Towards Decoding Selective Attention from Single-Trial EEG Data in Cochlear Implant Users

Waldo Nogueira, Giulio Cosatti, Irina Schierholz, Maria Egger, Bojana Mirkovic, Andreas Büchner

***Abstract*—Previous results showed that it is possible to decode an attended speech source from EEG data via the reconstruction of the speech envelope in normal hearing (NH) listeners. However, so far it is unknown how the performance of such a decoder is affected by the decrease in spectral resolution and the electrical artifacts introduced by a cochlear implant (CI) in users of these prostheses. NH-listeners and bilateral CI-users participated in the present study. Speech from two audio books, one uttered by a male voice and one by a female voice, was presented to NH-listeners and CI-users. Participants were instructed to attend to one of the two speech streams presented dichotically while a 96-channel EEG was recorded. Speech envelope reconstruction from the EEG data was obtained by training decoders using a regularized least square estimation method. Decoding accuracy was defined as the percentage of accurately reconstructed trials for each subject. For NH listeners, the experiment was repeated using a vocoder to reduce spectral-resolution and to simulate speech perception with a CI in NH-listeners. The results showed a decoding accuracy of 80.9% using the original sound files in NH-listeners. The performance dropped to 73.2% in the vocoder condition and to 71.5% in the group of CI-users. In sum, although the accuracy drops when the spectral resolution becomes worse, the results show the feasibility to decode the attended sound source in NH-listeners with a vocoder simulation, and even in CI-users albeit more training data are needed.**

***Index Terms*—Cochlear implant, Selective attention, EEG, electroencephalography**

I. Introduction

Cochlear implants (CIs) are medical devices that are used to restore the sense of hearing in people with profound sensorineural hearing loss or complete deafness. They act as a kind of artificial cochlea, transforming the acoustic signal into an electric one, bypassing the damaged structures of the ear and directly stimulating the auditory nerve (for a review, see e.g. [33]). Over the past few decades, the CI sound processor has been extensively developed to further improve speech intelligibility outcomes (e.g. [34] [35]). Current technology provides CI users with good speech recognition in quiet [18] [36], but unsatisfactory speech understanding in more challenging listening environments with multiple speakers, background noise or reverberation (i.e. the cocktail party problem; [4]).

Part of this problem is caused by the limitations in binaural hearing, including sound localization and speech intelligibility, experienced by CI users. As a consequence, CI users lose the capability of identifying and understanding a particular speech stream in a noisy environment. The investigation of neural speech-tracking using electroencephalography (EEG) and the identification of the attended speaker in multi-talker scenarios from multi-channel scalp-EEG recordings [24] [27] have demonstrated that EEG could feasibly inform future CI algorithms about the listeners focus of attention. This information would allow CIs for instance to adapt noise suppression algorithms or to align directional microphones towards the attended sound source.

One constraint of CIs is the limited spectral information they deliver. Although only four spectral channels are sufficient to understand speech in quiet [29], speech perception in more difficult listening conditions requires a greater number of spectral channels [30]. The spectral information very likely is limited by channel interactions occurring when different electrodes stimulate overlapping populations of neurons (e.g. [12]). The smeared spectral information may, as a consequence, also cause smeared cortical responses, which may decrease the accuracy of detecting the attended speaker from the EEG signal in a cocktail party type scenario. The smeared spectral information may, as a consequence, also cause smeared cortical responses, which may decrease the accuracy of detecting the attended speaker from the EEG signal in a cocktail party type scenario. Vocoders that smear spectro-temporal fine structure while keeping the temporal envelope, diminish top-down attention to differentially process different speech streams as measured through EEG (Kong et al. 2015) and magnetoencephalography (MEG; Ding et al. 2013).

CIs produce different electrical artifacts depending upon the manufacturer, CI sound coding strategies, fitting parameters and individual maps (e.g. [37] [32]). These artifacts overlay with cortical responses, making it difficult to isolate one from the other (e.g. [40] [28]). McLaughlin et. al [22] showed that recorded EEG signals in CI users consist of a neural response, a high frequency artifact and a low frequency artifact. The high frequency artifact can be completely attenuated by low-pass filtering. The low frequency artifact is related to the stimulation pulses and its shape is similar to that of the acoustic stimulus envelope. This fact could impair the possibility to decode selective attention in CI users, as the paradigm requires ongoing EEG recording while the CI is stimulated. For stimuli other than continuous speech, it could be shown that the artifact can be attenuated, obtaining only the actual neural response (see e.g. [21] [31]). The complexity of continuous speech with its spectral characteristics, however, makes it challenging to estimate and remove the artifact. A recent study investigated the electrical CI artifact in EEG recordings in

W. Nogueira, G. Cosatti, I. Schierholz, M. Egger and A. Büchner are with the Dept. of Otolaryngology, Cluster of Excellence Hearing4all, Hannover Medical University, Hannover, Germany.

B. Mirkovic is with the Neuropsychology Lab, Dept. of Psychology, Carl von Ossietzky University of Oldenburg, Cluster of Excellence Hearing4all, Oldenburg, Germany.

response to continuous speech using a head model. The results showed that the artifact in response to speech is smaller in magnitude than the artifact in response to non-speech stimuli, which is suggested to be related to signal inherent amplitude modulations [32]. A more recent study has shown that it is possible to measure neural tracking of speech envelope in response to ongoing electrical stimulation by creating a sound coding strategy with gaps in which electrical stimulation is periodically interrupted. During this stimulation gaps, artefact free EEG can be sampled and used to train a linear envelope decoder (e.g. [48]). In summary, the electrical CI artifact evoked by continuous speech stimuli may overlay with the cortical responses, to an extent similar to the artifact created by more simple/brief stimuli.

In the present study, we first investigated the feasibility of decoding selective attention with dichotic stimulation using single trial EEG-data in normal hearing (NH) listeners with reduced spectral resolution of the presented speech signals. In a subsequent step we evaluated if the attended speech source can be identified as well from single trial EEG-data in CI users, where the speech signal is not only degraded, but where also the electrical artifact of the implant might have an impact on the decoding accuracy. The first hypothesis of the study is that spectral smearing may cause a reduction in the accuracy to detect selective attention. The second hypothesis is that the introduction of electrical artifact as produced by a CI will limit detecting the attended speech source in a selective attention paradigm.

II. Methods

A. Participants

Participants of the present study included 12 NH listeners (6 male; mean age: 26, range: 18-33, SD: 4.4 years) and 12 bilateral implanted CI users (7 male; mean age: 60, range: 48-80, SD: 11.0 years). CI users were all good performers, with a mean performance of 79% in the Freiburg monosyllabic word test [14] and of 98.5% in the HSM sentence test in quiet [15]. All CI users had at least 1 year experience with their devices. NH listeners had age-appropriate hearing with a hearing loss of less or equal than 10 dB in the frequency range of 0.25 to 8 kHz. Subjects demographics and additional information can be obtained in Tables 1 and 2. All subjects were native German speakers. The Color-Word Interference Test after Stroop [3] was used as a measure of selective attention. Speech recognition in noise was assessed using the HSM in noise and the Göttinger sentence test (GÖSA, adapting background noise; [16]). Prior to the experiment, all participants provided written informed consent and the study was carried out in accordance with the Declaration of Helsinki principles, approved by the Ethics Committee of the Hannover Medical School.

B. Stimuli

Following the study of Mirkovic and colleagues [24], participants were presented with two German narrations (”A drama in the air” by Jules Verne and ”Two brothers” by the Grimm brothers). The story by Jules Verne was narrated by a German male speaker, the story by the Grimm brothers by a German female speaker. As in the study by [27] and [24], silent periods were limited to 0.5s to ensure the listeners could maintain attention to the correct story and attention was not captured by the other story. NH listeners were presented with two different stimulation conditions (original speech, vocoded speech), whereas CI users only were presented with original speech. The vocoder condition in the group of NH listeners was used to simulate hearing with a reduced spectral resolution, to test the hypothesis that lack of spectral resolution, as it occurs in CI users, causes a reduction in the accuracy to detect selective attention. Note however that a vocoder is obviously not a CI. Stimuli for NH listeners were delivered via inserted earphones (3M E-A-RTONE 3A, 50 Ohm). CI users received auditory stimulation via two audio cables directly attached to the speech processor. Stimulus presentation was controlled by the Presentation software (Neurobehavioral System, version 16.5). In order to adjust the loudness to an individual moderate level of ∼60-70 dB(A), participants performed a loudness scaling on a seven-point loudness-rating scale (with 1 equivalent to very soft and 7 equivalent to extremely loud).

C. Procedure

Subjects were instructed to focus attention to one of the two concurrent stories while ignoring the other one. The story to be attended was randomized between participants. During the task, participants were instructed to keep their eyes open and to maintain fixation at the front. Stories were presented in 24 segments of 2 minutes duration each, resulting in a total task duration of 48 minutes. The whole task was subdivided into 6 blocks with 4 segments each. For CI subjects, all 6 blocks contained the original speech signal. NH listeners were presented with 3 blocks of original and 3 blocks of vocoded speech in alternating order, starting either with the original or the vocoded condition. Each participant respectively attended to the same story throughout the experiment, but the side from which the attended speech stream was presented changed after each segment to exclude effects of side of presentation. Before each segment, participants were instructed which side to attend. The starting side of the attended speech stream was randomized between participants. Within the breaks, i.e. after each 2 minutes segment, participants had to answer eight multiple-choice questions, four related to the attended and four to the unattended story, with four possible answers each.

D. Vocoder

Vocoder simulations, utilized in this study, were designed to model both, the processing typically performed in a CI, and the spread of excitation that may occur in an electrically stimulated cochlea [26]. The vocoder does not aim at reproducing the exact speech intelligibility of real CI users, but instead it aims at showing the effect of spectral smearing on decoding selective attention. Each token was digitally sampled at 16 kHz. The short-time Fourier transform was computed with a resolution of 256 bins, and a temporal overlap of 75%. Next, individual bins were grouped into 22 non-overlapping, logarithmically spaced analysis channels. The envelope of

TABLE I
DEMOGRAPHICS OF NH LISTENERS.

ID	Sex	Age	Stroop Median Time (s)	Stroop *T*-value (age norm)	GOESA (dB SNR for 50%SRT)	Attended Story	Stimulation Order
NH1	M	33	92	46.6		F	Orig/Voc
NH2	M	28	79	50.0	-4.40	M	Orig/Voc
NH3	M	24	56	62.0	-5.20	F	Orig/Voc
NH4	F	28	68	55.0	-5.30	M	Orig/Voc
NH5	M	25	59	60.0	-6.50	M	Voc/Orig
NH6	F	28	55	62.5		M	Orig/Voc
NH7	M	30	92	46.6	-6.70	M	Voc/Orig
NH8	M	18	64	52.0	-5.90	F	Voc/Orig
NH9	M	23	79	50.5	-5.20	F	Voc/Orig
NH10	F	28	62	58.0	-6.60	F	Orig/Voc
NH11	F	23	59	61.0	-5.50	F	Orig/Voc
NH12	F	19	65	58.5	-5.70	M	Voc/Orig

Note. Stroop = Color-Word Interference Test after Stroop, here: interference subtest. HSM = Hochmair-Schulz-Moser sentence test (+10 dB SNR). GOESA = Goettinger sentence test in adaptive noise. F = Female. M = Male. Orig = Original speech quality. Voc = Vocoder speech quality.

TABLE II
DEMOGRAPHICS OF CI USERS.

ID	Sex	Age	Etiology	Age at Onset of Profound Deafness (years) (Left/Right)	Duration of Deafness (months) (Left/Right)	CI Experience	Stroop Median Time (s)	Stroop *T*-Value (age norm)	HSM in Noise (%)	GOESA (dB SNR for for % SRT)	Attended Story
CI1	M	80	Acute HL	66/66	5/108	170/67			45.28		F
CI2	M	67	Genetic	59/25	1/266	103/237			70.00		M
CI3	M	66	Unknown	61/61	26/8	29/47			60.00	6.00	F
CI4	F	51	Unknown	37/33	50/57	119/160	59	65.5	85.00	-.90	M
CI5	F	56	Unknown	47/47	14/1	94/107	81	55.8	64.15	5.50	M
CI6	F	49	Unknown	42/42	1/1	72/72	76	55.0	71.70	-.40	M
CI7	M	49	Unknown	16/16	292/185	95/202	90	51.0	89.62	2.10	M
CI8	F	47	Unknown	1/46	503/0	64/6	70	58.0	73.60	10.90	F
CI9	M	68	Unknown	58/47	31/147	92/108	93	56.6	69.81	3.4	F
CI10	M	69	Unknown	65/59	2/63	50/61	72	65.0	94.34	5.90	F
CI11	M	69	Acute HL	59/49	1/142	119/102	100	55.0	80.19	5.50	F
CI12	F	48	Unknown	46/47	1/1	17/7	81	52.8	96.23	9.80	M

Note. Stroop = Color-Word Interference Test after Stroop, here: interference subtest. HSM = Hochmair-Schulz-Moser sentence test (+10 dB SNR). GOESA = Goettinger sentence test in adaptive noise. F = Female. M = Male. HL = Hearing loss.

each channel was computed on a frame-by-frame basis by computing the square root of the total energy in the channel. Noise bands were amplitude modulated based on the envelope computed in each channel. Noise bands were generated by filtering white Gaussian noise through a filter bank having the same center frequencies as the analysis bands used by the CI processing. The rate of the drop-off of the noise bands away from the center frequency was set to 25 dB/octave to simulate the effect of spread of excitation that may occur in an electrically stimulated cochlea.

E. EEG Recording

Continuous EEG data were recorded in an electromagnetically shielded booth using a BrainAmp System (BrainProducts GmbH, Gilching, Germany) and 96 Ag/AgCl electrodes mounted in a customized, infracerebral electrode cap with an equidistant electrode layout (Easycap GmbH, Herrsching, Germany). The nose tip was used as reference and the ground was placed slightly anterior to Fz. Recordings were performed with a sampling rate of 1000 Hz and an online filter of .02 to 250 Hz. Impedances of the electrodes were maintained below 20 kΩ before data acquisition.

F. Pre-processing

EEG data were pre-processed offline using MATLAB (Mathworks Inc., Natick, MA) and the MATLAB toolbox EEGLAB [9], following the procedure of [27] and [24]. Accordingly, raw data were band-pass filtered (2-8 Hz) and down-sampled (64 Hz), before they were subjected to speech reconstruction. Speech envelopes of the two narratives were obtained by applying the Hilbert transform via an FFT on the respective speech streams, that is, original female, original male, vocoder female and vocoder male. In a further step, a low-pass filter (8 Hz) was applied and the signal was similarly down-sampled to 64 Hz.

G. Speech Reconstruction

The speech reconstruction followed the process described in [24], [27] and [2]. According to this, the EEG data were used to reconstruct an estimate of the attended speech signal using a linear reconstruction model. In a first step, the pre-processed EEG data were segmented using an EEG rectangular analysis window of 60 s. A single trial is defined by this window length. Accordingly, there were 48 trials per CI user and 24 trials per condition (original, vocoder) for the NH listeners.

Previous research on NH individuals [19] has shown that EEG activity reflects the envelope of the speech approximately at time lags from around 100 ms to around 250 ms. However, this is not known for CI users. Accordingly, the best time-lag was explored between 0 and 607 ms using a lag window length of 16 ms. This is a non-overlapping sliding window to cover the range from 0 to 607 ms. The first window covers therefore from 0 to 15 ms. A window length of 16 ms was chosen, as it provided the best results when comparing windows of 16, 32 and 64 ms duration.

The neural response at time sample $k = 0 \dots K-1$ of the electrode $n = 0 \dots N-1$ is denoted as $y_n[k]$; the spatio-temporal filter, also termed decoder, at specific time $l = 0 \dots L-1$ and electrode n is denoted $w_{n,l}$. The reconstructed attended signal (same process for the unattended signal) is estimated as follows:

$$\hat{x}_{a,u}[k] = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} w_{n,l} y_n[k+\Delta+l], \tag{1}$$

where $\hat{x}_{a,u}[k]$ denotes the reconstructed attended or unattended signal at time sample $k = 0 \dots K-1$, and Δ models the latency or lag. In vector notation would be:

$$\hat{x}_a = \mathbf{W}_a^T \mathbf{Y}[k], \tag{2}$$

$$\hat{x}_u = \mathbf{W}_u^T \mathbf{Y}[k]. \tag{3}$$

In the next subsections the sub-indices a and u are omitted to simplify the notation and $\mathbf{W}$ and $\mathbf{Y}$ are defined as:

$$\mathbf{W} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \dots \ \mathbf{w}_N^T]^T \tag{4}$$

$$\mathbf{w}_n = [w_{n,0}^T \ w_{n,1}^T \ \dots \ w_{n,L-1}^T]^T \tag{5}$$

$$\mathbf{Y}[k] = [\mathbf{y}_1[k]^T \ \mathbf{y}_2[k]^T \ \dots \ \mathbf{y}_N[k]^T]^T \tag{6}$$

$$\mathbf{y}_n[k] = [y_n[k+\Delta] \ y_n[k+\Delta+1] \ \dots \ y_n[k+\Delta+L-1]] \tag{7}$$

During training, the filter $\mathbf{W}_a$ is estimated using least squares error between $x_a[k]$ and $\hat{x}_a[k]$:

$$J_{LS}(\mathbf{W}_a) = E\{|x_a[k] - \mathbf{W}_a^T \mathbf{Y}[k]|^2\}, \tag{8}$$

where $x_a[k]$ is the Hilbert envelope extracted from the attended audio signal. To avoid overfitting, regularization is applied, using the norm of the coefficients

$$J_{RLS}(\mathbf{W}_a) = E\{|x_a[k] - \mathbf{W}_a^T \mathbf{Y}[k]|^2\} + \lambda \mathbf{W}_a^T \mathbf{W}_a, \tag{9}$$

with λ being the regularization parameter.

Minimizing the previous Equation by applying the derivative with respect to the coefficients $\mathbf{W}_a$ results in:

$$\mathbf{W}_a^T = (R_{x_a Y} + \lambda I)^{-1} \cdot R_{yy} \tag{10}$$

where $R_{x_a Y}$ and R_{YY} are defined as follows:

$$R_{x_a Y} = \sum_{m=0}^{K} x_a[k] y[k+m], \tag{11}$$

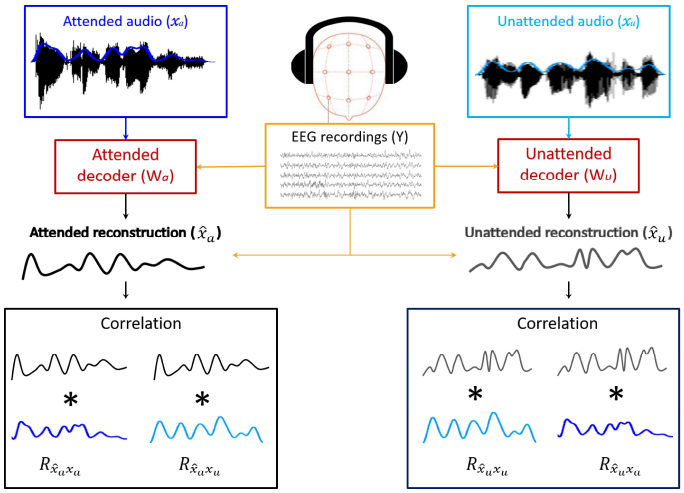


Fig. 1. Decoding strategy illustration. Data from all electrode channels are decoded simultaneously to give an estimate of the amplitude envelope of the input speech stream. The correlation between the reconstructed signals $\hat{x}_{a,u}$ and both the attended and unattended speech streams $x_{a,u}$ is then calculated for each trial, following a leave-one-out cross-validation approach.

$$R_{YY} = \sum_{m=0}^{K} y[k] y[k+m]. \tag{12}$$

Note that the filter or decoder $\mathbf{W}_u$ can be obtained following the same procedure as used to estimate $\mathbf{W}_a$.

Selective attention is decoded based on the correlation coefficient between the reconstructed attended and the original attended $C_{x_a \hat{x}_a}$ signal and the correlation coefficient between the reconstructed attended and the original unattended signal $C_{x_u \hat{x}_a}$ for each lag Δ:

$$C_{x_a \hat{x}_a} = \frac{E\{(x_a[k] - \mu_{x_a[k]})(\hat{x}_a[k] - \mu_{\hat{x}_a[k]})\}}{\sigma_{x_a[k]} \sigma_{\hat{x}_a[k]}}, \tag{13}$$

$$C_{x_u \hat{x}_a} = \frac{E\{(x_u[k] - \mu_{x_u[k]})(\hat{x}_a[k] - \mu_{\hat{x}_a[k]})\}}{\sigma_{x_u[k]} \sigma_{\hat{x}_a[k]}}. \tag{14}$$

The highest correlation coefficient with the reconstructed signal, i.e. $\arg\max_{x_a, x_u}(C_{x_a \hat{x}_a}, C_{x_u \hat{x}_a})$ indicates which is the attended source by the listener. This procedure is repeated for all trials (i.e. 24 times for NH original speech and vocoder and 48 times for CI users). The accuracy for detecting the attended speaker is obtained as the number of times that the signal is correctly decoded divided by the total number of trials. Note that the unattended decoder can also be used to decode both the attended and the unattended speech stream, however, it typically delivers lower accuracies than the attended decoder [27].

A classical leave-one-out cross-validation approach was used to train and test the decoder. Each test-trial was evaluated using the (averaged) decoder obtained from the average of the decoders trained on every other trial. Figure 1 shows an illustration of the decoding process.

III. Results

A. Behavioral Data

Overall, behavioral data from the questionnaire show that participants followed the instructions and attended the correct

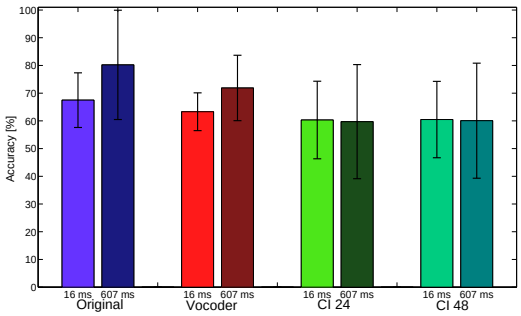


Fig. 2. Average decoding accuracies and standard deviation of the mean across subjects for original and vocoded speech using 24 minutes of recording in NH listeners, or 24 and 48 minutes in CI users (CI 24 and CI 48). Accuracy values were obtained using the attended decoder. For each condition, the results were obtained with a long lag window of 607 ms or averaging the decoding accuracy obtained with short non-overlapping lag windows of 16 ms covering the range from 0 to 607 ms.

story. NH listeners achieved a mean of 85 % (standard error of mean; SEM: 3%) correctly answered questions when listening to the original speech condition. For the vocoded speech condition, the performance dropped to 79% (SEM: 4%), which is still significant above chance level. The difference in performance for the original and vocoded speech condition in NH listeners was not statistically significant. CI users achieved a mean of 56% (SEM: 5%) correctly answered questions, which was also significant above chance level. Accordingly, NH listeners performed significantly better than CI users even in the vocoded speech condition ($t(22)$ = -3.84, p = 0.001). Accuracy for the questions related to the unattended story was below chance level in any of the groups, indicating that participants were only guessing the answers to the unattended story and did not follow it.

B. Overall Decoding Accuracy

Average decoding accuracies across NH subjects in the original and the vocoder condition and across CI users are shown in Figure 2 using the attended decoder with a long lag window of 607 ms and with a short lag window of 16 ms. For the long window only one lag was computed covering the range from 0 to 607 ms. For the short window, the accuracy results were averaged for each decoder covering the range from 0 to 607 ms in windows of 16 ms. The regularization parameter λ was set to 0.001 as it gave the best accuracy results for the NH, vocoder and CI group. Figure 9 in section B of the Appendix presents the accuracy results for different values of the regularization parameter λ.

Figure 3 presents the accuracy values across lags using 16 ms window for NH, vocoder and CI users. The left panels present the accuracy values obtained with the attended decoder predicting the attended speech, whereas the right panels show the accuracy values using the unattended decoder to predict the unattended speech. Chance level (29.2% - 70.8%) was determined using a binomial test at 5% significance level. Figure 3 shows a maximum decoding accuracy of 80.9% (lag 250-266 ms), 73.2% (lag 218-234 ms) and 71.5% (lag 410-426 ms) for the NH listeners in the original and vocoder condition,

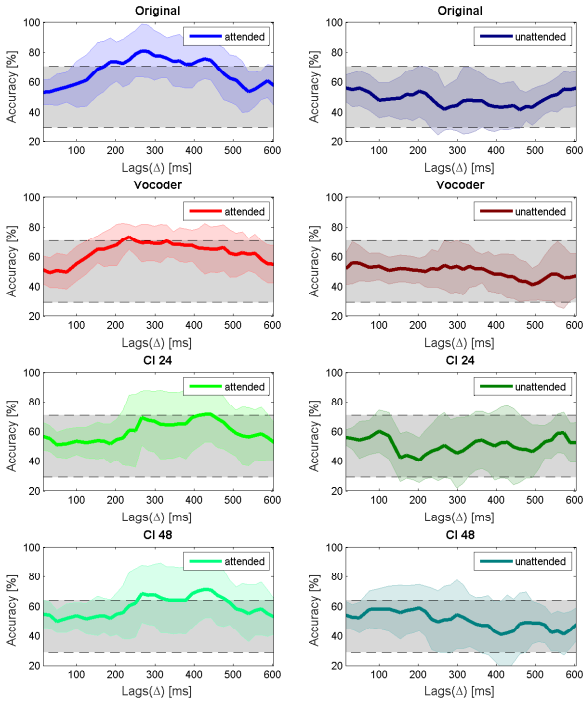


Fig. 3. Average decoding accuracies and standard deviation of the mean for original and vocoded speech using 48 minutes in NH listeners, 24 and 48 minutes in CI users (CI 24 and CI 48). Accuracy values obtained using the attended decoder, predicting the attended speech are shown in the left part in light color, while accuracy values obtained using the unattended decoder, predicting the unattended speech are shown in the right part in darker color. Note that the accuracy values obtained by predicting the attended or unattended signal with a particular decoder (attended or unattended) are complementary. The correlation coefficients for the CI 48 and the CI 24 minutes condition were very similar.

and the CI 24 group, respectively. The peak accuracy for CI users was just below chance level, whereas the accuracy for NH listeners in the original and vocoder condition was significantly above chance level. For both, the original and vocoder condition, the highest accuracy was obtained at time lags ranging from 256 to 272 ms. These results are consistent with previous works from [24] and [27]. In contrast to the NH listeners, CI users showed two main peaks. A first one at the same time lag ranging from 256 to 272 ms and a second peak at a time lag ranging from 416 to 432 ms. As only the attended decoder achieved results above chance level, the unattended decoder was excluded from the statistical analysis. For the attended decoder a repeated measures analysis of variance (ANOVA) with the within-subject factors Lag (Δ) and Condition (original, vocoder) revealed a significant main effect of Lag ($F(34,374)$=9.69; p<0.001) and a marginal significant Lag x Condition interaction ($F(34,374)$=1.44; p=0.058). The main effect Condition was not significant. From these results, one could conclude that decreasing spectral resolution with a vocoder may cause a drop in decoding selective attention for certain lags.

A mixed ANOVA with within-subject factor Lag (Δ) and

the between-subject factor Group (original, CI-24) confirmed a significant main effect of Lag ($F(34,748)=9.76$; $p<0.001$) and a significant Lag x Group interaction ($F(34,748)=2.37$; $p<0.001$), but no significant main effect of Group. The interaction effect means that for specific lags the decoding accuracy of each group was significantly different from each other.

An additional mixed ANOVA with within-subject factor Lag (Δ) and the between-subject factor Group (vocoder, CI-24) confirmed a significant main effect of Lag ($F(34,748)=7.00$; $p<0.001$) and a significant Lag x Group interaction ($F(34,748)=1.93$; $p=0.001$), but no significant main effect of Group. The interaction effect means that for specific lags the decoding accuracy between vocoder and CI was significantly different from each other.

From these results, one could conclude that the drop in accuracy observed in the CI group may be caused by the lack of spectral resolution known from the vocoder condition, however one cannot yet exclude that the electrical artifact introduced by the CI also decreases the accuracy of detecting selective attention.

C. Effect of Amount of EEG Recording Time in CI Users

Selective attention accuracy in CI users was computed for two different amounts of recording time: 24 and 48 minutes. The two bottom panel rows in Figure 3 show the mean accuracy across lags for the CI group when training the decoder with either 24 or 48 minutes. The left panel presents the mean accuracy values obtained with the attended decoder predicting the attended speech, whereas the right panel shows the accuracy values using the unattended decoder to predict the unattended speech. A mixed ANOVA with factors Lag (Δ) and Condition (24-minutes decoder, 48-mintues decoder) showed a significant main effect of Lag ($F(34,374)=4.22$; $p<0.001$), no significant effect of Condition and no significant interaction. This suggests that increasing the amount of recording time beyond 24 minutes does not improve decoding accuracy in CI users. However, training the decoder with 48 minutes results in a lower chance level than with 24 minutes of training. Accordingly, the results, in contrast to the 24 minutes condition, become significant above chance level. These results are consistent with results observed in NH listeners where an increase in recording time beyond 24 minutes did not significantly improve selective attention accuracy [24]. All in all, the results provide some evidence that selective attention decoding in CI users by means of EEG is feasible, given a sufficient recording time.

IV. Further Analysis

A. Cortical Activity and Electrical Artifact: Analysis of Correlation Coefficients

It is important to analyze the influence of the electrical artifact created by the CI in the EEG recording. In theory the artifact should be related to some extent to the incoming sound, as the CI transmits envelope information in each electrode. Therefore, it is hypothesized that the EEG recordings from CI users may be stronger correlated with the incoming sound than for NH listeners in the original or the vocoder condition.

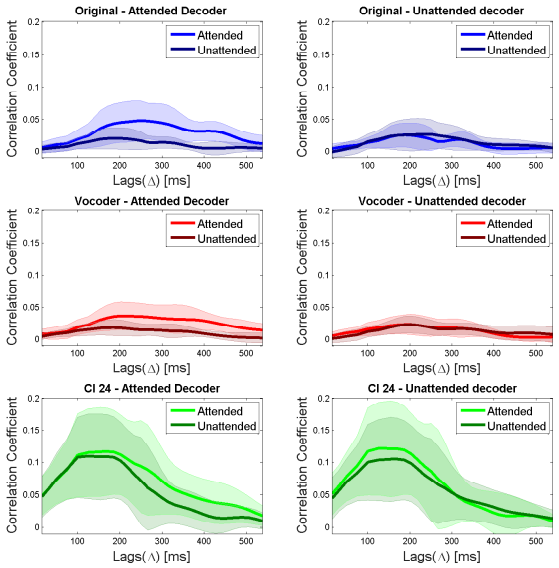


Fig. 4. Correlation coefficients for NH listeners in the original (blue), and vocoder (red) condition; and for the CI users (green). In each subplot lighter colors refer to the correlation coefficients for the attended speech and darker colors refer to the correlation coefficients for the unattended speech. The thick lines represent the mean values and the shaded areas the standard deviation across subjects.

This hypothesis was investigated by comparing the correlation coefficients between the original and the reconstructed sound using the attended ($C_{x_a\hat{x}_a}$ and $C_{x_u\hat{x}_a}$; Left panels of Figure 4) and the unattended decoder ($C_{x_u\hat{x}_u}$ and $C_{x_a\hat{x}_u}$; Right panels of Figure 4) for each group (original, vocoder, CI). Correlation coefficients were estimated for each time lag Δ by averaging them across subjects.

Figure 4 shows that the difference in correlation coefficient between the attended and the unattended signals is larger for the attended than for the unattended decoder. In NH listeners (original and vocoder condition) the highest correlation values are obtained between 200 and 300 ms, while CI subjects obtain the highest correlation coefficients between around 100 and 200 ms. However, for CI users the largest difference in correlation values between the attended and the unattended speech is achieved for lags ranging from 250 to 350 ms and from 400 to 450 ms, which correspond to the highest accuracies reported in Figure 3. The large correlation values observed in CI users compared to those of the NH listeners in the original and the vocoder conditions may indicate that the EEG recordings contain artifacts related to the envelope of the incoming speech. If the EEG signal contains mainly artifact related to the incoming sound, the decoder is in practice conducting an autocorrelation of the speech envelope. The known rapid decrease in the speech autocorrelation explains the decrease in the correlation coefficient with increasing lag after around 200 ms shown in Figure 4.

To test the previous hypothesis, the decoder was trained with a simulated EEG signal where the 96 electrode recordings were replaced by the sum of the two original speech sounds processed by a CI (Figure 5). This situation models a worst

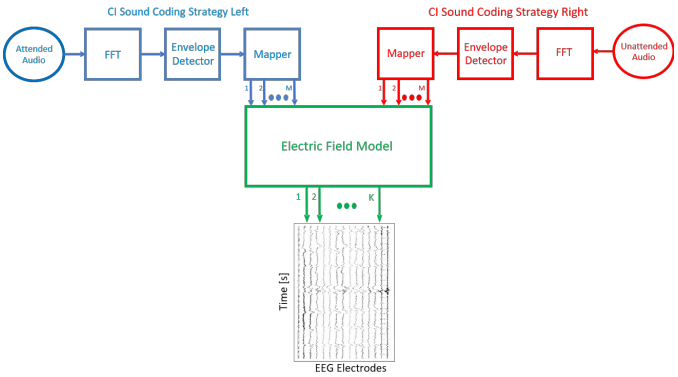


Fig. 5. Block diagram of the electrical artifact model and the sound coding strategy used in the left and right CI sound processor.

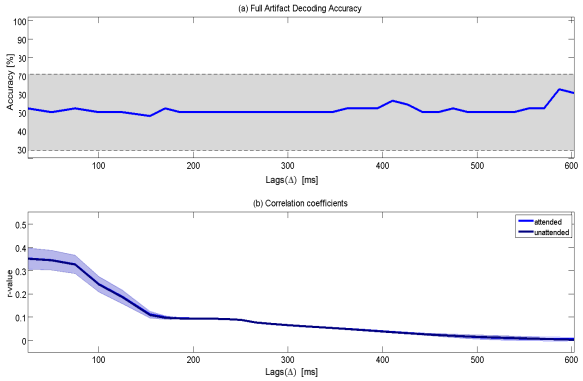


Fig. 6. a) Decoding accuracy using the attended decoder trained with a simulated EEG signal containing only artifact. b) Correlation coefficients for the attended speech (light blue) and the unattended speech (dark blue) using the attended decoder. The shaded area denotes the standard deviation across trials (48 trials for 48 minutes) and the dark line shows the mean value. Note, the correlation coefficients for the attended and the unattended speech sounds overlap. Both, correlation coefficients for the attended and the unattended speech sounds drop rapidly with increasing lag, reaching a value of 0.1 after around 150 ms.

case scenario where full artifact caused by both CIs reaches all 96 EEG electrodes simultaneously. The same sound coding strategy implemented in the vocoder was used in this model. Note that only the latency introduced by the sound coding strategy, i.e. the algorithmic latency, is considered without modeling the latency introduced by the CI electronics. The sound coding strategy model was used to estimate the current delivered to each electrode over time for the attended and the unattended speech signal. For each EEG electrode e located at position x_e, y_e, z_e, the voltage was estimated using the analytical solution for the voltage in a medium due to current I_i applied on a CI electrode i located at position x_i, y_i, z_i:

$$V_e(x_e, y_e, z_e) = \frac{I_i}{4\pi\sigma\sqrt{(x_i - x_e)^2 + (y_i - y_e)^2 + (z_i - z_e)^2}} \quad (15)$$

It was assumed that the system was linear and therefore the voltage created by each CI electrode i on each side (left and right) was added up to each other to estimate the electrical artifact in the EEG. The conductivity of the brain was set to σ=0.33 S/m [41]. The simulated EEG signal was then used to obtain the attended decoder and to reconstruct the original signals. Figure 6 presents the accuracy values and the corresponding correlation coefficients for time lags between 0 and 607 ms for the attended decoder simulating full artifact in the EEG recording. Figure 6a shows that the decoding accuracy never exceeds the chance level if the EEG recording only contains artifact as simulated by the model. Moreover, Figure 6b shows that the correlation coefficients between the reconstructed speech and the attended or the unattended speech are almost identical (both curves show a complete overlap). The correlation coefficients remain almost constant up to a lag of 80 ms, but decay very fast thereafter, reaching a value of around 0.1 at 150 ms. Beyond these time lags, the correlation coefficient decay slows down, reaching a value of around 0 at approximately 600 ms, i.e. here the reconstructed and the original attended or unattended speech signals become uncorrelated. Compared to the strong correlation coefficient decay observed in the model results, the correlation coefficient in CI users decayed much slower, specially for time lags of 250 ms (Figure 4). This slower decay may be explained by the cortical activity overlaid with the electrical artifact. It is also interesting to note that the results in NH and CI users show an increased difference between the correlation coefficients for the attended and the unattended speech for time lags of 200-250 ms (Figure 4). In the modeled results, however, no difference in correlation coefficients between the attended and the unattended speech was observed, as both coefficients are fully contaminated by the artifact and the cortical activity is not modeled. Based on the present analysis, we suggest that selective attention in CI users can be successfully decoded from cortical activity for lags beyond around 200-250 ms.

B. *Electrical Artifact*

To give more insight about the CI electrical artifact, the top panels of Figure 7 present the power spectral density for each EEG electrode for a single trial (1 minute recording time) from 0 Hz to 12 Hz for the NH (original speech) and for the CI listener groups. For both groups, it can be observed that the power spectral density is dominated by artifacts caused by eye movements. On the bottom panels the power spectral density is shown after applying independent component analysis (ICA) to remove eye artifacts. It can be observed that the power spectral density for the CI group is larger than for the NH group, probably because it contains CI electrical artifact. Moreover, the power spectral density in CI users shows larger variability than for the NH listeners. The topographical maps above the power spectral density curves demonstrate that indeed for the CI group, the EEG electrodes with locations close to the CI present much higher amplitude than the other EEG electrodes.

V. Discussion

This study investigated whether selective attention can be successfully decoded by means of single-trial EEG data in CI

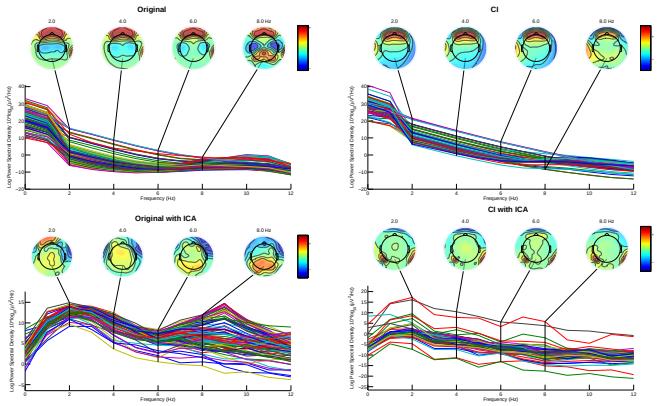


Fig. 7. Power spectral density of each EEG electrode across frequency. The colored curves present the power spectral density averaged across subjects for each EEG electrode. The topographical maps present the power spectral density across head location. The left panels present the results for NH subjects using original speech, whereas the right panels present the results for CI users. On the top panels no artefact removal is applied and in the topographical maps one can observe a large power in the electrodes located on the eyes. The topographical maps after applying independent component analysis (ICA) are presented on the bottom panels to improve the analysis with reduced eye artefact.

users. Two main obstacles were expected for this task. First, the lack of spectral resolution in CI users and second, the electrical artifact introduced by the CI in the EEG signal, which may cover the true cortical activity. Results of the current study demonstrate that selective attention can be successfully decoded in CI users although the accuracy is lower than in NH subjects. The results provided evidence that the lack of spectral resolution has a negative effect on the decoding accuracy of selective attention, but also the artefact might likely contribute to the observed decrease in decoding accuracy.

The effect of a reduced spectral resolution without the additional effect of an electrical artifact on the decoding accuracy was investigated in NH subjects listening to both, original and vocoded sounds. Decoding accuracy dropped from 80.9% in the original condition to 73.2% in the vocoded condition. Previous studies have shown that a degradation of frequency information, even if it does not affect speech intelligibility [42], plays an important role in stream separation and strongly effects top-down modulation of neural activity (e.g. [11] [46] [47]). This fact could explain the trend towards poorer decoding accuracies for NH listeners in the vocoder condition compared to the original speech condition, although the vocoder also causes a decrease in speech performance. The mean decoding accuracy observed for NH listeners (80.9%) was well above chance level, but slightly below the accuracies reported by previous studies (88-89%; [24] [27]). This small difference may be caused by two reasons. First, the amount of training time for NH listeners in the current study was only 24 minutes compared to the 30 or 48 minutes used by [24] and [27], respectively. Second, these previous studies presented the two stories always on the same side to each subject, whereas the current study alternated the presentation side.

To investigate the combined effect of reduced spectral resolution and the electrical artifact of the CI on decoding selective attention, we likewise investigated a group of CI users. The decoding accuracy in this population (71.5%) was similar to the one obtained for NH listeners in the vocoder condition (73.2%). To evaluate the effect of amount of training data, the decoding accuracy in CI users was determined for two different recording times: 24 and 48 minutes. Increasing the amount of recording time beyond 24 minutes did not significantly improve the absolute decoding accuracy. It however resulted in a lower chance level, so that the decoding accuracy for 48 minutes became significant above chance level. The current results provide first evidence that selective attention can be successfully decoded in CI users, even if an electrical artifact is present. The fact that the decoding accuracies for NH subjects in the vocoder condition and the ones of the CI users were similar, further suggests that rather spectral smearing and consequently worse speech understanding but not the electrical artifact might be a reason for the decrease in decoding accuracy. No significant correlation between behavioral speech performance and decoding accuracy was found.

Previous studies on cortical responses to attended continuous speech in NH listeners have reported peak latencies or lags of the highest decoding accuracy ranging from short delays from 100 to 150 ms ([8], [17]) to longer delays from 150 to 300 ms (e.g. [1], [27], [24]). NH listeners (original and vocoder condition) in the present study showed the maximum decoding accuracy at lags from 220 to 270 ms, which is consistent with these previous findings. For the CI users, however, two main peaks for the decoding accuracy were observed. The first one occurring at a time lag from 256 to 272 ms and a second one occurring at a time lag from 416 to 432 ms. In summary, the uracy in CI users is lower and delayed (second peak) in comparison to NH listeners (original and vocoder condition).

Although there is no clear evidence that different peaks may relate to word processing (e.g. [51] some direct efforts made an attempt to relate EEG/MEG to word and conceptual processing that go beyond envelope measures (e.g. [50]). In the present work it has been shown that CI users present two main peaks, the first one at a time lag ranging from 256 to 272 ms and the second at a time lag ranging from 416 to 432 ms. Both peaks also seem to be present in NH, but are just not as prominent. However, in NH the first peak is always higher than the second one in both the original and the vocoder conditions. This observation is in agreement with the results reported by [24] which suggested that the first peak may be related to word processing while the second one may be related to sentence/conceptual processing. In our results, we observed that in NH the second peak is lower in vocoded than in the original speech. Interestingly, the double peak structure is more prominent in the CI group. In CIs it is possible that the first peak is lower than in NH listeners because of the spectral degradation they receive, but the second peak is still higher since the understanding is almost as good as in NH listeners (note that the accuracy, while low is still just as high as in NH at this lag). Also, since the CI users are more used to spectral degradation they may be more successful in dealing with it than the NH group in the vocoder condition.

The large correlation between the EEG reconstructed and the original speech observed in CI users compared to the correlation for NH listeners in both, the original and vocoder

condition may indicate that the EEG recordings contain artifact related to the envelope of the incoming speech. However, the correlation drops rapidly with increasing lags. This decrease with lag may be expected, as the autocorrelation of speech decreases rapidly. If the EEG signal contains mainly artifact related to the incoming sound, the decoder is in practice conducting a correlation of the signal with itself. This effect was demonstrated using a decoder trained with simulated EEG signals where the 96 electrode recordings were replaced by the sum of the two original speech sounds used in the selective attention experiment. Results in this simulation demonstrate that the decay in correlation coefficient with increasing lag was faster than the decay observed in real EEG recordings in CI subjects. From this analysis, we suggest that the effect of the artifact on the correlation coefficient used to detect selective attention is small in comparison to the effect of the cortical response for the range of lags (200-250 ms) shown to provide highest accuracy in NH subjects.

Previous work has shown that the ability to pay auditory selective attention is not predicted by age (e.g. [43]), although care has to be taken when mixing data from different age populations ([49]). Section A in the Appendix presents results comparing selective attention decoding in the same young group of NH listeners with an elderly group of normal hearing listeners. The decoding accuracy in both groups was similar and in both cases the accuracy was higher than the accuracy obtained by normal hearing listeners using the vocoder or by CI users.

The selective attention decoding process is a very promising method to improve speech intelligibility in hearing impaired people. This method could be used to steer signal processing algorithms (such as beamformers, noise reduction algorithms) and it could be combined with source separation algorithms [26]. However, many other aspects must be investigated before being able to use these methods in daily life devices. Although the effect of the CI artifact in decoding selective attention seems to be smaller than expected, there is still need to further investigate its effects and keep it as minimal as possible. For example, low stimulation rate sound coding strategies or high EEG sampling rates could be used to minimize the CI artifact [48]. ICA has been suggested as a successful method for artifact removal [40] [41]. However, this method requires manual and subjective decisions and is a time-consuming technique. Moreover, the current study has used a very simple paradigm to simulate the cocktail party effect. The results presented here are therefore not easily applicable to daily life situations and more realistic sound environments need to be explored to better understand the selective attention processes in both NH and CI users [7]. It is important to consider that selective attention decoding was performed using a high-density EEG consisting of 96 electrodes. New technologies with a minimized number of EEG electrodes could be used to overcome the lack of portability and the long set-up process of the EEG cap. The potential use of an around the ear device (cEEGrid, [6] [25]), an in the ear canal device [5] [10] or of a device that measures cortical potentials through the CI via intracochlear electrodes [25] [23] or of additional electrodes implanted during the CI surgical procedure offers promising methods that however need further research to reach a more complete portable system.

Conclusion

This work has shown that it is possible to decode selective attention in CI users. Two main limitations were foreseen, on the one hand the worse spectral resolution obtained by CI users and on the other hand the electrical artifact introduced by CIs in the recorded EEG. A reduction in spectral resolution modeled by presenting vocoded sounds to NH listeners caused a decrease in decoding accuracy. The electrical artifact became less relevant to decode selective attention with increasing the delay or lag in the recorded EEG signal. This fact enables decoding selective attention in CI users if sufficient training data is available.

Acknowledgment

The authors would like to thank Hanna Dolhopiatenko for her support in analyzing the data and the subjects who have participated in the experiments. We are also thankfukl to Pascale Sandmann and Christoph Kantzke for fruitful initial discussions and to Mayra Windeler for her support in measurements. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2177/1 - Project ID 390895286.

References

[1] Aiken, S. J., Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. Hearing Research, 245, 1-2, 35-47.

[2] Aroudi, A., Mirkovic, B., De Vos. M., Doclo, S. (2016). Auditory attention decoding with EEG recordings using noisy acoustic reference signals, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 694-698.

[3] Bümler, G., (1984). Farbe-Wort-Interferenztest (FWIT) nach JR Stroop: Handanweisung.

[4] Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am., 25, 975979.

[5] Bleichner, M. G., Lundbeck, M., Selisky, M., Minow, F., Jager, M., Emkes, R., et al. (2015). Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? Physiol. Rep. 3:e12362. doi: 10.14814/phy2.12362

[6] Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015). Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. Sci. Rep. 5, 16743. doi: 10.1038/srep16743

[7] Das, N., Van Eyndhoven. S., Francart, T. , Bertrand, A. (2016). Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses. IEEE 38th Annual Int. Conf. of the Engineering in Medicine and Biology Society (EMBC), 7780.

[8] Ding, N., Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. National Academy of Sciences, 109, 29, 11854–11859.

[9] Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, Journal of Neuroscience Methods, 134, 1, 9-21.

[10] Fiedler, L., Obleser, J., Lunner, T., Graversen, C. (2016) Ear-EEG allows extraction of neural responses in challenging listening scenariosa future technology for hearing aids?. 2016 IEEE 38th Annual Int. Conf. of the Engineering in Medicine and Biology Society (EMBC), 5697700.

[11] Fritz, J., Elhilali, M., Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hearing Research, 206, 159176.

[12] Fu, Q.-J., and Nogaki, G. (2005). Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing. JARO: Journal of the Association for Research in Otolaryngology, 6(1), 1927. http://doi.org/10.1007/s10162-004-5024-3

[13] Gilley, P.M., Sharma, A., Dorman, M., Finley, C.C., Panch, A.S., Martin, K.(2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. Clin. Neurophysiol. 117, 17721782. doi:10.1016/j.clinph.2006.04.018

[14] Hahlbrock, K.H. (1970). Sprachaudiometrie: Grundlagen und praktische Anwendung einer Sprachaudiometrie fr das deutsche Sprachgebiet. Georg Thieme Verlag, Stuttgart.

[15] Hochmair-Desoyer, I., Schulz, E., Moser, L., Schmidt, M. (1997). The HSM Sentence Test as a Tool for Evaluating the Speech Understanding in Noise of Cochlear Implant Users. Am. J. Otol. 18, S83.

[16] Kollmeier, B., Wesselkamp, M., (1997). Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. J. Acoust. Soc. Am., 102, 24122421. doi:10.1121/1.419624

[17] Koskinen, M., Seppa, M. (2014). Uncovering cortical MEG responses to listened audiobook stories, Neuroimage, 100, 26370.

[18] Krueger, B., Joseph, G., Rost, U., Strau-Schier, A., Lenarz, T., Buechner, A. (2008). Performance Groups in Adult Cochlear Implant Users: Speech Perception Results From 1984 Until Today. Otol. Neurotol. 29, 509512.

[19] Lalor, E. C. and Foxe, J. J. (2010), Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience, 31, 189-193.

[20] Loizou, P.C. (1998). Mimicking the human ear. IEEE Signal Process. Mag. 15, 101130.

[21] Martin, B. A. (2007). Can the Acoustic Change Complex Be Recorded in an Individual with a Cochlear Implant? Separating Neural Responses from Cochlear Implant Artifact. J Am Acad Audiol 18, 126140.

[22] McLaughlin, M., Lopez Valdes, A., Reilly, R.B., Zeng, F.G. (2013). Cochlear implant artifact attenuation in late auditory evoked potentials: A single channel approach. Hear. Res. 302, 8495.

[23] McLaughlin, M., Lu, T., Dimitrijevic, A., Zeng, F. G. (2012). Towards a closed-loop cochlear implant system: application of embedded monitoring of peripheral and central neural activity, 20, 4, 443-54.

[24] Mirkovic, B., Debener, S., Jaeger, M., De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. J. Neural Eng. 12, 46007.

[25] Mirkovic, B., Bleichner, M. G., De Vos, M. and Debener, S. (2016). Target Speaker Detection with Concealed EEG Around the Ear, Front. Neurosci., 10, 349, .

[26] Nogueira, W., Gajcki, T., Krger, B., Janer, J., Bchner, A. (2016). Development of a Sound Coding Strategy based on a Deep Recurrent Neural Network for Monaural Source Separation in Cochlear Implants, ITG Speech Communication, Padeborn, Germany.

[27] O' Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb. Cortex, 25, 16971706.

[28] Sandmann, P., Eichele, T., Buechler, M., Debener, S., Jncke, L., Dillier, N., Hugdahl, K., Meyer, M. (2009). Evaluation of evoked potentials to dyadic tones after cochlear implantation. Brain, 132, 19671979. doi:10.1093/brain/awp034

[29] Shannon, R. V, Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. Science, 270, 5234, 303-304.

[30] Shannon R. V., Fu Q. J., Galvin III J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation, Acta oto-laryngologica, 124, 50-54.

[31] Viola, F.C., Thorne, J.D., Bleeck, S., Eyles, J., Debener, S. (2011). Uncovering auditory evoked potentials from cochlear implant users with independent component analysis. Psychophysiology 48, 14701480.

[32] Wagner, L., Maurits, N., Maat, B., Baskent, D., Wagner, A.E. (2018). The cochlear implant EEG artifact recorded from an artificial brain for complex acoustic stimuli. IEEE Trans. Neural Syst. Rehabil. Eng.

[33] Wilson, B.S., Dorman, M.F. (2008). Cochlear implants: A remarkable past and a brilliant future. Hear. Res. 242, 321.

[34] Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., Rabinowitz, W.M. (1991). Better speech recognition with cochlear implants. Nature 352, 236238.

[35] Wouters, J., McDermott, H.J., Francart, T. (2015). Sound Coding in Cochlear Implants: From electric pulses to hearing. IEEE Signal Process. Mag. 32, 6780.

[36] Zeng, F.-G., Rebscher, S., Harrison, W., Sun, X., Feng, H. (2008). Cochlear Implants: System Design, Integration, and Evaluation. IEEE Rev. Biomed. Eng. 1, 115142.

[37] X. Li et al. (2010), Characteristics of Stimulus Artifacts in EEG Recordings Induced by Electrical Stimulation of Cochlear Implants Biomedical Engineering and Informatics, IEEE

[38] K. Tremblay et al. (1998). The time course of auditory perceptual learning: neurophysiological changes during speechsound training. Neuroreport, 9(16), 3557-3560.

[39] B. Van Dun. (2014). Towards a clinically viable way to record Cortical Auditory Evoked Potentials (CAEPs) in cochlear implant (CI) clients. 8th International Symposium on Objective Measures in Auditory Implants, Toronto, Canada.

[40] Gilley, P. M., Sharma, A, Dorman, M, Finley, C. C., Panch, A. S., Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. Clinical Neurophysiology, 117(8), 1772-1782.

[41] Viola, F. C., De Vos, M., Hine, J., Sandmann, P., Bleeck, S., Eyles, J., Debener, S. (2012). Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials. Hearing research, 284(1), 6-15.

[42] Zion Golumbic, E. M., Poeppel, D. and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective Brain Lang. 122, 15161.

[43] Ruggles, D, Bharadwaj, H, Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. Curr Biol. 22(15):1417-22.

[44] Stothart, G, Kazanina, N. (2016). Auditory perception in the aging brain: the role of inhibition and facilitation in early processing. Neurobiol Aging. 47:23-34.

[45] Woods, D.L. (1992) Auditory selective attention in middle-aged and elderly subjects: an event-related brain potential study, Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 84(5):456-468.

[46] Kong YY, Somarowthu A, Ding N. (2015) Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech. J Assoc Res Otolaryngol. 16(6):783-96.

[47] Ding N, Chatterjee M, Simon JZ. (2013). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage. 88:41-6.

[48] Somers, B. Verschueren, E., Francart, T. (2018), Neural Tracking of the Speech Envelope in Cochlear Implant Users, J. Neural. Eng. (In Press).

[49] Karns, C. M., Isbell, E., Giuliano, R. J., Neville, H. J. (2015). Auditory attention in childhood and adolescence: An event-related potential study of spatial selective attention to one of two simultaneous stories, Develpmental Cognitive Neuroscience, 13. 53-67.

[50] Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech, 28(5): 803-809.

[51] Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach, Clinical Neurophysiology, 118(2):237-254.

[52] Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Frontiers in human neuroscience, 10, 604. doi:10.3389/fnhum.2016.00604.

APPENDIX

A. *Effect of age in decoding selective attention*

The NH and the CI group that participated in current the study were not balanced in age (NH mean age: 25.58 years vs CI mean age: 59.9167). A previous study recommended to be careful combining data from different age populations [49]. For this reason, the selective attention decoding was repeated following the same procedure in a new group of 4 NH listeners with a mean age of 75 years. Figure 8 presents decoding accuracy for both groups of NH listeners. From these

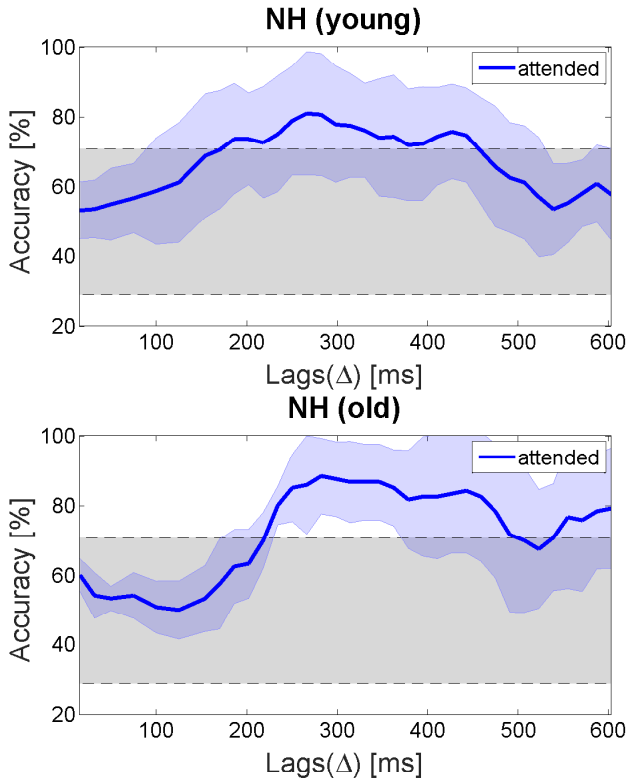


Fig. 8. Decoding accuracy for two groups of NH listeners with different mean ages across lags.

pilot results it can be seen that there is no large difference in decoding selective attention between both groups differing in age. These results are in agreement with previous works in which selective attention ability is not predicted by age [43].

B. *Effect of regularization parameter*

The regularization parameter λ can be used to avoid overfitting in the least squares optimization method used to decode selective attention. The exact value of λ has an influence in the overall accuray and needs to be optimized. Figure 9 presents the effect of the λ on decoding accuracy. The λ value of 0.001 obtained the highest accuracy for all groups.

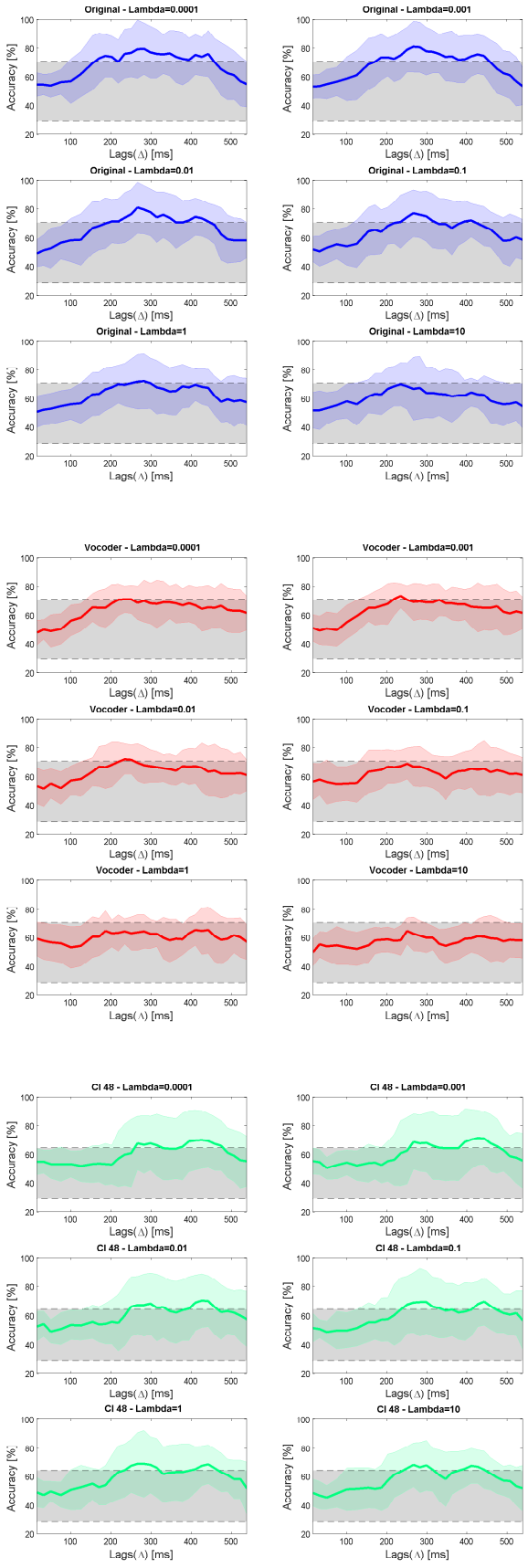
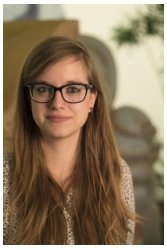


Fig. 9. Decoding accuracy for the normal hearing (NH) listeners using original or vocoder sounds and CI users across lags for different values of the regularization parameter. For the CI group we provide the analysis using 48 minutes.

Waldo Nogueira received his Dipl.-Ing. and Dr.-Ing. degree from the Polytechnic University of Catalonia and the Leibniz University of Hannover (LUH) in 2003 and 2008, respectively. In 2008 he joined the R&D labs of Advanced Bionics in Belgium and Germany. During 2011 he started a Post-Doc at the Music Technology Group of the Pompeu Fabra University in Barcelona. Since 2013 he is Junior Professor at the Hannover Medical School and the Cluster of Excellence Hearing4all. His main research interests focus on audio signal processing for auditory devices, modeling, psychoacoustics and electrophysiology with electric hearing.

Giulio Cosatti received his Master degree in Bioengineering in 2017 and his Bachelor degree in Electronic Engineering in 2014 from the University of Rome 3. He conducted his Master thesis at the German Hearing Center (DHZ) focused on selective attention in CI users. In 2017 he joined the Cluster of Excellence Hearing4all at the Hannover Medical School. Since the end of 2017 he is working on Computer System Validation at the Pharma Quality Europe (PQE).



Irina Schierholz received her Bachelors degree in Biology in 2011 and her Masters degree in Neurocognitive Psychology in 2013 from the University of Oldenburg, Germany. In 2017 she completed her doctorate degree (Dr. rer. nat.) in Auditory Science at the Medical School Hannover, Germany. Since 2017 she is a Postdoctoral Researcher at the German Hearing Center of the ENT clinics at the Medical School Hannover, Germany. Her research focuses on objective measures, especially EEG and pupillometry, in patients with auditory prostheses.

Maria Egger Maria Egger received her Bachelor degree in biomedical engineering in 2017 University of Applied Sciences Technikum, Vienna. She conducted her Bachelor thesis at the German Hearing Center (DHZ) focused on selective attention in CI users. Since 2018 she is working on emotion classification from physiological signals at the AIT Austrian Institute of Technology GmbH.

Bojana Mirkovic received her Bachelors degree in Electrical engineering in 2012 and Masters degree in Biomedical and Ecology Engineering in 2013 from University of Belgrade, Serbia. After working on development of mobile EEG solutions at mBrain-Train, Serbia (2013), she completed a PhD degree in Natural Science at the University of Oldenburg, Germany, in 2017. Her research interests are in the areas of biomedical signal processing, cognitive neuroscience, braincomputer interfaces and machine learning. Currently she is a Postdoctoral Researcher at the Neuropsychology Department at the University of Oldenburg investigating neural correlates of listening attention and listening demand.

Andreas Büchner received his M.S. degree in informatics in 1995 at the University of Hildesheim, Germany, and his Ph.D. degree (human biology) at the Medical University of Hannover in 2002. After developing signal processing and pattern recognition algorithms for medical imaging systems at the University of Hildesheim, he became a Research Scientist in the field of audiology at the Department of Otolaryngology at the Medical University of Hannover (MHH) in 1995. Since 2003, he has been the Scientific Director of the Hannover Hearing Center, subsidiary of the Department of Otolaryngology, MHH. At MHH, he quickly focused on the development and improvement of signal processing strategies for cochlear implants and their introduction into the clinical field. Other research interests are combination of acoustic and electric hearing (electroacoustic stimulation), tinnitus treatment, evaluation, and interpretation of electrically evoked action potentials of the auditory nerve.