Delta(but not theta)-band cortical entrainment
759 involves speech-specific processing. *European Journal of Neuroscience*, 9, 1–9.
760 doi:10.1111/ejn.13811

761 Obleser, J. and Kotz, S. A. (2011). Multiple brain signatures of integration in the
762 comprehension of degraded speech. *NeuroImage*, 55, 713–723.
763 doi:10.1016/j.neuroimage.2010.12.020

764 Ohlenforst, B., Zekveld, A.A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N.J.,
765 & Kramer, S.E. (2017). Impact of stimulus-related factors and hearing impairment on
766 listening effort as indicated by pupil dilation. *Hearing Research*, 351, 68–79.

767 Pernet, C. R., Wilcox, R. and Rousselet, G. A. (2012). Robust Correlation Analyses:
768 False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Front*
769 *Psychol.*, 3:606. doi: 10.3389/fpsyg.2012.00606.

770 Petersen, E. B., Wostmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of
 771 attended versus ignored speech is differentially affected by hearing loss. *Journal of*
 772 *Neurophysiology*, 117, 18–27. doi:10.1152/jn.00527.2016

773 Picton, T. W. (1992). The P300 Wave of the Human Event-Related Potential. *Journal of*
 774 *clinical Neurophysiology*, 9, 456–479

775 Picton, T. W. (2011). *Human Auditory Evoked Potentials* (San Diego: Plural Publishing
 776 inc.)

777 Shannon, R. V., Zeng, F.-G., Kamath, V., et al. (1995). Speech Recognition with Primarily
 778 Temporal Cues. *Science*, 270, 303–304. doi:10.1126/science.270.5234.303

779 Somers, B., Francart, T., and Bertrand, A. (2018). A generic EEG artifact removal
 780 algorithm based on the multi-channel Wiener filter. *Journal of neural engineering*, 15.
 781 doi:10.1088/1741-2552/aaac92

782 Vanthornhout, J., Decruy, L., and Francart, T. (2019). Effect of task and attention on
 783 neural tracking of speech. *Frontiers in Neuroscience*, 13, 977.
 784 doi:10.3389/fnins.2019.00977

785 Vanthornhout, J., Decruy, L., Wouters, J., and Francart, T. (2018). Speech intelligibility
 786 predicted from neural entrainment of the speech envelope. *JARO*, 19, 181–191.
 787 doi:10.1007/s10162-018-0654-z

788 Whiting, K.A., Martin, B.A., Stapells, D.R. (1998). The effects of broad-band noise

789 masking on cortical event-related potentials to speech sounds /ba/ and /da. *Ear &*
790 *Hearing*, 19 (3), 218–231.

791 Wu, Y.H., Stangl, E., Zhang, X., Perkins, J., & Eilers, E. (2016). Psychometric Functions
792 of Dual-Task Paradigms for Measuring Listening Effort. *Ear & Hearing*, 37, 660–670.

Accepted version

793 **Figure legends**

794 **Figure 1.** Overview of the experimental setup using the linear decoder analysis. We
795 presented the Matrix sentences and a natural story at different Signal-to-Noise Ratio's
796 (SNR). Participants listened to the speech while their EEG was measured. To obtain a
797 measure of neural envelope tracking we correlated the reconstructed envelope with the
798 acoustic envelope after band-pass filtering (BP filter). We compared the envelope
799 tracking results with the behavioral speech intelligibility (SI) scores.

800 **Figure 2.** A comparison between the Matrix sentences and the natural story reveals that
801 the story is more difficult to understand when adding background noise.

802 **Figure 3.** Neural envelope tracking increases with increasing speech intelligibility and
803 by using natural speech as a stimulus. The shading represents two times the standard
804 error of the fit and the dotted line is the significance level of the correlation (± 0.019).

805 **Figure 4.** Electrode selection: 64 active electrodes placed according to the 10-20
806 electrode system. The locations of the electrodes that were selected for the calculation
807 of the occipital-temporal TRF are indicated in red.

808 **Figure 5.** Topographies for the natural story and the Matrix sentences at different SNRs
809 and different time lags varying from 0 until 200 ms. Significant differences between the
810 Matrix sentences and the natural story are highlighted in red.

811 **Figure 6.** Time-course of the temporal-occipital TRFs over participants for the Matrix

812 sentences and the natural story. TRF samples significantly different from zero are
813 highlighted in bold.

814 **Figure 7.** Latency and amplitude of the negative peak of the temporal-occipital TRF
815 between 50 and 300 ms per participant over speech intelligibility.

Accepted version

Figure 1. Overview of the experimental setup using the linear decoder analysis. We presented the Matrix sentences and a story at different Signal-to-Noise Ratio's (SNR). Participants listened to the speech while their EEG was measured. To obtain a measure of neural envelope tracking we correlated the reconstructed envelope with the acoustic envelope after band-pass filtering (BP filter). We compared the envelope tracking results with the behavioral speech intelligibility (SI) scores.

Figure 2. A comparison between the Matrix sentences and the story reveals that the story is more difficult to understand when adding background noise.

Figure 3. Neural envelope tracking increases with increasing speech intelligibility and by using natural speech as a stimulus. The shading represents two times the standard error of the fit and the dotted line is the significance level of the correlation (± 0.019).

Figure 4. Electrode selection: 64 active electrodes placed according to the 10-20 electrode system. The locations of the electrodes that were selected for the calculation of the occipital-temporal TRF are indicated in red.

Figure 5. Topographies for the story and the Matrix sentences at different SNRs and different time lags varying from 0 until 200 ms. Significant differences between the Matrix sentences and the story are highlighted in red.

Figure 6. Time-course of the temporal-occipital TRFs over participants for the Matrix sentences and the story. TRF samples significantly different from zero are highlighted in bold.

Figure 7. Latency and amplitude of the negative peak of the temporal-occipital TRF between 50 and 300 ms per participant over speech intelligibility.