

Neural envelope tracking as a measure of speech understanding in cochlear implant users

Eline Verschueren^{a,*}, Ben Somers^a and Tom Francart^a

^aResearch Group Experimental Oto-rhino-laryngology (ExpORL), Department of Neurosciences, KU Leuven - University of Leuven, 3000 Leuven, Belgium

Correspondence*:

Eline Verschueren

Herestraat 49, bus 721, 3000 Leuven, Belgium

eline.verschueren@kuleuven.be

1 Declarations of interest: none

2 **ABSTRACT**

3 The speech envelope is essential for speech understanding and can be reconstructed from the
4 electroencephalogram (EEG) recorded while listening to running speech. This so-called neural
5 envelope tracking has been shown to relate to speech understanding in normal hearing listeners,
6 but has barely been investigated in persons wearing cochlear implants (CI). We investigated the
7 relation between speech understanding and neural envelope tracking in CI users.

8 EEG was recorded in 8 CI users while they listened to a story. Speech understanding was varied
9 by changing the intensity of the presented speech. The speech envelope was reconstructed from
10 the EEG using a linear decoder and then correlated with the envelope of the speech stimulus as
11 a measure of neural envelope tracking which was compared to actual speech understanding.

12 This study showed that neural envelope tracking increased with increasing speech
13 understanding in every participant. Furthermore behaviorally measured speech understanding
14 was correlated with participant specific neural envelope tracking results indicating the potential
15 of neural envelope tracking as an objective measure of speech understanding in CI users. This
16 could enable objective and automatic fitting of CIs and pave the way towards closed-loop CIs that
17 adjust continuously and automatically to individual CI users.

18 **Keywords:** cochlear implant; artifact rejection; neural decoding; natural speech; EEG; speech understanding

1 INTRODUCTION

19 Speech is characterized by fast and slow modulations. The slow modulations are also called the envelope
20 of speech, reflecting the different syllable, word and sentence boundaries known to be essential for
21 speech understanding (Shannon et al., 1995). Previous studies have shown that the brain tracks the speech
22 envelope and that it is possible to reconstruct the envelope from brain responses in normal hearing listeners
23 using electroencephalography (EEG) or magnetoencephalography (Aiken and Picton, 2008; Luo and
24 Poeppel, 2007; Ding and Simon, 2011; Ding et al., 2015; Meyer et al., 2017). The correlation between
25 this reconstructed envelope and the real speech envelope reflects a measure of neural envelope tracking.
26 Recently, researchers were able to establish a link between increasing neural envelope tracking and
27 increasing speech understanding using speech versus non-speech stimuli (Molinaro and Lizarazu, 2017),
28 priming and vocoders (Di Liberto et al., 2018) or by adding background noise to the speech signal (Ding
29 and Simon, 2013; Ding et al., 2014; Vanthornhout et al., 2018), underlining the application potential of
30 neural envelope tracking as an objective measure of speech understanding.

31 Besides the promising results in normal hearing listeners, neural envelope tracking has been measured in
32 listeners with a hearing impairment by Petersen et al. (2017). They showed that the amount of hearing loss
33 could be related to neural tracking of the to-be-ignored speech, diminishing the difference between the
34 attended and unattended speech stream in persons with increasing hearing loss.

35 Despite encouraging results in both normal hearing and hearing impaired listeners, for cochlear implant
36 (CI) users, the link between neural envelope tracking and speech understanding has not been established.
37 A CI is an implanted hearing aid that gives severely hearing-impaired to deaf persons the opportunity to
38 (re)gain access to sound through electrical stimulation of the auditory nerve (Loizou, 1998; Wouters et al.,
39 2015). Numerous studies have demonstrated good speech understanding for CI users in quiet, some even
40 similar to normal hearing participants. However, when noise is added, speech understanding drops far
41 below scores of normal hearing listeners and a large variation is seen between CI users (van Wieringen and
42 Wouters, 2008; Shannon et al., 2011; Loizou et al., 2000). To better characterize and eventually overcome
43 these speech understanding problems with the help of objective and automatic fitting of the CI, an objective
44 way to measure speech understanding is of great interest. For this reason it would be useful to investigate
45 neural envelope tracking, a potential objective measure of speech understanding, in CI users. In addition,
46 CI users have to rely almost entirely on slow temporal information, i.e., the speech envelope, to understand
47 speech. Therefore CI users seem excellent candidates for using neural envelope tracking.

48 However, a mayor problem when measuring neural envelope tracking in CI users are the stimulation
49 artifacts in the EEG, e.g., Somers et al. (2018b); Deprez et al. (2017); Hoffman and Wouters (2010). When
50 a CI electrically stimulates the auditory nerve, large electric potentials from the CI appear in the EEG. Just
51 like the brain responses, the modulations in these CI artifacts follow the envelope of the presented speech
52 signal. As a consequence it is difficult to distinguish between the desired brain response and the unwanted
53 artifact, leading to false positives. Recently a novel artifact removal method has been developed by Somers

54 et al. (2018b) which works by leaving out small groups of stimulation pulses during the presentation of a
55 speech signal. This creates short artifact-free EEG windows during which the ongoing brain responses can
56 be measured. This method was validated in CI users.

57 In the current study we investigated the effect of speech understanding on neural envelope tracking using
58 EEG in CI users by applying the newly proposed artifact removal method by Somers et al. (2018b). We
59 hypothesized that neural envelope tracking will increase with speech understanding.

2 MATERIAL AND METHODS

60 2.1 Participants

61 Eight participants aged between 46 and 75 years took part in the experiment after providing informed
62 consent. Participants had Flemish as their mother tongue and were all experienced CI users (>8 months)
63 using Cochlear Ltd. devices. None of them experienced tinnitus, except for S4 when the CI is off. Relevant
64 details are shown in table 1. The study was approved by the Medical Ethics Committee UZ Leuven /
65 Research (KU Leuven) with reference S57102.

Table 1. Relevant participant details

Name	Age (years)	Gender	CI experience	CI side	CI device	Dynamic range*	Pulse width
S1	58	F	1.1	R	CI522	65	37
S2	75	M	1.3	L	CI522	61	37
S3	73	F	0.7	L	CI522	67	37
S4	61	F	1.1	L	CI522	40	37
S5	58	M	2.0	L	CI522	78	37
S6	71	M	4.7	L	CI24RE	53	50
S7	46	F	6.2	R	CI24RE	33	25
S8	56	F	1.8	R	CI512	62	37

66 * Dynamic range is the average of all electrode specific dynamic ranges per participant.

67 2.2 Preprocessing of the stimuli

68 All stimuli were presented directly to the participant's CI using a research speech processor (L34)
69 provided by Cochlear Ltd., in combination with APEX 3 software (Francart et al., 2008) and the Nucleus
70 Implant Communicator (NIC). To enable this direct stimulation, we converted audio files into sequences of
71 electrical stimulation pulses using the Nucleus Matlab Toolbox (NMT) (Swanson and Mauch, 2006) in
72 MATLAB (version R2016b) which simulates the signal processing done by a clinical CI using the ACE
73 strategy. Such a sequence is obtained by splitting the audio waveform into frequency bands which are
74 mapped to 22 electrodes covering a frequency range from 188 to 7938 Hz. Next, an electric pulse train

75 is modulated with the envelope of the signal in each of these frequency bands. Finally, every modulated
76 pulse train is sent to one of the 22 electrodes implanted inside the cochlea to stimulate the auditory nerve
77 and enable hearing. All electrical stimulation is programmed based on the participant's clinical settings.
78 More concrete, the lowest audible stimulation level (threshold (T)) and the loudest stimulation level that
79 can be comfortably listened to (C) are set per participant per electrode. Next, the amplitudes of the pulse
80 sequences in each channel are mapped to electric currents, i.e., current units, that evoke sound sensations
81 that are just between T and C level. Other parameter settings are pulse rate (fixed at 900 pulses per second),
82 stimulation mode (MP1+2 = monopolar), number of active electrodes (22 electrodes, except for participant
83 S1: 18 electrodes) and pulse width (see table 1). To vary speech understanding we varied the stimulation
84 levels by shifting all stimulation levels between T and C level by an equal number of current units. In
85 this way we maintain the dynamic range of the speech envelope but vary the stimulation levels and thus
86 audibility and consequently speech understanding (figure 1).

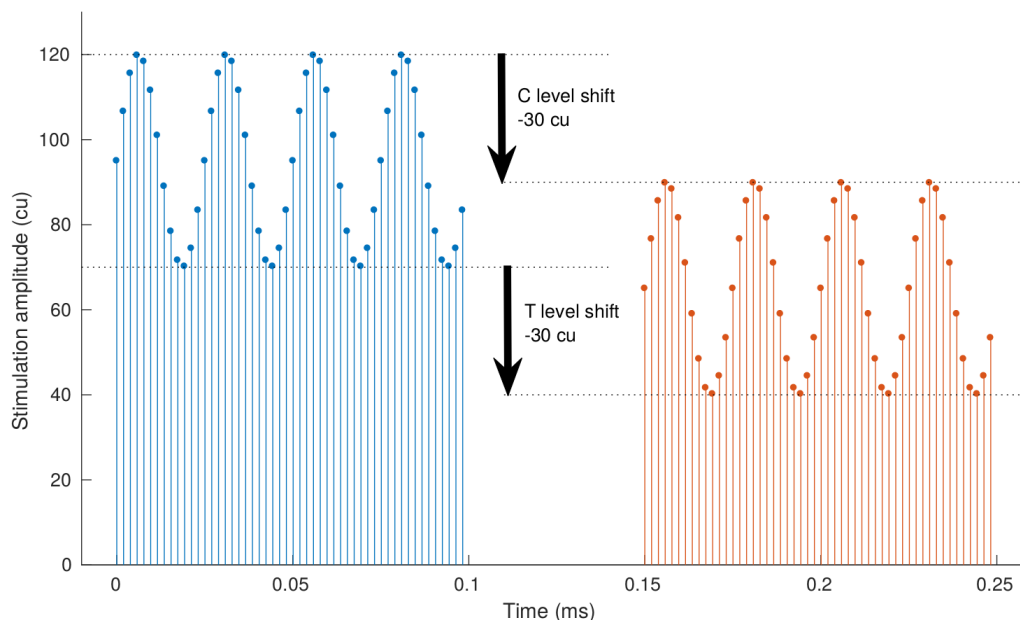


Figure 1. Illustration of a stimulation level shift for a sinusoidally modulated pulse train. A stimulation level shift is obtained by varying all the stimulation levels between the lowest audible (threshold (T)) and most comfortable level (C) with a fixed amount of current units (cu). Shifting the stimulation levels of the CI varies speech understanding while maintaining the dynamic range of the speech envelope.

87 In addition, to tackle the problem of CI stimulation artifacts when using EEG, we applied a new artifact
88 removal method proposed and validated by Somers et al. (2018b). In brief, this method obtains EEG free
89 of stimulation artifacts by periodically interrupting the electrical stimulation by leaving out small groups of
90 stimulation pulses as shown in figure 2 and further referred to as a stimulus with 'dropped pulses'. Within
91 these interruptions artifact-free EEG can be sampled. The stimulus interruptions were 4 ms long at a rate of

92 40 Hz. This is short enough to preserve speech understanding and long enough to remove the artifact. Only
93 the EEG within the stimulation gaps, i.e., artifact free EEG, is further analyzed.

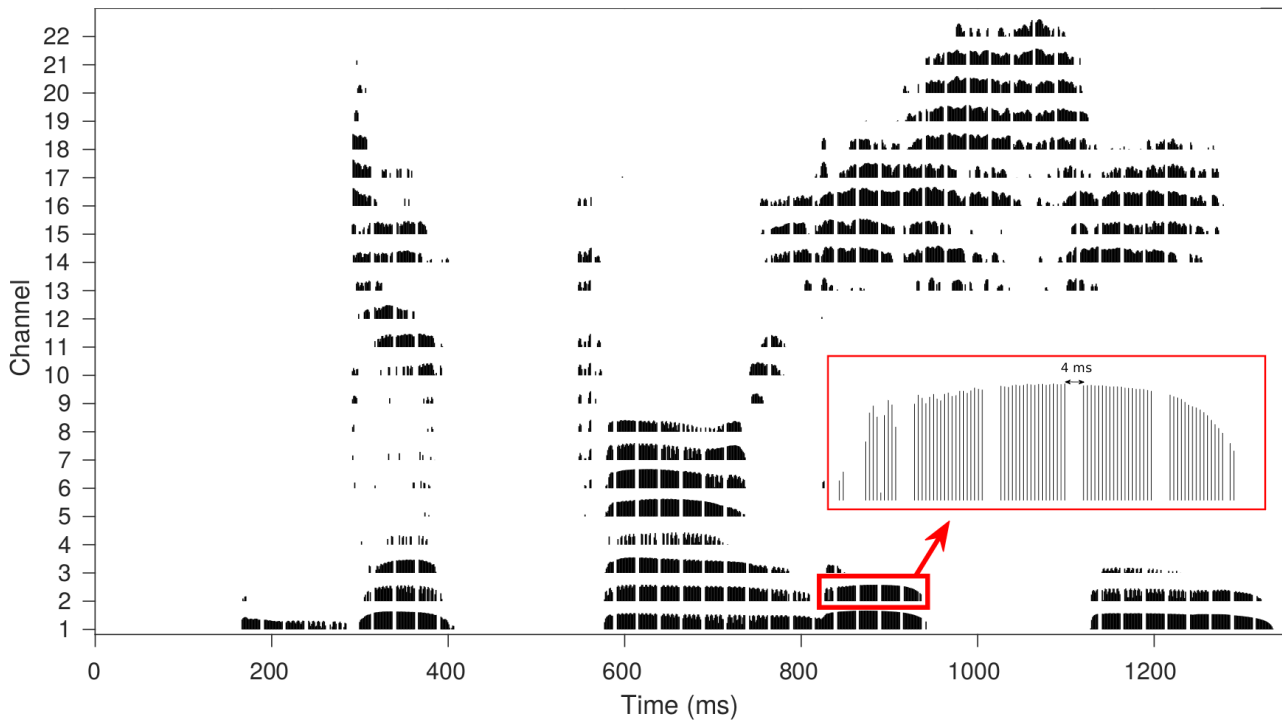


Figure 2. Illustration of gap insertion in the electrodiagram for the Flemish word ‘politie’ (police). Every channel contains an electric pulse train modulated with the speech envelope of the corresponding frequency band. The periodically inserted gaps to obtain artifact free EEG are 4ms long at a rate of 40 Hz.

94 2.3 Behavioral experiment

95 Before the start of the experiment, we checked whether the participant felt comfortable listening to the
96 stimulus with dropped pulses by presenting them with a sentence including stimulation gaps of 4 ms at a
97 rate of 40 Hz. These values were chosen based on the findings of Somers et al. (2018b). If the participant
98 could not tolerate the distortions induced by the gaps, we decreased the gap rate and/or length until a
99 comfortable level was reached.

100 After selecting the optimal gap parameters, speech understanding was measured behaviorally in order
101 to compare neural envelope tracking results with actual speech understanding. We used the Leuven
102 intelligibility sentence test (LIST) (van Wieringen and Wouters, 2008), uttered in Flemish by Wivine
103 Decoster (female speaker). This speech material is suitable to measure speech understanding in CI users
104 or persons with a severe hearing loss due to the low speech rate (2.5 syllables/second) and the key word
105 scoring. Every participant started with 1 training list at an input level equivalent to an acoustic signal of 60
106 dB SPL at the CI microphone, which will be referred to in the sequel as the baseline settings. Thereafter a
107 number of lists, each containing 10 sentences, were presented at different stimulation level shifts to vary

108 speech understanding. Stimulation levels were only shifted to lower values, as stimulating current units
109 (cu) above C levels would make speech too loud and could harm the participant. Stimulation level shifts
110 varied from 0 cu (i.e., baseline settings) to -50 cu (almost inaudible as C levels shift under original T levels
111 for most participants depending on their dynamic range). Participants had to recall the sentence they heard.
112 By counting the correctly recalled key words, a percentage correct per presented stimulation level shift
113 was calculated. Additionally a list without stimulation gaps was presented at baseline settings to check the
114 influence of inserting gaps on speech understanding.

115 2.4 EEG experiment

116 Recordings were made in a soundproof and electromagnetically shielded room. A 64-channel BioSemi
117 ActiveTwo EEG recording system was used at a sample rate of 16384 Hz. Participants sat in a comfortable
118 chair and were asked to move as little as possible during the recordings. The stories used during the EEG
119 measurement were the stories 'Marfoesjka en de Vorst' (translated Russian fairytale) and 'Luna van de
120 boom' (Bart Moeyaert), both narrated in Flemish by Wivine Decoster, the speaker of the LIST sentences.
121 The stories had a total length of 48 minutes. The stories were cut in different parts: 3 longer parts of ± 8
122 minutes and 10 parts of ± 2.4 minutes which were presented in chronological order. The long parts were
123 presented at baseline settings and used to train the decoder on. The short parts were presented at stimulation
124 level shifts from a fixed list in random order, containing 0cu, -5cu -10cu, -15cu, -20cu, -30cu, -40cu, to
125 vary speech understanding as shown in the experiment overview in figure 3. Every stimulation level shift
126 was applied twice, on a different part of the story, to analyze test-retest reliability.

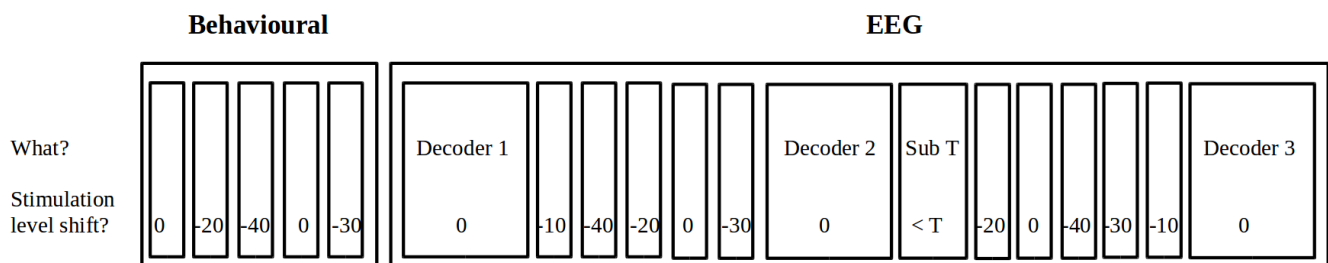


Figure 3. Overview of an example test session starting with the behavioral part where we measured speech understanding in a standardized way with the LIST sentences. In the second part, the EEG part, 2 stories were presented containing 3 blocks of 8 minutes without shift to train the decoder on and 10 blocks of 2.4 minutes at stimulation level shifts from a fixed list in random order, containing 0cu, -5cu -10cu, -15cu, -20cu, -30cu, -40cu, to vary speech understanding. In the middle a 5 minute part of the story was presented sub threshold (inaudible).

127 Additionally, we presented 5 minutes of the story below the participant's T level (referred to as sub-T),
128 meaning all stimulation was lower than the lowest stimulation level audible for that participant, which
129 resulted in no sound sensation. In the sub-T condition there are CI stimulation artifacts that are highly

130 correlated with the stimulus envelope (the CI does stimulate) in the absence of neural responses (the
131 participant cannot hear any sound). This condition can be used to check if CI artifacts are effectively
132 removed. If a significant neural envelope tracking response is found in this condition, this is caused by
133 remaining CI artifacts. If no significant response is found in this condition, we can assume that the CI
134 artifacts are removed.

135 To maximize the participants attention, content questions were asked after each part of the story. To
136 measure speech intelligibility of the story, we could not ask the participant to recall every sentence similar
137 to the behavioral LIST experiment. Therefore we used a rating method where the participants were asked
138 to rate their speech understanding on a scale from 0 to 100% following the question ‘Which percentage of
139 the story did you understand?’. After each trial a summary of the story presented in the previous trial was
140 shown on the screen in front of the participant to ensure comprehension of the storyline and keep them
141 motivated.

142 **2.5 Signal processing**

143 A measure of neural envelope tracking was calculated by correlating the stimulus envelope with the
144 envelope reconstructed from the EEG. All signal processing was done in MATLAB (version R2016b).

145 **2.5.1 Stimulus envelope**

146 In studies evaluating neural envelope tracking in acoustic hearing, the acoustic stimulus envelope is
147 chosen as the reference envelope. In this study, investigating electrical hearing, using the electrical envelope
148 as a reference is more appropriate (Somers et al., 2018b). The electrical envelope namely takes into account
149 all the preprocessing done by the CI before the speech reaches the cochlea, making this a more realistic
150 reference. The electric speech envelope was extracted from the stimulus by combining the magnitudes
151 of the electrical pulses over all stimulation channels. As a next step, the electric speech envelope was
152 band-pass filtered in the delta (0.5-4 Hz) frequency band.

153 **2.5.2 Envelope reconstruction**

154 As a first step CI artifacts were removed by retaining only samples within the stimulus gaps as mentioned
155 in paragraph 2.2. Other common EEG artifacts such as eye blink and muscle artifacts were removed using a
156 multi-channel Wiener filter (Somers et al., 2018a). After artifact rejection, the signal was bandpass filtered,
157 similar to the electric speech envelope. The mTRF Matlab toolbox (Lalor et al., 2006, 2009) was used to
158 compute the linear decoder which reconstructs the speech envelope from the EEG recordings. As speech
159 elicits neural responses with some delay, the linear decoder combines EEG channels and their time shifted
160 versions to optimally reconstruct the speech envelope. If g is the linear decoder and R the shifted neural
161 data, the reconstruction of the speech envelope $\hat{s}(t)$ was obtained by $\hat{s}(t) = \sum_n \sum_\tau g(n, \tau) R(t + \tau, n)$
162 with t the time ranging from 0 to T , n the recording electrode ranging from 1 to N and τ the number of
163 post-stimulus samples used to reconstruct the envelope. The decoder was calculated using ridge regression
164 by solving $g = (RR^T)^{-1}(RS^T)$ with S the speech envelope. As we used an integration window from 0

165 until 250 ms post-stimulus, the decoder matrix g was a 64 (EEG channels) x 11 (time delays) matrix. The
166 decoder was created using a combination of the long segments (3 x 8 min), not including the short 2.4
167 minute trials.

168 To investigate neural envelope tracking we did two different analyses. First, to check whether we managed
169 to eliminate the artifact from the EEG data, we did a leave-one-out cross-validation on the 24 minutes
170 of speech at baseline settings for every sample inside the inserted stimulation gap. We hypothesized that
171 correlations calculated for samples that contain artifacts will be higher as the artifact resembles the signal
172 of interest. When using samples further in the gap, where the artifact has died out, correlations will be
173 smaller, similar to correlations of previous studies with acoustic listeners, only containing the possible
174 neural response. This leave-one-out cross-validation was done by splitting the data in equal parts of 2
175 minutes. The first part (2 minutes) was selected as the testing data, while the rest was concatenated to
176 create one decoder used to reconstruct the envelope for the testing part. Next, the second part was selected
177 as the testing data and another decoder was trained on the remaining parts et cetera. After checking if the
178 samples were artifact-free, we concatenated the 24 minutes of EEG data at baseline settings to create one
179 decoder per participant to apply on the 10 short segments with different stimulation level shifts, resulting
180 in 10 outcome correlation measures per participant, i.e., 1 per trial.

181 **2.6 Statistical Analysis**

182 Statistical analysis was performed using R (version 3.3.2) software. The significance level was set at
183 $\alpha=0.05$ unless otherwise stated.

184 To analyze the behavioral results and test-retest for envelope reconstruction we used the nonparametric
185 Wilcoxon signed-rank test for dependent samples.

186 To investigate the relationship between the shift in stimulation levels and neural envelope tracking
187 or speech understanding per participant, a linear mixed effect (LME) model was constructed of neural
188 envelope tracking/speech understanding in function of stimulation level shift (continuous variable) with a
189 random slope and intercept per participant. To check if every chosen effect benefited the model the Akaike
190 Information Criterion (AIC) was calculated. The model with the lowest AIC was selected and its residual
191 plot was analyzed to assess the normality assumption of the LME residuals. Degrees of freedom (df),
192 t-values, and p-values are reported in the results section.

193 To calculate the correlation between envelope reconstruction and the speech reception threshold (SRT)
194 we used a Pearson's correlation.

3 RESULTS

195 **3.1 Behavioral speech intelligibility**

196 For 7 out of 8 participants, the gap settings of 4 ms and 40 Hz were chosen. One participant (S7) reported
197 that the speech was too distorted, a gap length of 3 ms was chosen for this participant. All participants

198 reported that the speech with dropped pulses sounded more robotic, but speech understanding was not
199 affected. We checked this by comparing the recall scores for LIST sentences with and without dropped
200 pulses at baseline settings and found no significant difference between the two ($p=0.7352$, $CI(95\%) = [-$
201 13.00% ; 6.00%], $n=7$, Wilcoxon signed-rank test). In the following analysis S8 is included to investigate the
202 artifacts, but excluded to investigate the possible link of neural envelope tracking with speech understanding
203 as she participated in the pilot study and only listened to the stories at baseline settings.

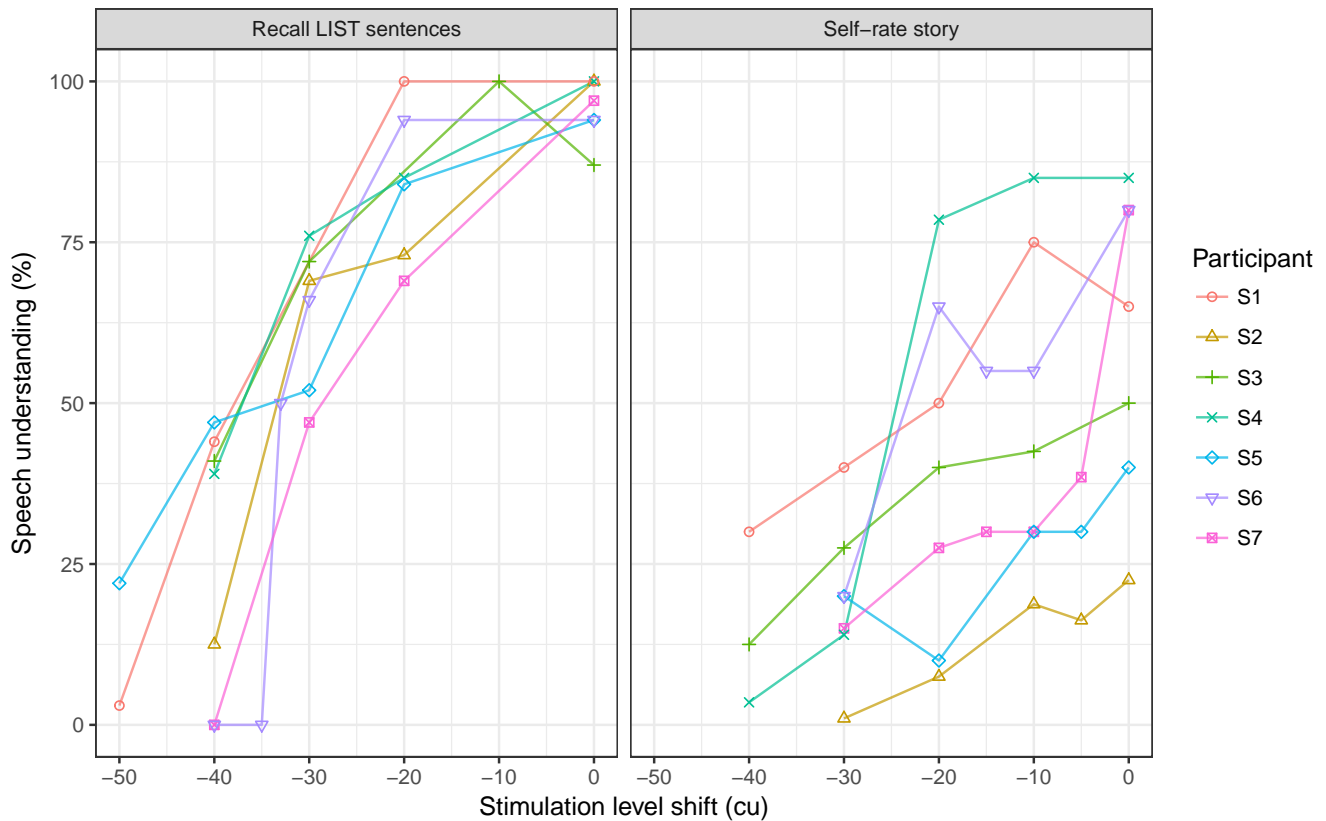


Figure 4. Speech understanding decreases with increasing stimulation level shift. More variation between participants is present for self-rated speech understanding of the story compared to the recalled scores of the LIST sentences.

204 To investigate the effect of shifting stimulation levels on speech understanding we presented LIST
205 sentences at different stimulation levels and asked the participants to recall the sentences. Figure 4 shows
206 that speech understanding decreases when stimulation levels decrease for the LIST sentences (fixed
207 effect stimulation level shift, $df = 23$, $t = 8.74$, $p < 0.0001$, LME). During the EEG a story was presented
208 and we asked the participants to rate their speech understanding. Similar to the LIST sentences, speech
209 understanding for the story also decreased with decreasing stimulation levels (fixed effect stimulation
210 level shift, $df = 28$, $t = 5.11$, $p < 0.0001$, LME). The variation between participants was larger for the story
211 (self-rated) than for the LIST sentences (recall) as shown in figure 4.

212 3.2 Neural envelope tracking

213 3.2.1 Influence of CI stimulation artifacts on neural envelope tracking

214 To investigate the presence of CI stimulation artifacts, we did two types of analysis. First we investigated
215 the magnitude of the correlation between the real and the reconstructed envelope for each sample inside
216 the stimulation gap. We hypothesized that correlations calculated for samples that contain artifacts will be
217 especially high as the artifact resembles the signal of interest. When using samples where the artifact has
218 died out, correlations will be smaller, similar to correlations of previous studies with acoustic listeners,
219 only containing the possible neural response.

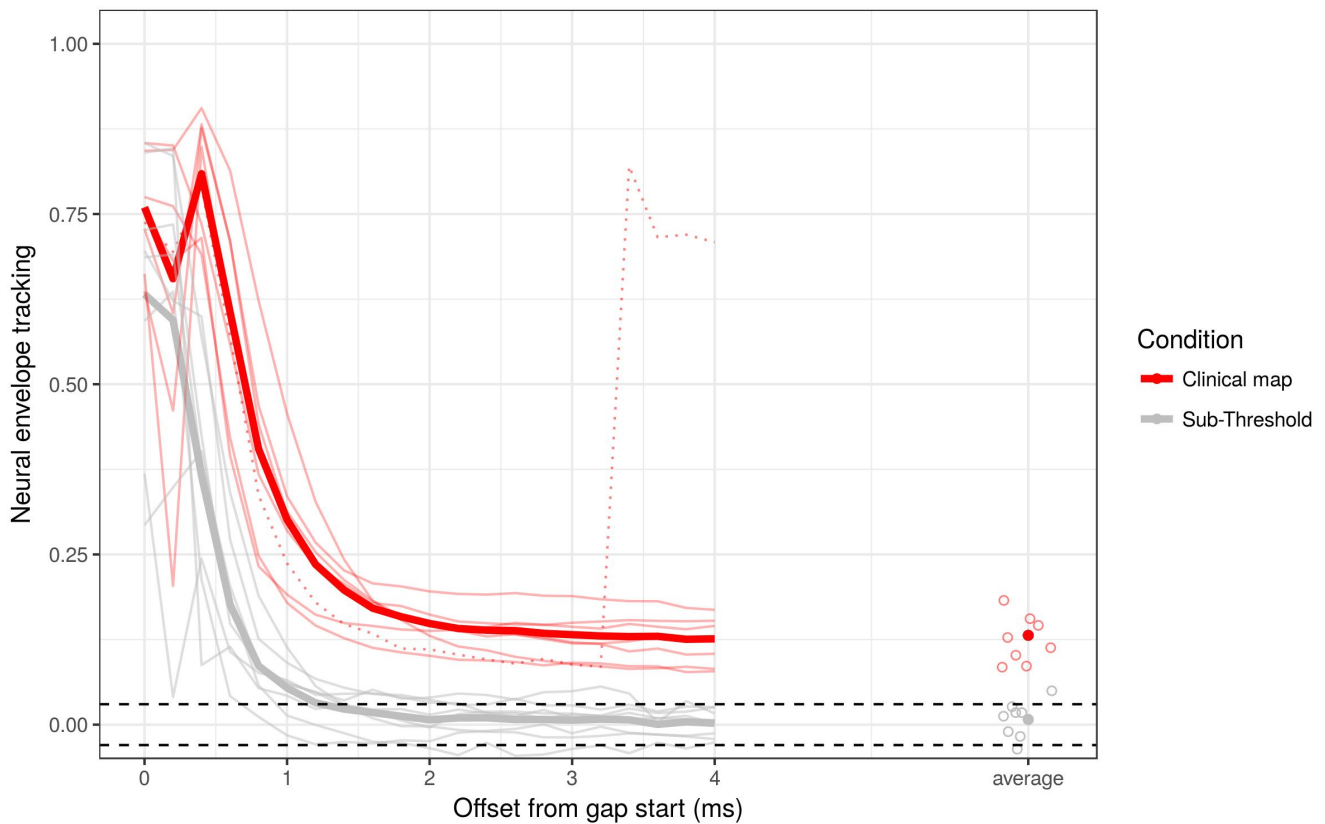


Figure 5. Neural envelope tracking results over the gap length. The red lines are the mean result per participant at baseline settings. The black lines are the result of the condition where speech was presented sub threshold. The bold lines represent the mean over participants. The dashed lines are the 95% significance level of the correlation. The circles at the right hand side show the results of the average condition. The dotted red line shows the results of S7 with a deviant gap length.

220 Figure 5 shows the average magnitude of the correlation for every sample inside the gap per participant at
221 baseline settings (red color). The bold line represents the mean over participants. The correlation between
222 the stimulus envelope and reconstructed envelope consisting of samples at the start of the gap is high. When

223 using samples at a later point inside the gap, the correlation decreases until it reaches a plateau at the end
224 of the gap. This asymptotic decay inside the gap is present for every participant. The circles at the right
225 hand side of the graph indicate the average condition. This condition uses the average of the samples of the
226 last millisecond of the gap to reconstruct the envelope, reducing the EEG noise and resulting in a more
227 reliable reconstruction. Important to note is that one participant (S7) listened to speech with stimulation
228 gaps of only 3 ms, resulting in deviant results in function of gap length. The results of this participant are
229 shown on the graph as a dotted red line, but are not included to calculate the mean value of the group.

230 Next we calculated the correlation between the real and reconstructed envelope in the sub-T condition. In
231 this condition the CI was stimulated, but the recipient could not hear any sound. Results of this analysis are
232 shown in figure 5 in grey. Correlations at the beginning of the gap are still significant, but when taking
233 a sample far enough from the start, the correlation is below or within significance level for 6 out of 7
234 participants. The significance level of the correlation is indicated by the dashed horizontal lines. It is
235 calculated by correlating random permutations of the real and reconstructed envelope 1000 times and
236 taking percentile 2.5 and 97.5 to obtain a 95% confidence interval.

237 3.2.2 Relation between speech understanding and neural envelope tracking

238 To investigate whether speech understanding is related with neural envelope tracking in CI users, we
239 varied the stimulation levels by reducing them with a fixed value as shown in figure 1. To measure neural
240 envelope tracking, we calculated the Spearman correlation between the reconstructed envelope and the
241 stimulus envelope for an average of the samples within the last millisecond of the presented stimulation
242 gap to exclude artifacts. Conducting a test-retest analysis showed no significant difference between test and
243 retest correlations ($p=0.59$, $CI(95\%) = [-0.011; 0.019]$, Wilcoxon signed-rank test), therefore we averaged
244 the correlation of the test and retest conditions resulting in one correlation per participant per stimulation
245 level shift, except for participant S5 who only participated in the test condition.

246 Figure 6 shows that the more the stimulation levels (lower x-axis) decrease, the more the correlation
247 between the real and the reconstructed envelope, i.e., neural envelope tracking, also decreases (fixed
248 effect stimulation level shift, $df = 28$, $t = 4.60$, $p=0.0001$, LME). In addition, the results of the sub-T
249 condition are also shown. Similar to the sub-T results in figure 5, neural envelope tracking in the sub-T
250 condition is below or within significance level for 6 out of 7 participants. As an extra factor average speech
251 understanding across participants of the LIST sentences is also included in the figure on the upper x-axis to
252 show the interplay between decreasing stimulation levels, decreasing speech understanding and decreasing
253 neural envelope tracking (fixed effect speech understanding on neural envelope tracking, $df = 28$, $t = 3.56$,
254 $p=0.0013$, LME).

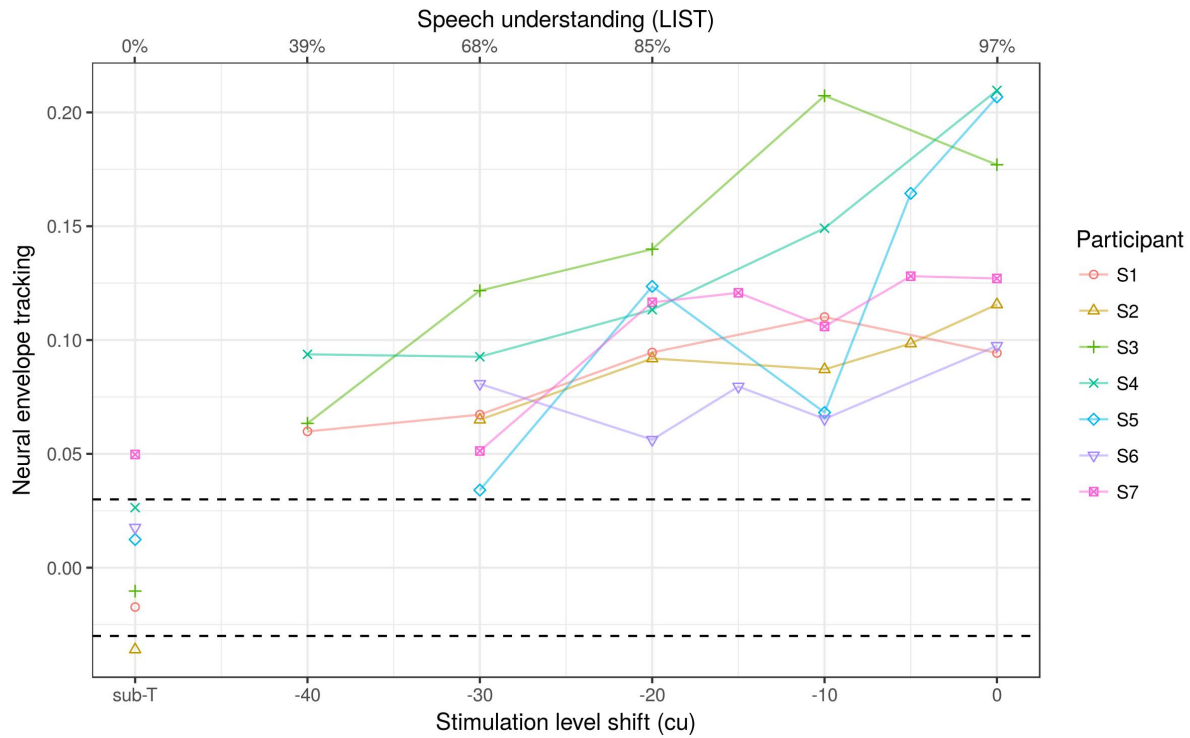


Figure 6. Neural envelope tracking increases with increasing stimulation level shift. The dashed black lines are the 95% significance level of the correlation. The values for speech understanding are the median values for the LIST sentences over participants.

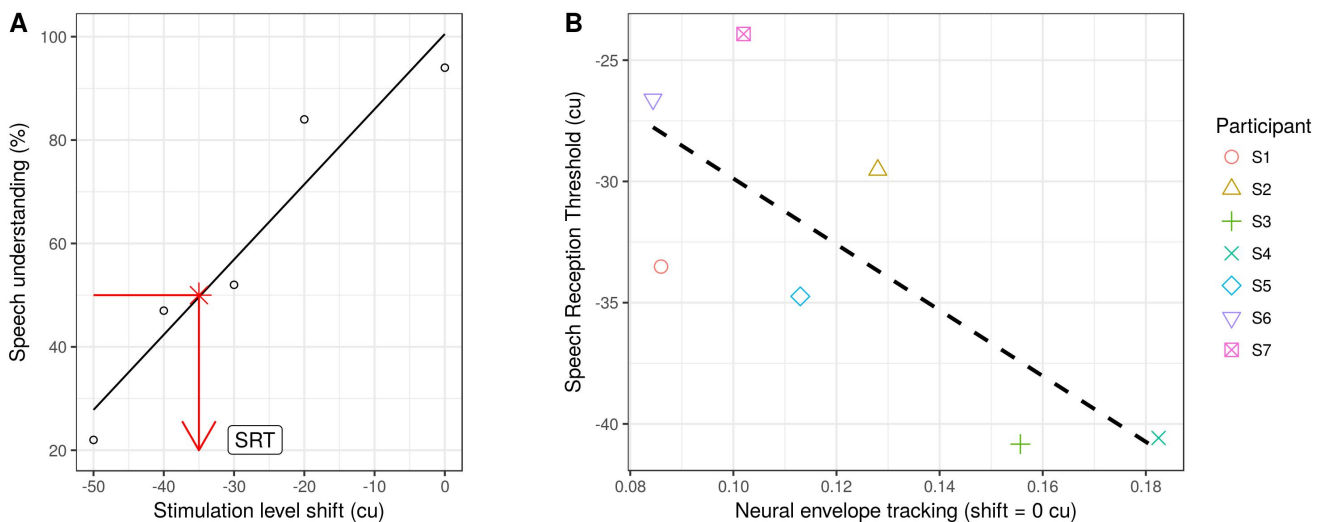


Figure 7. Neural envelope tracking correlates with speech understanding. Panel A shows how the behavioral speech understanding results of 1 participant (S5). A linear function is fitted through the data and the stimulation level shift corresponding to 50% speech understanding is labeled as the SRT. Panel B shows that the better (negative) the SRT, the higher neural envelope tracking is at baseline settings.

255 In the previous analysis we investigated the relation between speech understanding and neural envelope
256 tracking on a group level. To have a closer look at how participant specific speech understanding relates
257 to neural envelope tracking, we calculated the speech reception threshold (SRT) per participant, i.e., the
258 stimulation level shift yielding 50% speech understanding. This 50% point was calculated by fitting a
259 linear function on the data and solving the equation: $SRT = \frac{50\% - intercept}{slope}$ as shown in figure 7A. Next,
260 we correlated the SRT of every participant with the neural envelope tracking score on the 24 minutes of
261 speech at baseline settings, represented as circles on the right hand side in figure 5. The more negative the
262 SRT, the better the participant understands speech, the higher neural envelope tracking at baseline settings
263 (Pearson correlation = -0.76, p=0.048, figure 7B).

4 DISCUSSION

264 In this study we investigated if neural envelope tracking is related to speech understanding in CI users
265 similar to normal hearing listeners. To that end, we recorded the EEG of 8 CI users listening to a story
266 at varying levels of speech understanding by shifting the stimulation levels. An envelope reconstruction
267 analysis was conducted and compared to speech understanding results. We found increasing neural envelope
268 tracking with increasing stimulation levels and corresponding speech understanding which supports the
269 hypothesis that neural envelope tracking is related with speech understanding in CI users.

270 4.1 Speech understanding influenced by stimulation level shifts

271 As a first step we checked if the chosen method to vary speech understanding, i.e., shifting stimulation
272 levels, achieved the desired outcome by presenting several lists of LIST sentences at various stimulation
273 level shifts. Similar to results in normal hearing listeners where decreasing stimulus intensity is accompanied
274 by decreasing speech understanding (van Wieringen and Wouters, 2008), we found that decreasing
275 stimulation levels resulted in decreasing speech understanding. In addition, speech understanding did
276 decrease gradually with decreasing stimulation levels, giving the opportunity to investigate neural envelope
277 tracking at a wide range of speech understanding levels. Next we checked if the introduced stimulation
278 gaps, necessary to remove the artifact, affected speech understanding. This was not the case, however they
279 did affect the quality of the speech signal, possibly resulting in more listening effort which has to be taken
280 into account when comparing results of CI users to results of normal hearing listeners.

281 Besides the standardized recall test of the LIST sentences, we also asked the participants to rate their
282 speech understanding of the presented story during the EEG. As shown in figure 4, the variation between
283 participants is larger for the self-rated story than for the recalled LIST sentences. This can be explained
284 by the different way speech understanding was measured with more reliable results for the recalled LIST
285 sentences, indicating the importance of standardized speech tests in addition to self-rated measures.

286 **4.2 Influence of CI stimulation artifacts on neural envelop tracking**

287 To remove CI stimulation artifacts we used a validated artifact removal technique by Somers et al. (2018b)
288 which leaves out small groups of stimulation pulses during the presentation of a speech signal. We were
289 able to show that the correlation inside the gap between the stimulus envelope and reconstructed envelope
290 decreased when using samples further in the gap until it reached a plateau at the end of the gap (figure 5).
291 Although it is likely that the artifact was removed because of the present decay over the gap length, this
292 can not be guaranteed. Despite this clear decay in all participants, one participant (S7) showed significant,
293 but very low, neural envelope tracking in the sub-T condition until the end of the gap where no neural
294 envelope tracking was expected. This could mean that a gap of 4 ms in this participant was not enough for
295 the stimulus artifact to decay. Another explanation could be that although the stimulation was sub-T, the
296 participant still perceived some sounds, resulting in very small neural responses.

297 **4.3 Relation between speech understanding and neural envelope tracking**

298 Next, we investigated the relation between neural envelope tracking and speech understanding by varying
299 the stimulation levels of the CI. Similar to research in normal hearing listeners that showed increasing
300 neural envelope tracking with increasing speech understanding (Molinaro and Lizarazu, 2017; Ding and
301 Simon, 2013; Ding et al., 2014; Vanthornhout et al., 2018; Di Liberto et al., 2018), we showed similar
302 results in CI users although using a different approach. We varied the stimulation levels of the CI because
303 using speech versus non-speech stimuli (Molinaro and Lizarazu, 2017) or priming (Di Liberto et al., 2018)
304 would only result in 2 speech understanding levels. Furthermore, adding background noise to the speech
305 signal (Ding and Simon, 2013; Ding et al., 2014; Vanthornhout et al., 2018) would not be compatible with
306 the artifact removal method as the noise would constantly be interrupted. Therefore we decided to directly
307 manipulate a parameter of the CI, namely the stimulation levels. By doing so, we not only varied speech
308 understanding but also investigated the effect of adjusting a CI parameter based on neural envelope tracking
309 which demonstrates the feasibility of fitting a CI based on this measure. Besides the group analysis, we
310 also investigated the relation between speech understanding and neural envelope tracking on a participant
311 specific level similar to Vanthornhout et al. (2018). We were able to show that participants with good speech
312 understanding (good SRTs) had enhanced neural envelope tracking at baseline settings, again showing the
313 potential of neural envelope tracking as an objective measure of speech understanding ion CI users.

314 A potential confound in our study is loudness. Varying stimulation levels not only varies speech
315 understanding, it additionally affects the loudness of the stimulation. Therefore, it is difficult to distinguish
316 if neural envelope tracking decreased with decreasing stimulation levels because of speech understanding
317 or loudness. However Ding and Simon (2012) showed no difference in neural envelope tracking for a
318 variation in stimulus intensity over 16 dB. Nevertheless, even if the found effect would be influenced by
319 loudness in addition to speech understanding, this would still be an interesting result in the context of
320 objective fitting, showing that neural envelope tracking can be used to adjust the parameter settings of the
321 CI.

322 **4.4 Implications for applied research**

323 This study is the first to show a link between neural envelope tracking and speech understanding in CI
324 users, indicating the potential of neural envelope tracking as a measure of speech understanding in CI users.
325 Further research in this field could enable the development of an objective measure of speech understanding
326 in CI users with application potential in the field of objective clinical measures and neuro-steered hearing
327 aids.

5 CONCLUSION

328 This study confirms that neural envelope tracking responses can be found in CI users in response to running
329 speech using appropriate CI artifact removal methods. Furthermore, these responses become weaker as the
330 stimulus is presented at less intelligible stimulation levels. Neural envelope tracking can serve as a measure
331 of speech understanding that directly relates to settings of the CI, and thus has application potential in
332 objective and automatic fitting of CIs.

333 Acknowledgements: The authors would also like to thank Wivine Decoster for narrating the stories.
334 Funding: This work was supported by the European Research Council (ERC) under the European Union's
335 Horizon 2020 research and innovation programme [grant number 637424 (Tom Francart)]; the KU Leuven
336 Special Research Fund [grant number OT/14/119]; and the Research Foundation Flanders (FWO) [grant
337 numbers 1S46117N (Ben Somers), 1S86118N (Eline Verschueren)].

REFERENCES

- 338 Aiken, S. J. and Picton, T. W. (2008). Human Cortical Responses to the Speech Envelope. *Ear & Hearing*
339 29, 139–157
- 340 Deprez, H., Gransier, R., Hofmann, M., Wieringen, A. V., Wouters, J., and Moonen, M. (2017). Biomedical
341 Signal Processing and Control Characterization of cochlear implant artifacts in electrically evoked
342 auditory steady-state responses. *Biomedical Signal Processing and Control* 31, 127–138. doi:10.1016/j.
343 bspc.2016.07.013
- 344 Di Liberto, G. M., Lalor, E. C., and Millman, R. E. (2018). Causal cortical dynamics of a predictive
345 enhancement of speech intelligibility. *NeuroImage* 166, 247–258. doi:10.1016/j.neuroimage.2017.10.
346 066
- 347 Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies
348 on the spectro-temporal fine structure. *NeuroImage* 88, 41–46. doi:10.1016/j.neuroimage.2013.10.054
- 349 Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2015). Cortical tracking of hierarchical
350 linguistic structures in connected speech. *Nature Neuroscience* 19, 158–64. doi:10.1038/nn.4186
- 351 Ding, N. and Simon, J. Z. (2011). Neural coding of continuous speech in auditory cortex during monaural
352 and dichotic listening. *Journal of Neurophysiology* 107, 78–89. doi:10.1152/jn.00297.2011
- 353 Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to
354 competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*
355 109, 11854–9. doi:10.1073/pnas.1205381109
- 356 Ding, N. and Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical
357 Representation of Speech. *Journal of Neuroscience* 33, 5728–5735. doi:10.1523/JNEUROSCI.5297-12.
358 2013
- 359 Francart, T., van Wieringen, A., and Wouters, J. (2008). APEX 3: a multi-purpose test platform for
360 auditory psychophysical experiments. *Journal of Neuroscience Methods* 172, 283–293. doi:http:
361 //dx.doi.org/10.1016/j.jneumeth.2008.04.020
- 362 Hoffman, M. and Wouters, J. (2010). Electrically Evoked Auditory Steady State Responses in Cochlear
363 Implant Users. *JARO* 282, 267–282. doi:10.1007/s10162-009-0201-z
- 364 Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., McDarby, G., and Foxe, J. J. (2006). The VESPA: A
365 method for the rapid estimation of a visual evoked potential. *NeuroImage* 32, 1549–1561. doi:10.1016/j.
366 neuroimage.2006.05.054

- 367 Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving Precise Temporal Processing
368 Properties of the Auditory System Using Continuous Stimuli. *Journal of Neurophysiology* 102, 349–359.
369 doi:10.1152/jn.90896.2008
- 370 Loizou, P. C. (1998). Mimicking the human ear. *Signal Processing Magazine, IEEE* , 101–130
- 371 Loizou, P. C., Poroy, O., and Dorman, M. (2000). The effect of parametric variations of cochlear implant
372 processors on speech understanding The effect of parametric variations of cochlear implant. *JASA* 108,
373 790–802. doi:10.1121/1.429612
- 374 Luo, H. and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in
375 human auditory cortex. *Neuron* 54, 1001–10. doi:10.1016/j.neuron.2007.06.004
- 376 Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., and Friederici, A. D. (2017). Linguistic Bias
377 Modulates Interpretation of Speech via Neural Delta-Band Oscillations. *Cerebral Cortex* 27, 4293–4302.
378 doi:10.1093/cercor/bhw228
- 379 Molinaro, N. and Lizarazu, M. (2017). Delta(but not theta)-band cortical entrainment involves speech-
380 specific processing. *European Journal of Neuroscience* 9, 1–9. doi:10.1111/ejn.13811
- 381 Petersen, E. B., Wöstmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus
382 ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology* 117, 18–27.
383 doi:10.1152/jn.00527.2016
- 384 Shannon, R. V., Cruz, R. J., and Galvin, J. J. (2011). Effect of Stimulation Rate on Cochlear Implant Users
385 ' Phoneme , Word and Sentence Recognition in Quiet and in Noise. *Audiol Neurotol* 19, 113–123.
386 doi:10.1159/000315115
- 387 Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech Recognition with
388 Primarily Temporal Cues. *Science* 270, 303–304. doi:10.1126/science.270.5234.303
- 389 Somers, B., Francart, T., and Bertrand, A. (2018a). A generic EEG artifact removal algorithm based on the
390 multi-channel Wiener filter. *Journal of neural engineering* 15. doi:10.1088/1741-2552/aaac92
- 391 Somers, B., Verschueren, E., and Francart, T. (2018b). Neural tracking of the speech envelope in cochlear
392 implant users. *Journal of neural engineering* , Accepted
- 393 Swanson, B. and Mauch, H. (2006). Nucleus Matlab Toolbox 4.20 software user manual. Cochlear Ltd,
394 Lane Cove NSW, Australia.
- 395 van Wieringen, A. and Wouters, J. (2008). LIST and LINT: Sentences and numbers for quantifying speech
396 understanding in severely impaired listeners for Flanders and the Netherlands. *International Journal of*
397 *Audiology* 47, 348–355. doi:10.1080/14992020801895144
- 398 Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., and Francart, T. (2018). Speech intelligibility
399 predicted from neural entrainment of the speech envelope. *JARO* 19, 181–191
- 400 Wouters, J., Mcdermott, H. J., and Francart, T. (2015). Sound coding in cochlear implants : from electric
401 pulses to hearing. *Signal Processing Magazine, IEEE* 32, 67–80. doi:10.1109/MSP.2014.2371671.