Optimal number and placement of EEG electrodes for measurement of neural tracking of speech

Jair Montoya-Martínez

KU Leuven, Department of Neurosciences, Research Group Experimental Oto-Rhino-Laryngology

E-mail: jair.montoya@kuleuven.be

Alexander Bertrand

KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics

E-mail: alexander.bertrand@esat.kuleuven.be

Tom Francart

KU Leuven, Department of Neurosciences, Research Group Experimental Oto-Rhino-Laryngology

E-mail: tom.francart@med.kuleuven.be

Abstract. Measurement of neural tracking of natural running speech from the electroencephalogram (EEG) is an increasingly popular method in auditory neuroscience and has applications in audiology. The method involves decoding the envelope of the speech signal from the EEG signal, and calculating the correlation with the envelope that was presented to the subject. Typically EEG systems with 64 or more electrodes are used. However, in practical applications, set-ups with fewer electrodes are required. Here, we determine the optimal number of electrodes, and the best position to place a limited number of electrodes on the scalp. We propose a channel selection strategy, aiming to induce the selection of symmetric EEG channel groups in order to avoid hemispheric bias. The proposed method is based on a utility metric, which allows a quick quantitative assessment of the influence of each group of EEG channels on the reconstruction error. We consider two use cases: a subject-specific case, where the optimal number and positions of the electrodes is determined for each subject individually, and a subject-independent case, where the electrodes are placed at the same positions (in the 10-20 system) for all the subjects. We evaluated our approach using 64-channel EEG data from 90 subjects. Surprisingly, in the subject-specific case we found that the correlation between actual and reconstructed envelope first increased with decreasing number of electrodes, with an optimum at around 20 electrodes, yielding 38% higher correlations using the optimal number of electrodes. In the subject-independent case, we obtained a stable decoding performance when decreasing from 64 to 32 channels. When the number of channels was further decreased, the correlation decreased. For a maximal decrease in correlation of 10%, 32 well-placed electrodes were sufficient in 87% of the subjects. Practical electrode placement recommendations are given for 8, 16, 24 and 32 electrode systems.

1. Introduction

To understand how the human brain processes an auditory stimulus, it is essential to use ecologically valid stimuli. An increasingly popular method is to measure neural tracking of natural running speech from the electroencephalogram (EEG). This method also has applications in domains such as audiology, as part of an objective measure of speech intelligibility (Vanthornhout et al., 2018; Lesenfants et al., 2019), and coma science (Braiman et al., 2018).

The relationship between the stimulus and the brain response can be studied using two different models (e.g., Crosse et al., 2016; Lalor and Foxe, 2010; Ding and Simon, 2012; Verschueren et al., 2019; Vanthornhout et al., 2018): in the forward model (also know as encoding model), we determine a linear mapping from the stimulus to the brain response. On the other hand, in the backward model (also known as stimulus reconstruction), we determine the linear mapping from the brain response to the stimulus. Backward models are referred to as decoding models, because they attempt to reverse the data generation process. Both the forward and backward models involve the solution of a linear least squares (LS) regression problem. The quality of the reconstruction is usually quantified in terms of correlation between the true signal and the reconstructed one. The benefit of the forward model is that the obtained models (also called temporal response functions) can be easily interpreted, and topographical information can be easily obtained. The benefit of the backward model is that through combination of information across EEG channels, better performance (higher correlations) can be obtained, but the model coefficients can not be easily interpreted. In this experimental paradigm, the most used stimulus representation is its slowly varying temporal envelope (e.g., Ding and Simon, 2011; Aiken and Picton, 2008), which is known to be one of the most important cues for speech recognition (Shannon et al., 1995).

While in research one can easily use EEG systems with 64 electrodes or more, for practical applications, such as objective measurement of speech intelligibility in the clinic, there are stronger constraints due to the cost of systems with a large number of channels and the time required to place the electrodes on the scalp. We therefore considered the following questions: for a smaller number of electrodes, (1) what is the optimal location of electrodes on the scalp and (2) what is the impact on decoding accuracy when we decrease the number of channels. We can consider two use cases: in one case the optimal number and position of electrodes is determined for each subject individually. This is probably mostly applicable in research or very specialised applications. Another, more practical use case is where the electrodes are placed at the same positions (in the 10-20 system) for all subjects, which would for instance be relevant in the design of an application-specific headset or electrode cap. Given its advantages in decoding accuracy over the forward model, we focused on the backward model.

We started from 64-channel recordings, and considered the question which subset of K channels allow to get the best decoding performance. This is a combinatorial problem, closely related to the column subset selection problem (Boutsidis et al., 2009), whose NP-hardness is an interesting open problem. In order to overcome this challenge, (Mirkovic et al., 2015; Fuglsang et al., 2017) used a channel selection strategy based on an iterative backward elimination approach, where at each iteration, the electrode with the lowest corresponding coefficient magnitudes in the decoder is removed from the next iteration (we will refer to this channel selection method as the decoder magnitude-based (DMB) method). This strategy assumes that important channels will have a large coefficient in the least squares solution. However, as pointed out by (Bertrand, 2018), this is an unsuitable assumption: for example, if the coefficients of one of the channels would all be scaled with a factor α, then the corresponding decoder coefficient in the LS solution would be scaled with α^{-1}, whereas the information content of that channel obviously remains unchanged.

In this work, we propose a channel selection strategy, aiming to induce the selection of *symmetric* EEG channel groups (see Figure 1), where, for channels located off the midline each group is composed of one channel located over the left hemisphere and its closest symmetric counterpart located over the right hemisphere. For channels located over the central line dividing both hemispheres, each group is composed of one channel located over the frontal lobe and its closest symmetric counterpart located either over the parietal or the occipital lobe. The rationale behind this channel selection strategy is to maintain symmetry. The symmetry criterion avoids bias to one hemisphere, which could be problematic as hemispheric differences are often found between subjects (e.g., Goossens et al., 2019; Van Eeckhoutte et al., 2018; Poelmans et al., 2012; Vanvooren et al., 2015). The proposed method is based on the utility metric (Bertrand, 2018), which allows a quick quantitative assessment of the influence of each group of channels on the reconstruction error. A similar channel selection strategy, also based on the utility metric, was proposed by (Narayanan and Bertrand, 2019) on an auditory attention decoding task, where the main goal was to optimize the topology of a wireless EEG sensor network (WESN), without imposing a symmetry constraint on the selected channels. We evaluated our approach using EEG data from 90 subjects. We aimed to minimize reconstruction error, and to minimize the intra-subject variability in reconstruction error.

2. Methods

2.1. Data collection

2.1.1. Participants Ninety Flemish-speaking volunteers participated in this study. They were recruited from our university student population to ensure normal language processing and cognitive function. Each participant reported normal hearing, which was verified by pure tone audiometry (thresholds lower than 25 dB HL for 125 Hz until 8000 Hz using MADSEN Orbiter 922–2 audiometer). Before each experiment, the participants signed an informed consent form approved by the Medical Ethics Committee UZ KU Leuven/Research (KU Leuven).

2.1.2. Experiment Each participant listened attentively to the children's story "Milan", written and narrated in Flemish by Stijn Vranken. The stimulus was 15 minutes long and was presented binaurally at 60 dBA without any noise. It was presented through Etymotic ER-3A insert phones (Etymotic Research, Inc., IL, USA) which were electromagnetically shielded using CFL2 boxes from Perancea Ltd. (London, UK). The acoustic system was calibrated using a 2-cm^3 coupler of the artificial ear (Brüel & Kjær, type 4192). The experimenter sat outside the room and presented the stimulus using the APEX 3 (version 3.1) software platform developed at ExpORL (Dept. Neurosciences, KU Leuven, Belgium) (Francart et al., 2008) and a RME Multiface II sound card (RME, Haimhausen, Germany) connected to a laptop. The experiments took place in a soundproof, electromagnetically shielded room.

2.1.3. EEG acquisition In order to measure the EEG responses, we used a BioSemi (Amsterdam, Netherlands) ActiveTwo EEG setup with 64 channels. The signals were recorded at a sampling rate of 8192 Hz, using the ActiView software provided by BioSemi. The electrodes were placed over the scalp according to the international 10-20 standard.

2.2. Signal processing

2.2.1. EEG pre-processing In order to decrease computation time, the EEG data was downsampled from 8192 Hz to 1024 Hz. Then, the EEG artifacts were removed by using the Sparse Time Artifact Removal method (STAR) (de Cheveigné, 2016), as well as a multi-channel Wiener filter algorithm (Somers et al., 2018). Next, the data was bandpass filtered between 0.5-4 Hz (delta band), using a Chebyshev filter with 80 dB attenuation at 10 % outside the passband. Finally, the data was downsampled to 64 Hz and re-referenced to Cz in the channel subset selection stage, and to a common-average reference (across the selected channels) in the decoding performance evaluation stage. The delta band was chosen because it yields the highest correlations and most information in the stimulus envelope is in this frequency band (Vanthornhout et al., 2018; Ding and Simon, 2014). However, this choice is application-dependent and it is straightforward to repeat our analysis with different filter settings.

2.2.2. Speech envelope The speech envelope was computed according to (Biesmans et al., 2017), who showed that good reconstruction accuracy can be achieved with a gammatone filterbank followed by a power law. We used a gammatone filterbank (Søndergaard et al., 2012; Søndergaard and Majdak, 2013), with 28 channels spaced by 1 equivalent rectangular bandwidth, with centre frequencies from 50 Hz to 5000 Hz. From each subband, we take the absolute value of each sample and raise it to the power of 0.6. The resulting 28 signals were then downsampled to 1024 Hz, averaged, bandpass filtered with a (0.5-4 Hz) Chebyshev filter to obtain the final envelope, and finally downsampled again to 64Hz. The power law was chosen as the human auditory system is not a linear system and compression is present in the system. The gammatone filterbank was chosen as it mimics the auditory filters present in the basilar membrane in the cochlea.

2.2.3. Backward model The backward model to decode a speech envelope from the EEG can be stated as a regularized linear least squares (LS) problem (O'sullivan et al., 2014):

$$J(\mathbf{X}) \triangleq \underset{\mathbf{w}}{\text{minimize}} \ \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2 \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{T\times(N\times\tau)}$ is the EEG data matrix concatenated with τ time-shifted (zero-padded) version of itself, $\mathbf{y} \in \mathbb{R}^{T\times 1}$ is the speech envelope, $\mathbf{w} \in \mathbb{R}^{(N\times\tau)\times 1}$ is the decoder, T is the total number of time samples, N is the number of channels, τ is the number of time samples covering the time integration window of interest, and λ is a regularization parameter. The solution to the backward problem ($\hat{\mathbf{w}}$) is usually referred to as a decoder. In order to choose the regularization parameter λ, we compute and sort the eigenvalues of the covariance matrix associated to $\mathbf{X}$. Then, we pick as λ the eigenvalue where the accumulated percentage of explained variance is greater than 99%.

2.2.4. Channel selection To select channels we used the utility metric (Bertrand, 2018), which quantifies the effective loss, i.e., the increase in the LS cost, if a group of columns (corresponding to one channel or a set of channels and all their $\tau - 1$ corresponding time-shifted version) would be removed and if the model (1) would be reoptimized afterwards:

$$U_g \triangleq J(\mathbf{X}_{-g}) - J(\mathbf{X}) \tag{2}$$

where $\mathbf{X}_{-g}$ denotes the EEG data matrix $\mathbf{X}$ after removing the columns associated with the g-th group of channels and their corresponding time-shifted versions. We will later on define how channels are grouped in our experiments (see Subsection 2.2.5).

Note that a naive implementation of computing U_g would require solving one LS squares problem like (1), for each possible removal of a candidate group, which could potentially lead to a large computational cost for problems with large dimensions and/or involving a large number of groups. Fortunately, this can be circumvented, as shown by (Bertrand, 2018), with a final computational complexity that scales linearly in the number of groups, given the solution of (1) when none of the channels are removed. The basic workflow for finding the best k groups of EEG channels can be summarized as follows (Narayanan and Bertrand, 2019): we compute the utility metric for each of the groups and remove the group with the lowest utility. Next, we recalculate the new values of the utility metric taking only into account the remaining groups, and once again we remove the one with the lowest value of utility. We continue iterating following these steps until we arrive to k groups.

We used the utility metric in two conditions: (1) in the subject-specific case where optimal electrodes are selected for each subject, and (2) in the generic case where the same set of electrodes is used for all subjects.

In the subject-specific case, we computed (for each subject i) the regularized covariance matrix $C^{(i)} = \frac{\mathbf{X}^{(i)\top}\mathbf{X}^{(i)}}{T} + \lambda\mathbf{I}$ ($\mathbf{I}$ denotes the identity matrix) and the cross-correlation vector $\mathbf{r}^{(i)} = \frac{\mathbf{X}^{(i)\top}\mathbf{y}}{T}$ in order to compute the optimal all-channel decoder $\hat{\mathbf{w}}^{(i)} = \left(C^{(i)}\right)^{-1}\mathbf{r}^{(i)}$. The utility metric for each (group of) channel(s) can be directly computed‡ from $\hat{\mathbf{w}}^{(i)}$ and $C^{(i)}$ (we refer to (Bertrand (2018)) and (Narayanan and Bertrand (2019)) for more details). We then ranked the groups according to their corresponding utilities, and removed the channel(s) corresponding to the group g with the lowest utility. We then repeated the same process with the matrix $X_{-g}^{(i)}$ in which the columns corresponding to the channels in group g were removed. We kept repeating this process until only k groups remained.

Next, during the decoding evaluation stage, we computed a decoder by solving the backward problem using the best k selected groups of channels for each subject. In this stage, we re-referenced the channels with respect to the common average across the selected channels and discarded the reference electrode Cz. We solved each backward problem using a 7-fold cross-validation approach, where 6 folds were used for training and 1 for testing. This corresponds to approximately 12 and 2 minutes of data, respectively. Using the decoder $\hat{\mathbf{w}}$, we computed the reconstructed envelope as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{w}}$ after which we computed the Spearman correlation between the reconstructed speech envelope ($\hat{\mathbf{y}}$) and the true one ($\mathbf{y}$). By following this procedure, for each subject, we ended up with 7

‡ We used the utility metric toolbox from (Narayanan and Bertrand (2019)) available at `https://github.com/mabhijithn/channelselect`

values of correlation (corresponding to the evaluation of the correlation using each one of the test folds), which can be arranged as an array $\mathbf{S} \in \mathbb{R}^{90\times k\times 7}$ (number of subjects $\times$ number of groups $\times$ number of test folds).

To compare with the literature, we also implemented the DMB approach, wherein we iteratively solved a backward problem for each subject, and at each iteration, the group of electrodes with the lowest corresponding coefficient magnitudes in the decoder was removed from the next iteration.

In the generic case, where the same set of electrodes is used for all subjects, we only used the utility metric. The evaluation consisted of the same two stages described above. The only difference was that, during the channel selection stage, we computed a grand average model by averaging the covariance matrices of all the subjects, which is equivalent to concatenating all the data from all the subjects in the data matrix X in (1). Finally, the decoding evaluation stage followed exactly the same steps described for the subject-specific case above, i.e., using a subject-*specific* decoder (yet, computed over electrodes that were selected in a subject-*independent* fashion).

2.2.5. Symmetric grouping of the EEG channels In addition to selecting individual channels to remove (no grouping of channels), we also evaluated a strategy in which symmetric groups of channels were removed, to avoid hemisphere bias effects across subjects. Each group is composed of two EEG channels (see Figure 1). For channels located on either side of the midline (Figure 1, groups with labels from 1 to 27), each group is composed by one channel located over the left hemisphere and its closest symmetric counterpart located over the right hemisphere. For channels located over the midline dividing both hemispheres (Figure 1, groups with labels from 28 to 31), each group is composed by one channels located over the frontal lobe and its closest symmetric counterpart located either over the parietal or the occipital lobe. Channel Cz does not belong to any group because it was used as a reference (in the channel subset selection stage). Channel Iz was not considered in order to preserve the symmetry with respect to the number of electrodes.

3. Results

3.1. Channel selection strategies: utility metric vs DMB

We compared the performance of the utility metric and DMB in the the subject-specific case, where the optimal electrode locations were determined for each subject individually. We compared the median of the correlation between $\mathbf{y}$ and $\hat{\mathbf{y}}$ for each subject, as well as the number of channels required to obtain it (from now on referred to as the optimal number of channels). Surprisingly, for both methods we observe a large increase in correlation when we use a reduced number of channels, with the optimum of the median around 20 and 36 channels, for the utility metric and DMB, respectively (see Figure 2a). This means that the evaluated strategies of removing electrodes can be used to improve

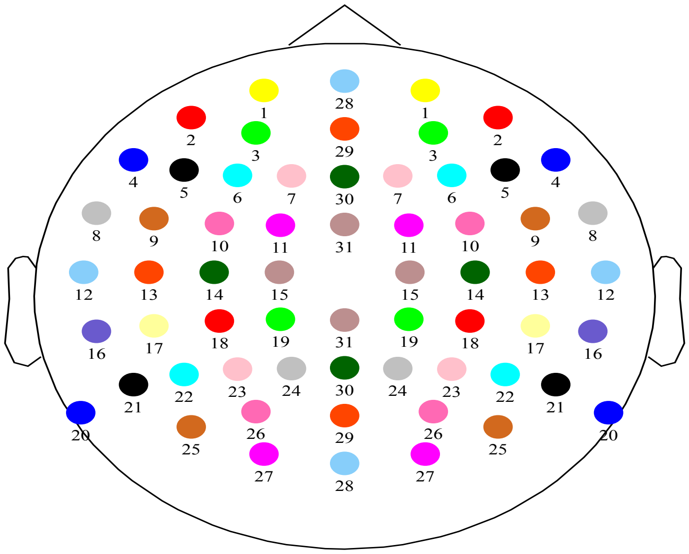


Figure 1: **Channel grouping strategy**. For channels located either over the left or right hemisphere (groups $1, 2, \ldots, 27$), each group is composed by one channel located over the left hemisphere and its closest symmetric counterpart located over the right hemisphere. For channels located over the central line dividing both hemispheres (groups $28, 29, 30, 31$), each group is composed by one channels located over the frontal lobe and its closest symmetric counterpart located either over the parietal or the occipital lobe.

the correlation metric in high-density EEG recordings.

We can see in Figure 2a that the utility metric globally outperforms the DMB approach, obtaining consistently higher correlations (median) across subjects. In Figure 2b, we can see that the utility metric also outperforms the DMB approach on an individual level, obtaining for every subject a higher value of maximal correlation, as well as requiring a smaller number of electrodes to obtain it. A Wilcoxon signed rank test showed that there was a significant difference (W=18, $p < 0.001$) between the correlation using the optimal number of channels according to the utility metric (median=0.22) compared to the one obtained using DMB (median=0.19). Another Wilcoxon signed rank test showed that there was also a significant difference (W=780.5, $p < 0.001$) between the optimal number of channels selected by the utility metric (median=10) compared to the optimal number selected by DMB (median=15). Because of the improved performance offered by the utility metric compared to DMB, we solely focus on the former in the remaining of the paper.

3.2. Channel selection based on the utility metric vs using all the channels

In this section, we compare the channel selection strategy based on the utility metric with the case where all the available channels are used. We compared both strategies in the subject-specific scenario, as well as the subject-independent (generic) one.

3.2.1. Subject-specific electrode locations Figure 3a shows the median correlation, computed as the median across folds followed by the median across subjects. Blue dashed lines show the 25-th (lower) and 75-th (upper) percentile. In this figure, we can see that at least 50% (median) of the subjects exhibit a higher value of correlation for 6 up to 64 channels.

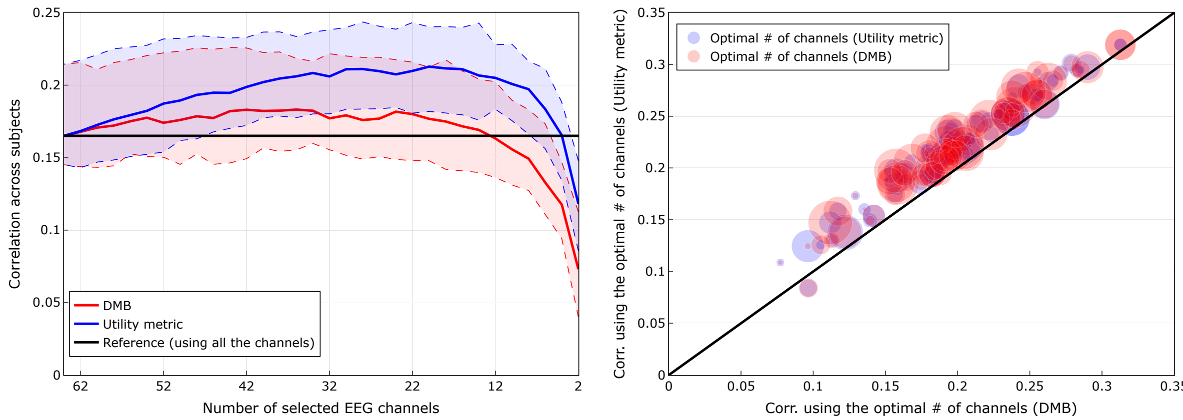
Figure 3b shows the standard deviation of the correlation, as a measure for within-subject variability, computed as the standard deviation across folds followed by the median across subjects. Blue dashed lines show the 25-th (lower) and 75-th (upper) percentile. In this figure we can see a largely stable standard deviation of the correlation around the reference value (standard deviation of the correlation when using all the 64 channels).

Figures 3a and 3b suggest that we could obtain a slightly higher correlation with a reduced number of channels. However, these are group results. Figure 3c shows, independently for each subject, the difference between the correlation when we use all the 64 channels and when we use a reduced number of channels. We can see that this effect is indeed consistently present for all subjects when we use a number of channels between 34 and 46. This behaviour can be seen more clearly in Figure 5a, where the percentage of subjects with a correlation greater or equal to 100%, 95% and 90% of the correlation obtained using all the channels (green, purple and cyan lines, respectively) is shown. Figure 5a clearly shows that for 98% of the subjects it is possible to reduce the number of channels to 32 and still be able to obtain a correlation higher than the one obtained using all the channels. Even if we go all the way down to 10 channels, we can see that 94%, 98% and 99% of the subjects is still able to get a correlation higher than 100%, 95% and 90% of the correlation obtained using all channels, respectively.

Figure 3d shows a comparison of the correlation obtained using the optimal number of channels (obtained through the utility metric) versus the correlation obtained using all 64 channels. In this figure we can see that for every subject the utility metric consistently yielded a higher value of correlation compared to using all the channels. A Wilcoxon signed rank test showed that there was a significant difference (W=0, $p < 0.001$) between the correlation using the optimal number of channels according to the utility metric (median=0.22) compared to the one obtained using all the channels (median=0.16), which is a 38% improvement.

So far we presented the results for the condition where we removed channels one by one. We also evaluated the symmetric grouping approach in the subject-specific case, but obtained worse results: median correlations with the optimal number of channels significantly decreased from 0.22 to 0.21 when moving from the channel-by-channel to the symmetric grouping strategy ($W = 223, p < 0.001$).

3.2.2. Subject-independent electrode locations We now consider the case where the same set of electrodes is used for all subjects. Figure 4a shows the correlation across subjects, computed as the median across folds followed by the median across subjects. In this figure, we can see that at least 50% (median) of the subjects exhibit a slightly higher



(a) Correlation across subjects, computed as the median across folds followed by the median across subjects. Dashed lines show the 25-th (lower) and 75-th (upper) percentile.

(b) Comparison of the correlation obtained using the optimal number of channels (number of channels where each subject obtained the highest correlation). Size of the markers is proportional to the optimal number of channels (one marker per subject).

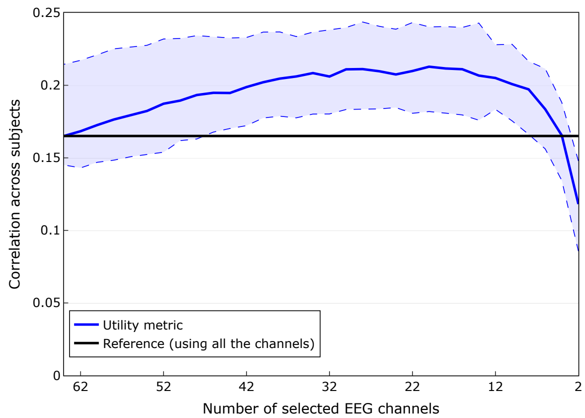
Figure 2: **Comparison of channel selection strategies: utility metric vs DMB (*subject-specific scenario*)**. A Wilcoxon signed rank test showed that there was a significant difference (W=18, $p < 0.001$) between the correlation obtained using the optimal number of channels according to the utility metric (median=0.22) compared to the one obtained using DMB (median=0.19). Another Wilcoxon signed rank test showed that there was also a significant difference (W=780.5, $p < 0.001$) between the optimal number of channels selected by the utility metric (median=10) compared to the one selected by DMB (median=15).

correlation for 20 up to 64 channels.

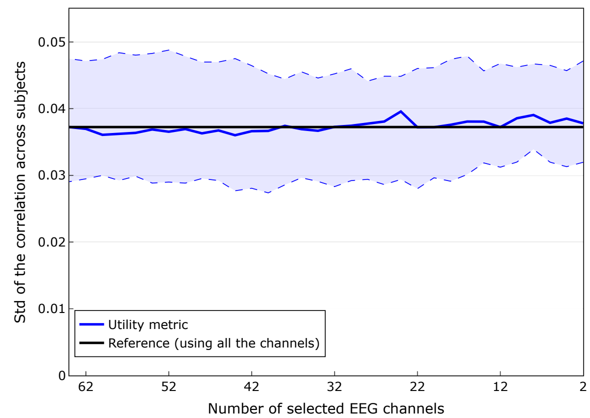
Contrary to the subject-specific electrode locations, we here found a benefit of using the symmetric channel grouping strategy: median correlations with the optimal number of channels significantly improved from 0.177 to 0.188 when moving from the channel-by-channel to symmetric grouping strategy ($W = 1000, p < 0.01$). In the figures and what follows, we only consider the results obtained with the symmetric grouping strategy.

Figure 4b shows the standard deviation of the correlation, as a measure of within-subject variability, computed as the standard deviation across folds followed by the median across subjects. In this figure we can see a largely stable standard deviation of the correlation around the reference value (standard deviation of the correlation when using all the 64 channels).

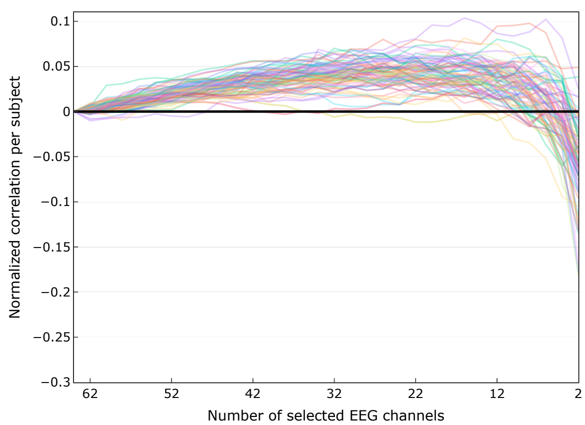
Figures 4a and 4b suggest that, similarly to the case with individual electrode locations, we could obtain a slightly higher value of correlation with a reduced number of channels. However, these are group results. Figure 4c shows, independently for each subject, the difference between the value of the correlation when we use all the 64



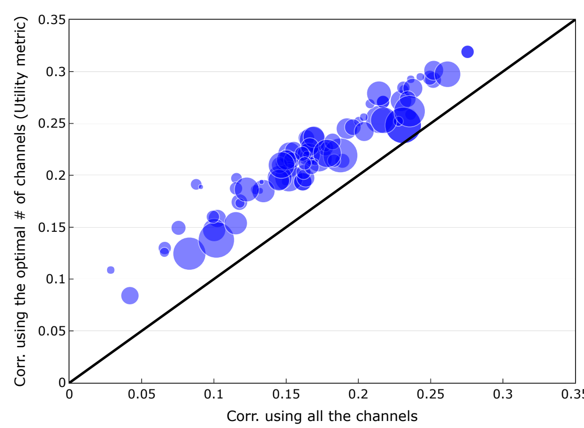
(a) Correlation computed as the median across folds followed by the median across subjects. Dashed lines show the 25-th (lower) and 75-th (upper) percentile.



(b) Standard deviation of the correlation coefficient, computed as the standard deviation across folds followed by the median across subjects. Dashed lines show the 25-th (lower) and 75-th (upper) percentile.

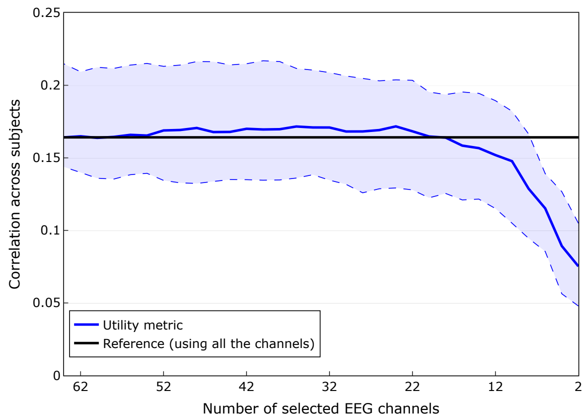


(c) Normalized correlation per subject (each line is a different subject), defined as the difference between the value of the correlation obtained when we use all the channels and the value of the correlation obtained when we use a reduced number of channels.

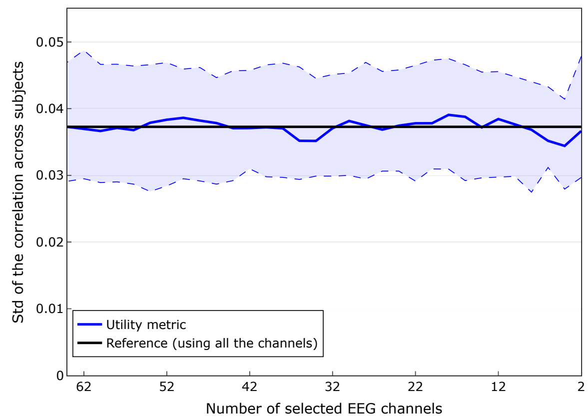


(d) Comparison of the correlation obtained using the optimal number of channels (number of channels where each subject obtained the highest correlation) vs the correlation obtained using all the channels. Size of the markers is proportional to the optimal number of channels (one marker per subject).

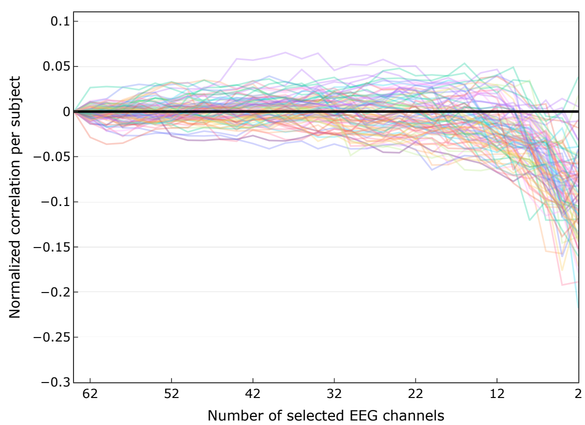
Figure 3: **Comparison of the channel selection based on the utility metric vs using all the channels (*subject-specific scenario*)**. A Wilcoxon signed rank test showed that there was a significant difference (W=0, $p < 0.001$) between the correlation obtained using the optimal number of channels suggested by the utility metric (median=0.22) compared to the one obtained using all the channels (median=0.16). Only results for the individual (non-grouped) channel-by-channel selection strategy are shown as these provided the best results for the subject-specific scenario.



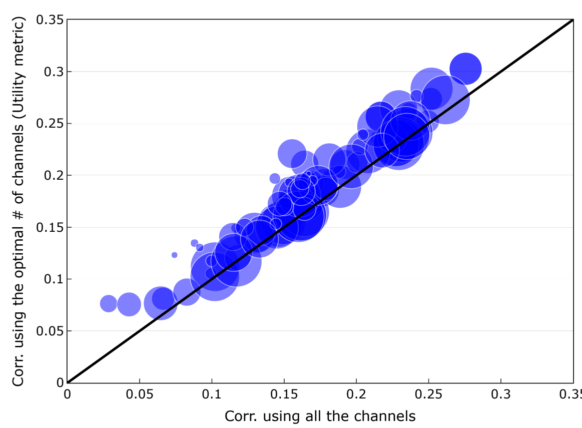
(a) Correlation across subjects, computed as the median across folds followed by the median across subjects. Dashed lines show the 25-th (lower) and 75-th (upper) percentile.



(b) Standard deviation of the correlation coefficient, computed as the standard deviation across folds followed by the median across subjects. Dashed lines show the 25-th (lower) and 75-th (upper) percentile.



(c) Normalized correlation per subject (each line is a different subject), defined as the difference between the value of the correlation obtained when we use all the channels and the value of the correlation obtained when we use a reduced number of channels.



(d) Comparison of the correlation obtained using the optimal number of channels (number of channels where each subject obtained the highest correlation) vs the correlation obtained using all the channels. Size of the markers is proportional to the optimal number of channels (one marker per subject).

Figure 4: **Comparison of the channel selection based on the utility metric vs using all the channels (*subject-independent scenario*)**. A Wilcoxon signed rank test showed that there was a significant difference (W=0, $p < 0.001$) between the correlation obtained using the optimal number of channels suggested by the utility metric (median=0.19) compared to the one obtained using all the channels (median=0.16). Only results for the symmetric channel grouping strategy are shown, as these provided the best results for the subject-independent scenario.

channels and the value of the correlation when we use a reduced number of channels. We can see that this effect is not consistently present for all subjects (if that would have been the case, all the lines would have appeared above 0 when we use a reduced number of channels n_k, $20 \leq n_k < 64$). Nevertheless, a certain percentage of subjects do exhibit a higher value of the correlation when using a reduced number of channels. Figure 5b helps us to quantify this behaviour, by showing the percentage of subjects with a correlation greater or equal to 100%, 95% and 90% of the correlation obtained using all the channels (green, purple and cyan lines, respectively). In this figure we can see that for 52%, 70% and 87% of the subjects it is possible to reduce the number of channels to 32 and still be able to obtain a correlation higher than 100%, 95% and 90% of the correlation obtained using all channels, respectively. The percentage of subjects can increase to 56%, 78% and 91%, respectively, if we increase the number of channels from 32 to 36.

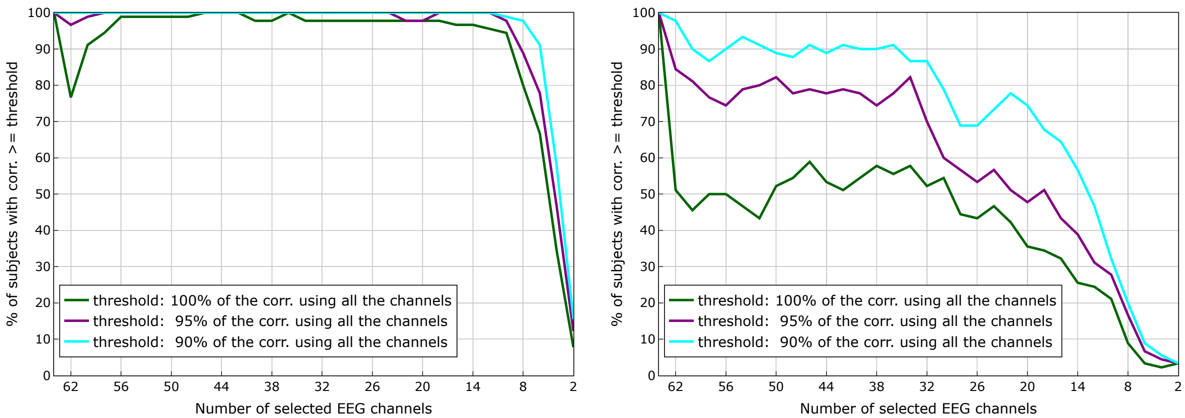
Figure 4d shows a comparison of the correlation obtained using the optimal number of channels suggested by the utility metric versus the correlation obtained using all 64 channels. In this figure we can see that, similar to the subject-specific scenario, the utility metric consistently obtained, for every subject, a higher value of correlation compared to correlation obtained when using all the channels. A Wilcoxon signed rank test showed that there was a significant difference (W=0, $p < 0.001$) between the correlation obtained using the optimal number of channels suggested by the utility metric (median=0.19) compared to the one obtained using all the channels (median=0.16).

Figures 6a, 6b, 6c and 6d show the best 8, 16, 24 and 32 channels selected by the utility metric. Next to each group of channels (formed exactly by two electrodes, see Figure 1), a number is shown which is computed as $N - p + 1$ where N is the total number of groups and p is the iteration at which the group was discarded in the greedy removal procedure. The lower this number, the more important the group, as it was retained for a longer number of iterations in the backwards greedy removal process due to its high influence in the LS cost (see Section 2.2.4). As we can see, the selected channels are mostly clustered over the left and right temporal lobes, which agrees with the empirical evidence which suggests that channels located close to auditory cortex are important for picking up electrical brain activity evoked as response to an auditory stimulus.

4. Discussion

Based on 64-channel EEG recordings, we determined the effect of reducing the number of available channels and the optimal electrode locations on the scalp for 4 frequently-used numbers of channels. This was based on a novel utility-based metric, by which we avoided the computationally intractable number of combinations that underlies the problem at hand.

(Mirkovic et al., 2015; Fuglsang et al., 2017) tackled the channels subset selection problem in the context auditory attention decoding (identify the attended speech stream



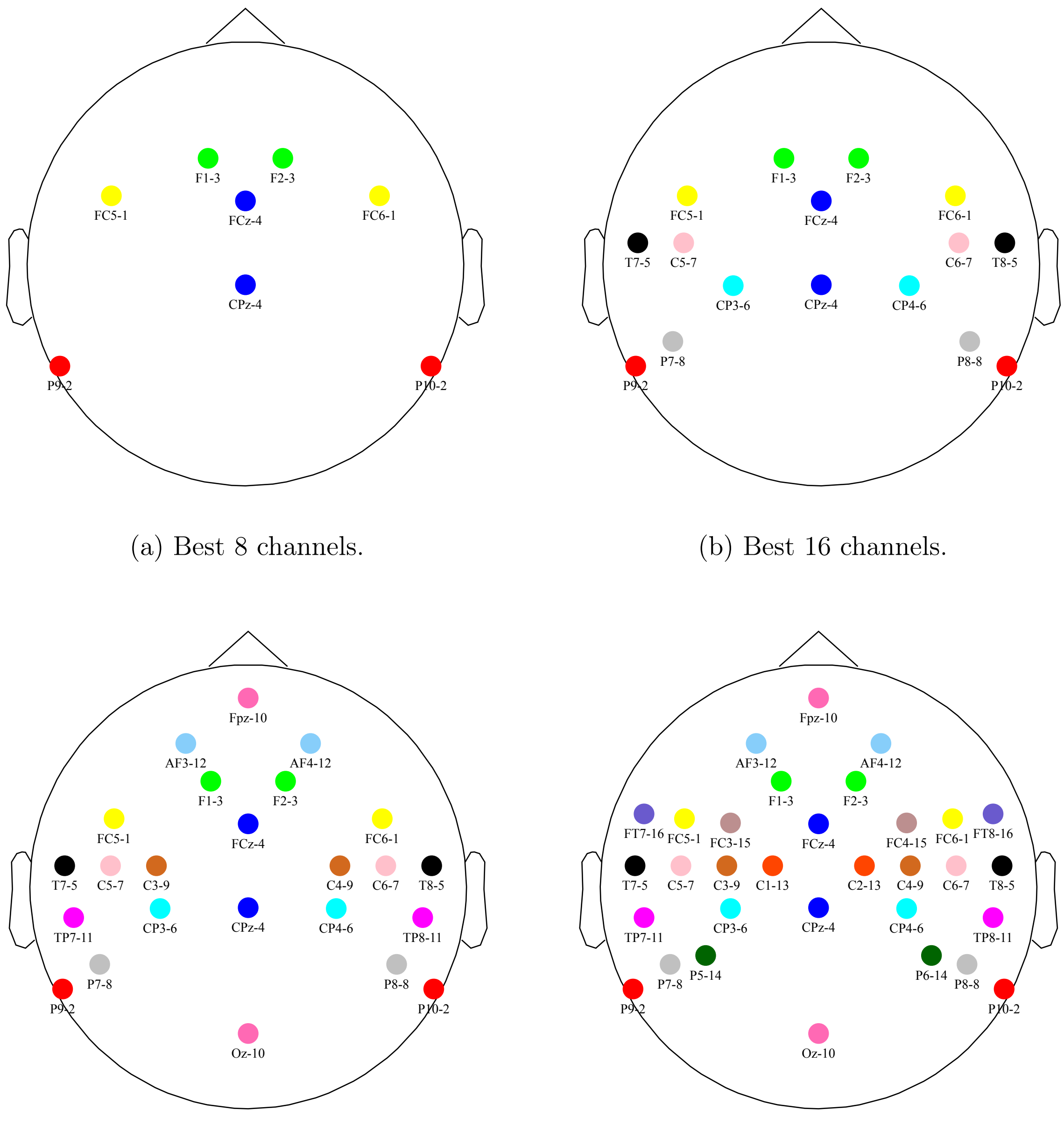
(a) Subject-specific scenario. (b) Subject-independent scenario.

Figure 5: **Percentage of subjects with a correlation greater or equal to 100%, 95% and 90% of the correlation obtained using all the channels**. In the subject-specific scenario we can see that for 98% of the subjects is possible to reduce the number of channels to 32 and still be able to obtain a correlation higher than the one obtained using all the channels. In the subject-independent scenario we can see that for 52%, 70% and 87% of the subjects is possible to reduce the number of channels to 32 and still be able to obtain a correlation higher than 100%, 95% and 90% of the correlation obtained using all channels, respectively. The percentage of subjects can increase to 56%, 78% and 91%, respectively, if we increase the number of channels from 32 to 36.

in a multi-speaker scenario). (Mirkovic et al., 2015; Fuglsang et al., 2017) processed the EEG recordings from 12 and 29 subjects, acquired using an EEG system with 96 and 64 channels, respectively. They found that, on average, the decoding accuracy dropped when using a number of channels less than 25. Both studies used the same channel selection strategy, which is based on an iterative backward elimination approach, where at each iteration, the channel with the lowest average decoder coefficient is removed from the next iteration. This strategy assumes that important channels will have a large coefficient in the LS solution. However, as explained in the introduction, this is not necessarily a suitable assumption. They did not report optimal electrode positions.

(Narayanan and Bertrand, 2019) also analyzed the channel subset selection problem in the context of auditory attention decoding, using a channel selection strategy based on the same utility metric discussed in the present study, but without imposing the symmetric grouping approach discussed in Section 2.2.5. They found that, on average, the decoding accuracy remained stable when using a number of channels greater or equal to 10. The (asymmetric) channels reported in their study correspond with the ones reported in this study in the sense that mostly channels around the left and right temporal lobes were selected.

Instead of attention decoding accuracy, we assessed the correlation between actual and reconstructed envelope (in a single-speaker scenario), which can be used as a metric for speech intelligibility (Vanthornhout et al., 2018; Lesenfants et al., 2019). For subject-



(a) Best 8 channels.

(b) Best 16 channels.

(c) Best 24 channels.

(d) Best 32 channels.

Figure 6: **Practical electrode placement recommendations.** The number next to each group of channels (formed by two electrodes, see Figure 1) indicates the ranking of the group with respect to its influence on the LS cost (see text). The lower this number, the more important the group.

specific electrode locations, we found similar differences between the DMB and utility metric: using the DMB metric, on average 14 electrodes were required to avoid a drop in correlation below the 64-channel case, and using the utility metric, only 6 electrodes were required. On top of this, we found a substantial increase in correlation when reducing the number of electrodes from 64 to 32-20. This indicates that application of the proposed channel selection approach may be practically useful.

The stable or sometimes even improved performance after reducing the number of channels could be attributed to the removal of noisy or irrelevant channels that do not contribute significantly to the reconstruction of the target speech envelope. As explained in Section 2.2.3, the backward problem is usually solved by using a regularized Ridge regression approach, which shrinks the magnitude of many decoder components to prevent overfitting (finding solutions that minimize the reconstruction error while satisfying, at the same time, the condition of having a small norm value). We recalculated the optimal regularization parameter for each number of channels. Reducing the number of channels has a similar regularization effect; it reduces the degrees of freedom by discarding irrelevant channels, making the model less prone to overfitting.

In the case where the same channels were selected for all subjects, the initial increase in correlation with decreasing number of channels was smaller and not present for all subjects. Therefore in this case our strategy is mainly useful to come up with a practical number and location of electrodes.

4.1. Selected channels

Based on the literature, we expect that most of the signals of interest originate from auditory cortex (e.g., Brodbeck et al., 2018; Pasley et al., 2012). We indeed see that channels that cover dipoles originating in this area are always selected with high priority. For higher numbers of channels, other areas are covered where auditory related responses have been shown to originate from, such as the inferior frontal cortex and the premotor cortex (Das et al., 2018; Lesenfants et al., 2019), and possibly channels that aid in the suppression of large irrelevant sources.

Note that channels that are typically prone to large artifacts, such as those close to the eyes (ocular artifacts) and in areas where the electrode-skin contact tends to be worse (lower portion of the occipital lobe) do not tend to be selected.

4.2. Applications

The backward model has been proposed in applications where an objective measure of speech intelligibility is needed. Our suggested electrode positions could be used to configure an electrode cap or headset for this specific application. We chose to run our calculations with the speech envelope as the stimulus feature and for the delta band (0.5-4Hz), as these parameters are most commonly used. Note that when deviating from these parameters, the selection should be re-run. In particular, when higher-order stimulus features are used, we expect significant changes in topography and therefore optimal electrode positions.

In cases where one has the opportunity to make an individual selection of electrode positions after the recording, our algorithm can be straightforwardly applied, and can lead to large increases in correlation.

5. Conclusion

In this work, the effect of selecting a reduced number of EEG channels was investigated within the context of the stimulus reconstruction task. We proposed a utility-based greedy channel selection strategy, aiming to induce the selection of symmetric EEG channel groups, while maximizing the covered area over the scalp. We evaluated our approach using 64-channel EEG data from 90 subjects. When using individual electrode selections for each subject, we found that the correlation between the actual and reconstructed envelope first increased with decreasing number of electrodes, with an optimum at around 20 electrodes. This means that the proposed method can be used in practice to obtain higher correlations. When using a generic electrode placement that is the same for all subjects, we obtained a stable decoding performance when using all 64 channels down to 32, suggesting that it is possible to get an optimal reconstruction of the speech envelope from a reduced number of EEG channels. Practical electrode placement recommendations are given for 8, 16, 24 and 32 electrode systems.

6. Acknowledgments

The authors would like to thank to Abhijith Mundanad Narayanan for sharing his code to compute the utility metric, as well as for the insightful discussions about the mathematical properties of the utility metric. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 637424, ERC starting Grant to Tom Francart).

References

Aiken, S. J. and Picton, T. W. (2008). Human cortical responses to the speech envelope, *Ear and hearing* **29**(2): 139–157.

Bertrand, A. (2018). Utility Metrics for Assessment and Subset Selection of Input Variables for Linear Estimation [Tips & Tricks], *IEEE Signal Processing Magazine* **35**(6): 93–99.

Biesmans, W., Das, N., Francart, T. and Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**(5): 402–412.

Boutsidis, C., Mahoney, M. W. and Drineas, P. (2009). An improved approximation algorithm for the column subset selection problem, *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, pp. 968–977.

Braiman, C., Fridman, E. A., Conte, M. M., Voss, H. U., Reichenbach, C. S., Reichenbach, T. and Schiff, N. D. (2018). Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury, *Current Biology* **28**(23): 3833–3839.

Brodbeck, C., Presacco, A. and Simon, J. Z. (2018). Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension, *NeuroImage* **172**: 162–174.

Crosse, M. J., Di Liberto, G. M., Bednar, A. and 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.

Das, P., Brodbeck, C., Simon, J. Z. and Babadi, B. (2018). Cortical localization of the auditory temporal response function from meg via non-convex optimization, *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, IEEE, pp. 373–378.

de Cheveigné, A. (2016). Sparse time artifact removal, *Journal of neuroscience methods* **262**: 14–20.

Ding, N. and Simon, J. Z. (2011). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening, *Journal of neurophysiology* **107**(1): 78–89.

Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers, *Proceedings of the National Academy of Sciences* **109**(29): 11854–11859.

Ding, N. and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations, *Frontiers in human neuroscience* **8**: 311.

Francart, T., Van Wieringen, A. and Wouters, J. (2008). APEX 3: a multi-purpose test platform for auditory psychophysical experiments, *Journal of neuroscience methods* **172**(2): 283–293.

Fuglsang, S. A., Dau, T. and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes, *Neuroimage* **156**: 435–444.

Goossens, T., Vercammen, C., Wouters, J. and van Wieringen, A. (2019). The association between hearing impairment and neural envelope encoding at different ages, *Neurobiology of Aging* **74**: 202–212.

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**(1): 189–193.

Lesenfants, D., Vanthornhout, J., Verschueren, E. and Francart, T. (2019). Data-driven spatial filtering for improved measurement of cortical tracking of multiple representations of speech, *Journal of Neural Engineering* .

Mirkovic, B., Debener, S., Jaeger, M. and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications, *Journal of neural engineering* **12**(4): 046007.

Narayanan, A. M. and Bertrand, A. (2019). Analysis of miniaturization effects and channel selection strategies for EEG sensor networks with application to auditory attention detection, *IEEE Transactions on Biomedical Engineering* .

O'sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A. and Lalor, E. C. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG, *Cerebral Cortex* **25**(7): 1697–1706.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T. and Chang, E. F. (2012). Reconstructing speech from human auditory cortex, *PLoS biology* **10**(1): 1–13.

Poelmans, H., Luts, H., Vandermosten, M., Ghesquière, P. and Wouters, J. (2012). Hemispheric asymmetry of auditory steady-state responses to monaural and diotic stimulation, *Journal of the Association for Research in Otolaryngology* **13**(6): 867–876.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M. (1995). Speech recognition with primarily temporal cues, *Science* **270**(5234): 303–304.

Somers, B., Francart, T. and Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel Wiener filter, *Journal of neural engineering* **15**(3): 036007.

Søndergaard, P. L., Torrésani, B. and Balazs, P. (2012). The linear time frequency analysis toolbox, *International Journal of Wavelets, Multiresolution and Information Processing* **10**(04): 1250032.

Søndergaard, P. and Majdak, P. (2013). The auditory modeling toolbox, *The technology of binaural listening*, Springer, pp. 33–56.

Van Eeckhoutte, M., Wouters, J. and Francart, T. (2018). Objective binaural loudness balancing based on 40-hz auditory steady-state responses. part i: Normal hearing, *Trends in Hearing* **22**.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z. and Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope, *Journal of the Association for Research in Otolaryngology* pp. 1–11.

Vanvooren, S., Hofmann, M., Poelmans, H., Ghesquière, P. and Wouters, J. (2015). Theta, beta and gamma rate modulations in the developing auditory system, *Hearing research* **327**: 153–162.

Verschueren, E., Somers, B. and Francart, T. (2019). Neural envelope tracking as a measure of speech understanding in cochlear implant users, *Hearing research* **373**: 23–31.