Citation/Reference	Eline Verschueren, Ben Somers, Tom Francart Neural envelope tracking as a measure of speech understanding in cochlear implant users Hearing Research 373:23-31 March 201
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	https://www.sciencedirect.com/science/article/pii/S0378595518304878
DOI	10.1016/j.heares.2018.12.004
Journal homepage	https://www.journals.elsevier.com/hearing-research
Author contact	eline.verschueren@kuleuven.be , tom.francart@kuleuven.be

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
10 from the EEG using a linear decoder. Next, the reconstructed envelope was correlated with the
11 envelope of the speech stimulus as a measure of neural envelope tracking which was compared
12 to actual speech understanding.

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

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

1 INTRODUCTION

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

Neural envelope tracking has also been measured in listeners with a hearing impairment by Petersen et al. (2017). They showed that neural tracking of the to-be-ignored speech increased with hearing loss. However, the attended speech was unaffected, resulting in a diminishing difference between the attended and unattended speech stream in persons with increasing hearing loss.

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

However, a major problem when measuring neural envelope tracking in CI users is the occurrence of stimulation artifacts in the EEG (e.g., Somers et al. (2019); Deprez et al. (2017); Hoffman and Wouters (2010)). When a CI electrically stimulates the auditory nerve, large electric potentials from the CI appear in the EEG. Just like the brain responses, the modulations in these CI artifacts follow the envelope of the presented speech signal. As a consequence it is difficult to distinguish between the desired brain response and the unwanted artifact. Recently a novel artifact removal method has been developed by Somers et al.

(2019) which works by leaving out small groups of stimulation pulses during the presentation of a speech signal. This creates short artifact-free EEG windows during which the ongoing brain responses can be measured. This method was validated in CI users.

In the current study we investigated the effect of speech understanding on neural envelope tracking using EEG in CI users by applying the newly proposed artifact removal method by Somers et al. (2019). We hypothesized that neural envelope tracking will increase with speech understanding.

2 MATERIAL AND METHODS

2.1 Participants

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

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

2.2 Preprocessing of the stimuli

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

76 modulated with the envelope of the signal in each of these frequency bands. Finally, every modulated pulse
77 train is mapped to an appropriate stimulus intensity and applied to one of the 22 implanted electrodes in the
78 cochlea to stimulate the auditory nerve and enable hearing. The intensities of the pulses in each electrode's
79 pulse train are mapped between that electrode's Threshold (T) and Comfortable (C) level¹. Pulse rate was
80 fixed at 900 pulses per second, stimulation mode was set to monopolar and the number of active electrodes
81 was 22 (except for participant S1: 18 electrodes). All remaining parameter settings can be found in Table I.
82 All parameters required in the conversion from audio to electrical stimulation, such as the T and C level for
83 every electrode, pulse rate and stimulation mode, were derived from the participant's clinical settings.

Figure 1. Illustration of gap insertion in the electrodegram 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 4 ms long at a rate of 40 Hz.

84 To tackle the problem of CI stimulation artifacts when using EEG, we applied a new artifact removal
85 method proposed and validated by Somers et al. (2019). In brief, this method obtains EEG free of
86 stimulation artifacts by periodically interrupting the electrical stimulation by leaving out small groups of
87 stimulation pulses as shown in Figure 1, further referred to as a stimulus with ‘dropped pulses’. Within
88 these interruptions artifact-free EEG can be sampled. The stimulus interruptions were 4 ms long at a rate of

¹ The T-level, or threshold level, is the lowest audible stimulation level. The C-level or comfortable level, is the loudest stimulation level that can be comfortably listened to.

89 40 Hz. This is short enough to preserve speech understanding and long enough to remove the artifact. Only
90 the EEG within the stimulation gaps, i.e., artifact free EEG, is further analyzed.

Figure 2. 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.

91 To vary speech understanding we directly manipulated a parameter of the CI, namely the stimulation
92 levels. We varied the stimulation levels by shifting all stimulation levels between T and C level by an equal
93 number of current units. In this way we maintained the dynamic range of the speech envelope but varied the
94 stimulation levels and thus audibility and consequently speech understanding (Figure 2). We selected this
95 approach, as using speech versus non-speech stimuli (Molinaro and Lizarazu, 2017) or priming (Di Liberto
96 et al., 2018), on the one hand, would only result in 2 speech understanding levels. Adding background
97 noise to the speech signal (Ding and Simon, 2013; Ding et al., 2014; Vanthornhout et al., 2018), on the
98 other hand, would not be compatible with the artifact removal method as the noise would constantly be
99 interrupted.

100 2.3 Behavioral experiment

101 Before the start of the experiment, we checked whether the participant felt comfortable listening to the
102 stimulus with dropped pulses by presenting them with a sentence including stimulation gaps of 4 ms at a

103 rate of 40 Hz. These values were chosen based on the findings of Somers et al. (2019). If the participant
104 could not tolerate the distortions induced by the gaps, we decreased the gap rate and/or length until a
105 comfortable level was reached.

106 After selecting the optimal gap parameters, speech understanding was measured behaviorally in order
107 to compare neural envelope tracking results with actual speech understanding. We used the Leuven
108 intelligibility sentence test (LIST) (van Wieringen and Wouters, 2008), uttered in Flemish by Wivine
109 Decoster (female speaker). This speech material is suitable to measure speech understanding in CI users
110 or persons with a severe hearing loss due to the low speech rate (2.5 syllables/second) and the key word
111 scoring. Every participant started with 1 training list where all stimulation levels were mapped between the
112 actual T and C levels, which will be referred to in the sequel as the baseline settings. Thereafter a number
113 of lists, each containing 10 sentences, were presented at different stimulation level shifts to vary speech
114 understanding. Stimulation levels were only shifted to lower values, as stimulating current units (cu) above
115 C levels would make speech too loud and could harm the participant. Stimulation level shifts varied from
116 0 cu (i.e., baseline settings) to -50 cu (almost inaudible as C levels shift under original T levels for most
117 participants depending on their dynamic range). Participants had to recall the sentence they heard. By
118 counting the correctly recalled key words, a percentage correct per stimulation level shift was calculated.
119 Additionally a list without stimulation gaps was presented at baseline settings to check the influence of
120 inserting gaps on speech understanding.

121 2.4 EEG experiment

122 Recordings were made in a soundproof and electromagnetically shielded room. A 64-channel BioSemi
123 ActiveTwo EEG recording system was used at a sample rate of 16384 Hz. Participants sat in a comfortable
124 chair and were asked to move as little as possible during the recordings. The stories used during the EEG
125 measurement were the stories ‘Marfoesjka en de Vorst’ (translated Russian fairytale) and ‘Luna van de
126 boom’ (Bart Moeyaert), both narrated in Flemish by Wivine Decoster, the speaker of the LIST sentences.
127 The stories had a total length of 48 minutes. The stories were cut in different blocks: 3 longer blocks of $\approx$
128 8 minutes and 10 blocks of $\approx$ 2.4 minutes which were presented in chronological order. These blocks had
129 slightly different lengths because word and sentence boundaries were taken into account. The long blocks
130 were presented at baseline settings and used to train the decoder on. The short blocks were presented in
131 random order at 5 stimulation level shifts selected from a fixed list, containing 0 cu, -5 cu, -10 cu, -15 cu,
132 -20 cu, -30 cu, -40 cu, to vary speech understanding as shown in the experiment overview in Figure 3. The
133 selection of stimulation level shifts was based on the reported speech understanding of the participant.
134 Every selected stimulation level shift was applied twice, on a different part of the story, to analyze test-retest
135 reliability.

Figure 3. Overview of an example test session. The session starts with the behavioral part where speech understanding was measured with the LIST sentences. The numbers in each block indicate the stimulation level shifts in current units. In the second part, the EEG part, 2 stories were presented containing 3 longer blocks at baseline settings (0 cu shift) to calculate the decoder on. Ten shorter blocks were presented in random order at 5 different stimulation level shifts to vary speech understanding. In the middle a block of the story was presented subthreshold (sub-T).

136 Additionally, we presented 5 minutes of the story below the participant’s T level (referred to as sub-T),
 137 meaning all stimulation was lower than the lowest stimulation level audible for that participant, which
 138 resulted in no sound sensation. In the sub-T condition there are CI stimulation artifacts that are highly
 139 correlated with the stimulus envelope in the absence of neural responses. This condition can be used
 140 to check if CI artifacts are effectively removed. If significant neural envelope tracking is found in this
 141 condition, this is caused by remaining CI artifacts. If no significant response is found in this condition, we
 142 can assume that the CI artifacts are removed.

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

150 2.5 Signal processing

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

153 2.5.1 Stimulus envelope

154 In studies evaluating neural envelope tracking in acoustic hearing, the acoustic stimulus envelope is
 155 chosen as the reference envelope. In this study, investigating electrical hearing, using the electrical envelope
 156 as a reference is more appropriate (Somers et al., 2019). The electrical envelope namely takes into account

all the preprocessing done by the CI before the speech reaches the cochlea, making this a more realistic reference. The electric speech envelope was extracted from the stimulus by combining the magnitudes of the electrical pulses over all stimulation channels. As a next step, the electric speech envelope was band-pass filtered in the delta (0.5-4 Hz) frequency band.

2.5.2 Envelope reconstruction

This section details the signal processing applied to the EEG to obtain the reconstructed envelope, including artifact removal, filtering and calculating the linear decoder.

As a first step, CI artifacts were removed by only using EEG samples occurring within the 4 ms gaps inserted in the stimulus (see section 2.2). In practice, obtaining the artifact-free EEG samples was achieved by decimating the recorded EEG to the gap frequency of 40 Hz (Somers et al., 2019), i.e. only one sample per gap is retained. The offset of the retained sample relative to the gap onset is a parameter to be set by the experimenter. Due to the exponential decay of the artifacts, small offsets (0-2 ms) likely result in EEG with residual stimulus artifacts, whereas larger offsets (2-4 ms) will ensure that the artifact has largely decayed. Repeating the analysis for different offsets reveals the contribution of the decaying CI artifact to the outcome measure, as seen in Figure 5. Other common EEG artifacts such as eye blink and muscle artifacts were removed using a multi-channel Wiener filter (Somers et al., 2018). After artifact rejection, the signal was bandpass filtered, similar to the electric speech envelope.

The mTRF Matlab toolbox (Lalor et al., 2006, 2009) was used to compute the linear decoder which reconstructs the speech envelope from the EEG recordings. As speech elicits neural responses with some delay, the linear decoder combines EEG channels and their time shifted versions to optimally reconstruct the speech envelope. If g is the linear decoder and R the shifted neural data, the reconstruction of the speech envelope $\hat{s}(t)$ was obtained by $\hat{s}(t) = \sum_n \sum_\tau g(n, \tau) R(t + \tau, n)$ with t the time index, n ranging over the recording electrodes and τ ranging over the integration window, i.e., the number of post-stimulus samples used to reconstruct the envelope. The decoder was calculated by solving $g = (RR^T)^{-1}(Rs^T)$ with s the speech envelope and applying ridge regression to prevent overfitting. We used an integration window of 250 ms post-stimulus resulting in the decoder matrix g of 64 (EEG channels) x 11 (time delays within the integration window). The decoder was created using a combination of the 3 long segments of $\approx$ 8 minutes, not including the shorter trials.

To investigate neural envelope tracking we did two different analyses. First, to check whether the decimation approach was sufficient to remove the CI artifacts, we repeated a leave-one-out cross-validation on the 24 minutes of speech at baseline settings for varying decimation offset values, as described in the previous paragraph. We hypothesized that outcome correlations calculated in repeats with small offsets (containing artifact) will be higher, as the artifact is highly correlated with the speech envelope. In this case, the outcome measure is not neural but a false positive. When using samples further in the gap, where the artifact has died out, correlations will be smaller, similar to correlations of previous studies with acoustic listeners, only containing the possible neural response. The leave-one-out cross-validation was done by

splitting the data in equal blocks of 2 minutes. The first block (2 minutes) was selected as the testing data, while the rest was concatenated to create one decoder used to reconstruct the envelope for the testing part. Next, the second block was selected as the testing data and another decoder was trained on the remaining blocks et cetera. Secondly, after validating the CI artifact removal, we performed an analysis to investigate the effect of intelligibility on the neural envelope tracking. To this end, we concatenated the 24 minutes of EEG data at baseline settings to create one decoder per participant using the average of the EEG samples in the last millisecond of the stimulation gap, i.e. the average condition (see section 3.2.1). This decoder was applied on the 10 short segments with different stimulation level shifts, resulting in 10 outcome correlation measures per participant, i.e., 1 per trial.

2.6 Statistical Analysis

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

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

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

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

3 RESULTS

3.1 Behavioral speech intelligibility

For 7 out of 8 participants, the gap settings of 4 ms and 40 Hz were chosen. One participant (S7) reported that the speech was too distorted, a gap length of 3 ms was chosen for this participant. All participants reported that the speech with dropped pulses sounded more robotic, but speech understanding was not affected. We checked this by comparing the recall scores for LIST sentences with and without dropped pulses at baseline settings and found no significant difference between the two ($p=0.7352$, $CI(95\%) = [-13.00\%; 6.00\%]$, $n=7$, Wilcoxon signed-rank test). In the following analysis S8 is included to investigate which EEG samples are artifact-free, but excluded to investigate the possible link of neural envelope tracking with speech understanding as the participant 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.

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

3.2 Neural envelope tracking

3.2.1 Influence of CI stimulation artifacts on neural envelope tracking

We investigated the magnitude of the correlation between the real and the reconstructed envelope for each sample inside the stimulation gap. We hypothesized that correlations calculated for samples that contain artifacts will be especially high as the artifact resembles the signal of interest. When using samples

239 where the artifact has died out, correlations will be smaller, similar to correlations of previous studies with
240 acoustic listeners, 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 subthreshold. 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.

241 Figure 5 shows the average magnitude of the correlation for every sample inside the gap per participant at
242 baseline settings (red color). The bold line represents the mean over participants. The correlation between
243 the stimulus envelope and reconstructed envelope consisting of samples at the start of the gap is high. When
244 using samples at a later point inside the gap, the correlation decreases until it reaches a plateau at the end
245 of the gap. This asymptotic decay inside the gap is present for every participant. The circles at the right
246 hand side of the graph indicate the average condition. This condition uses the average of the samples of the
247 last millisecond of the gap to reconstruct the envelope, reducing the EEG noise and resulting in a more
248 reliable reconstruction. Important to note is that one participant (S7) listened to speech with stimulation
249 gaps of only 3 ms, resulting in deviant results in function of gap length. The results of this participant are
250 shown on the graph as a dotted red line, but are not included to calculate the mean value of the group.

Next we calculated the correlation between the real and reconstructed envelope in the sub-T condition. In this condition the CI was stimulated, but the recipient could not hear any sound. Results of this analysis are shown in Figure 5 in grey. Correlations at the beginning of the gap are still significant, but when taking a sample far enough from the start, the correlation is below or within significance level for 6 out of 7 participants. The significance level of the correlation is indicated by the dashed horizontal lines. It is calculated 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.

3.2.2 Relation between speech understanding and neural envelope tracking

To investigate whether speech understanding is related with neural envelope tracking in CI users, we varied the stimulation levels by reducing them with a fixed value as shown in Figure 2. To measure neural envelope tracking, we calculated the Spearman correlation between the reconstructed envelope and the stimulus envelope for an average of the samples within the last millisecond of the presented stimulation gap to exclude artifacts. Conducting a test-retest analysis showed no significant difference between test and retest correlations ($p=0.59$, $CI(95\%) = [-0.011; 0.019]$, Wilcoxon signed-rank test), therefore we averaged the correlation of the test and retest conditions resulting in one correlation per participant per stimulation level shift, except for participant S5 who only participated in the test condition.

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

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

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.

4 DISCUSSION

284 In this study we investigated if neural envelope tracking is related to speech understanding in CI users
285 similar to normal hearing listeners. To that end, we recorded the EEG of 8 CI users listening to a story
286 at varying levels of speech understanding by shifting the stimulation levels. An envelope reconstruction
287 analysis was conducted and compared to speech understanding results. The results of this study support
288 the hypothesis that increasing neural envelope tracking is related to increasing speech understanding in CI
289 users.

290 4.1 Speech understanding influenced by stimulation level shifts

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

301 Besides the standardized recall test of the LIST sentences, we also asked the participants to rate their
302 speech understanding of the presented story during the EEG. As shown in Figure 4, the variation between
303 participants is larger for the self-rated story than for the recalled LIST sentences. This can be explained by
304 the different way speech understanding was measured: self-rating by the participant (story) compared to
305 the standardized method where a percentage correct is calculated based on correctly recalled words (LIST
306 sentences). This indicates the importance of standardized speech tests in addition to self-rated measures.

307 4.2 Influence of CI stimulation artifacts on neural envelop tracking

308 To remove CI stimulation artifacts we used a validated artifact removal technique by Somers et al. (2019)
309 which leaves out small groups of stimulation pulses during the presentation of a speech signal. We were
310 able to show that the correlation inside the gap between the stimulus envelope and reconstructed envelope
311 decreased when using samples further in the gap until it reached a plateau at the end of the gap (Figure
312 5). Although it is likely that the artifact was removed because of the decay over the gap length, this can
313 not be guaranteed. Despite this clear decay in all participants, one participant (S7) showed significant,
314 but very low, neural envelope tracking in the sub-T condition until the end of the gap where no neural
315 envelope tracking was expected. This could mean that a gap of 4 ms in this participant was not enough
316 for the stimulus artifact to decay. Another explanation could be that although the stimulation was sub-T,
317 the participant still perceived some sounds, resulting in very small neural responses. Despite the good

performance of this artifact removal technique, a limitation of this approach is that the stimulation has to be manipulated which is possible in a controlled setting, such as a research or clinical setting, but not in a non-controlled environment.

4.3 Relation between speech understanding and neural envelope tracking

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

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

4.4 Implications for applied research

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

5 CONCLUSION

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

357 Acknowledgements: The authors would also like to thank Wivine Decoster for narrating the stories.

358 Funding: This work was supported by the European Research Council (ERC) under the European Union's

359 Horizon 2020 research and innovation programme [grant number 637424 (Tom Francart)]; the KU Leuven

360 Special Research Fund [grant number OT/14/119]; and the Research Foundation Flanders (FWO) [grant

361 numbers 1S46117N (Ben Somers), 1S86118N (Eline Verschueren)].

REFERENCES

- 362 Aiken, S.J., Picton, T.W., 2008. Human Cortical Responses to the Speech Envelope. *Ear & Hearing* 29,
- 363 139–157.
- 364 Deprez, H., Gransier, R., Hofmann, M., Wieringen, A.V., Wouters, J., Moonen, M., 2017.
- 365 Biomedical Signal Processing and Control Characterization of cochlear implant artifacts in electrically
- 366 evoked auditory steady-state responses. *Biomedical Signal Processing and Control* 31, 127–
- 367 138. URL: <http://dx.doi.org/10.1016/j.bspc.2016.07.013>, doi:[10.1016/j.bspc.2016.07.013](https://doi.org/10.1016/j.bspc.2016.07.013).
- 368
- 369 Di Liberto, G.M., Lalor, E.C., Millman, R.E., 2018. Causal cortical dynamics of a predictive enhancement
- 370 of speech intelligibility. *NeuroImage* 166, 247–258. URL: <https://doi.org/10.1016/j.neuroimage.2017.10.066>, doi:[10.1016/j.neuroimage.2017.10.066](https://doi.org/10.1016/j.neuroimage.2017.10.066).
- 371
- 372 Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies
- 373 on the spectro-temporal fine structure. *NeuroImage* 88, 41–46. URL: <http://dx.doi.org/10.1016/j.neuroimage.2013.10.054>, doi:[10.1016/j.neuroimage.2013.10.054](https://doi.org/10.1016/j.neuroimage.2013.10.054),
- 374 [arXiv:NIHMS150003](https://arxiv.org/abs/1508.00003).
- 375
- 376 Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic
- 377 structures in connected speech. *Nature Neuroscience* 19, 158–64. URL: <http://www.ncbi.nlm.nih.gov/pubmed/26642090>, doi:[10.1038/nn.4186](https://doi.org/10.1038/nn.4186), [arXiv:15334406](https://arxiv.org/abs/15334406).
- 378
- 379 Ding, N., Simon, J.Z., 2011. Neural coding of continuous speech in auditory cortex during monaural and
- 380 dichotic listening. *Journal of Neurophysiology* 107, 78–89. doi:[10.1152/jn.00297.2011](https://doi.org/10.1152/jn.00297.2011).
- 381
- 382 Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to
- 383 competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*
- 384 109, 11854–9. doi:[10.1073/pnas.1205381109](https://doi.org/10.1073/pnas.1205381109).
- 385
- 386 Ding, N., Simon, J.Z., 2013. Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical
- 387 Representation of Speech. *Journal of Neuroscience* 33, 5728–5735. doi:[10.1523/JNEUROSCI.5297-12.2013](https://doi.org/10.1523/JNEUROSCI.5297-12.2013).
- 388
- 389 Francart, T., van Wieringen, A., Wouters, J., 2008. APEX 3: a multi-purpose test platform for auditory
- 390 psychophysical experiments. *Journal of Neuroscience Methods* 172, 283–293. URL: <http://www.sciencedirect.com/science/article/pii/S0165027008002513>, doi:[http://dx.doi.org/10.1016/j.jneumeth.2008.04.020](https://doi.org/10.1016/j.jneumeth.2008.04.020).

391 Hoffman, M., Wouters, J., 2010. Electrically Evoked Auditory Steady State Responses in Cochlear Implant
 392 Users. *JARO* 282, 267–282. doi:[10.1007/s10162-009-0201-z](https://doi.org/10.1007/s10162-009-0201-z).

393 Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., Foxe, J.J., 2006. The VESPA: A method
 394 for the rapid estimation of a visual evoked potential. *NeuroImage* 32, 1549–1561. doi:[10.1016/j.](https://doi.org/10.1016/j.neuroimage.2006.05.054)
 395 [neuroimage.2006.05.054](https://doi.org/10.1016/j.neuroimage.2006.05.054).

396 Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving Precise Temporal Processing
 397 Properties of the Auditory System Using Continuous Stimuli. *Journal of Neurophysiology* 102, 349–
 398 359. URL: <http://jn.physiology.org/cgi/doi/10.1152/jn.90896.2008>, doi:[10.](https://doi.org/10.1152/jn.90896.2008)
 399 [1152/jn.90896.2008](https://doi.org/10.1152/jn.90896.2008).

400 Loizou, P.C., 1998. Mimicking the human ear. *Signal Processing Magazine, IEEE* , 101–130.

401 Loizou, P.C., Poroy, O., Dorman, M., 2000. The effect of parametric variations of cochlear implant
 402 processors on speech understanding The effect of parametric variations of cochlear implant. *JASA* 108,
 403 790–802. doi:[10.1121/1.429612](https://doi.org/10.1121/1.429612).

404 Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human
 405 auditory cortex. *Neuron* 54, 1001–10. doi:[10.1016/j.neuron.2007.06.004](https://doi.org/10.1016/j.neuron.2007.06.004).

406 Meyer, L., Henry, M.J., Gaston, P., Schmuck, N., Friederici, A.D., 2017. Linguistic Bias Modulates
 407 Interpretation of Speech via Neural Delta-Band Oscillations. *Cerebral Cortex* 27, 4293–4302. doi:[10.](https://doi.org/10.1093/cercor/bhw228)
 408 [1093/cercor/bhw228](https://doi.org/10.1093/cercor/bhw228).

409 Molinaro, N., Lizarazu, M., 2017. Delta(but not theta)-band cortical entrainment involves speech-
 410 specific processing. *European Journal of Neuroscience* 9, 1–9. doi:[10.1111/ejn.13811](https://doi.org/10.1111/ejn.13811),
 411 [arXiv:0608246v3](https://arxiv.org/abs/0608246v3).

412 Petersen, E.B., Wöstmann, M., Obleser, J., Lunner, T., 2017. Neural tracking of attended versus ignored
 413 speech is differentially affected by hearing loss. *Journal of Neurophysiology* 117, 18–27. URL:
 414 <http://jn.physiology.org/lookup/doi/10.1152/jn.00527.2016>, doi:[10.1152/](https://doi.org/10.1152/jn.00527.2016)
 415 [jn.00527.2016](https://doi.org/10.1152/jn.00527.2016).

416 Shannon, R.V., Cruz, R.J., Galvin, J.J., 2011. Effect of Stimulation Rate on Cochlear Implant Users '
 417 Phoneme , Word and Sentence Recognition in Quiet and in Noise. *Audiol Neurotol* 1922, 113–123.
 418 doi:[10.1159/000315115](https://doi.org/10.1159/000315115).

419 Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech Recognition with
 420 Primarily Temporal Cues. *Science* 270, 303–304. URL: [http://www.sciencemag.org/](http://www.sciencemag.org/cgi/doi/10.1126/science.270.5234.303)
 421 [cgi/doi/10.1126/science.270.5234.303](http://www.sciencemag.org/cgi/doi/10.1126/science.270.5234.303), doi:[10.1126/science.270.5234.303](https://doi.org/10.1126/science.270.5234.303),
 422 [arXiv:arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).

423 Somers, B., Francart, T., Bertrand, A., 2018. A generic EEG artifact removal algorithm based on the
 424 multi-channel Wiener filter. *Journal of neural engineering* 15. doi:[10.1088/1741-2552/aaac92](https://doi.org/10.1088/1741-2552/aaac92).

425 Somers, B., Verschueren, E., Francart, T., 2019. Neural tracking of the speech envelope in cochlear implant
 426 users. *Journal of neural engineering* 16, 016003. doi:[10.1088/1741-2552/aae6b9](https://doi.org/10.1088/1741-2552/aae6b9).

427 Swanson, B., Mauch, H., 2006. Nucleus Matlab Toolbox 4.20 software user manual. Cochlear Ltd, Lane
 428 Cove NSW, Australia. .

- 429 Vanthornhout, J., Decruy, L., Wouters, J., Simon, J.Z., Francart, T., 2018. Speech intelligibility predicted
430 from neural entrainment of the speech envelope. *JARO* 19, 181–191.
- 431 van Wieringen, A., Wouters, J., 2008. LIST and LINT: Sentences and numbers for quantifying speech
432 understanding in severely impaired listeners for Flanders and the Netherlands. *International Journal of*
433 *Audiology* 47, 348–355. doi:[10.1080/14992020801895144](https://doi.org/10.1080/14992020801895144).
- 434 Wouters, J., Mcdermott, H.J., Francart, T., 2015. Sound coding in cochlear implants : from electric pulses
435 to hearing. *Signal Processing Magazine, IEEE* 32, 67–80. doi:[10.1109/MSP.2014.2371671](https://doi.org/10.1109/MSP.2014.2371671).