KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Mundanad Narayanan A., Zink R., Bertrand A. "**EEG miniaturization limits for stimulus decoding with EEG sensor networks**" . J Neural Eng. 2021 Oct 4;18(5). doi: 10.1088/1741-2552/ac2629. PMID: 34517358.
Archived version	Author manuscript: published version (without the Journal's typesetting)
Published version	https://doi.org/10.1088/1741-2552/ac2629
Journal homepage	https://publishingsupport.iopscience.iop.org/journals/journal-of-neural-engineering/
Author contact	Email mailme@abhimundanad.com
IR	

(article begins on next page)

EEG miniaturization limits for stimulus decoding with EEG sensor networks

Abhijith Mundanad Narayanan[1,2], Rob Zink[1], and Alexander Bertrand*[1,2]

[1]*KU Leuven, Dept. of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics (STADIUS), Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[2]*Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium*

Abstract

Objective. Unobtrusive EEG monitoring in everyday life requires the availability of highly miniaturized EEG devices (mini-EEGs), which ideally consist of a wireless node with a small scalp area footprint, in which the electrodes, amplifier and wireless radio are embedded. By attaching a multitude of mini-EEGs at relevant positions on the scalp, a wireless 'EEG sensor network' (WESN) can be formed. However, each mini-EEG in the network only has access to its own local electrodes, thereby recording local scalp potentials with short inter-electrode distances. This is unlike using traditional cap-EEG, which by the virtue of re-referencing can measure EEG across arbitrarily large distances on the scalp. We evaluate the implications and limitations of such far-driven miniaturization on neural decoding performance. *Approach.* We collected 255-channel EEG data in an auditory attention decoding (AAD) task. As opposed to previous studies with a

*This work was carried out at the ESAT laboratory of KU Leuven and has received funding from FWO project no. G0A4918N and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 802895), and from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.
Corresponding author: A. Mundanad Narayanan (abhijith@esat.kuleuven.be)

lower channel density, this new high-density dataset allows emulation of mini-EEGs with inter-electrode distances down to 1 cm in order to identify and quantify the lower bound on miniaturization for EEG-based stimulus decoding. *Main Results.* We demonstrate that the performance remains reasonably stable for inter-electrode distances down to 3 cm, but decreases quickly for shorter distances if the mini-EEG nodes can be placed at optimal scalp locations and orientations selected by a data-driven algorithm. *Significance.* The results indicate the potential for the use of mini-EEGs in a WESN context for AAD applications and provide guidance on inter-electrode distances while designing such devices for neuro-steered hearing devices.

1 Introduction

The possibility to continuously record electroencephalography (EEG) in daily life holds great potential towards enabling a range of applications which depend on long-term neuromonitoring in real-life settings. For example, neuro-steered hearing aids [1]–[3], brain-machine/computer interfaces [4]–[7] and ambulant monitoring of brain disorders [8] are just a few of the many possible applications. In addition, EEG recordings in real-life settings will pave the way towards the translation of knowledge obtained from controlled laboratory experiments to natural settings [9]. Steady advancements in the development of wearable, miniaturized and concealable EEG sensor devices with small area footprints (mini-EEGs) are paving the way towards such applications. Several form factors for such mini-EEGs or patches have been proposed, such as, e.g., in-ear EEG sensors [10]–[13], flex-printed EEG sensors or patches [7], [14]–[17], polymer tattoo electrodes [18], [19], etc. However, several of the envisaged applications for such mini-EEGs currently rely on algorithms and methods that were developed based on standard cap-EEG, where scalp potentials are measured over large distances and with many electrodes. Therefore, it is essential to study the impact of EEG miniaturization for such applications, not only in terms of a reduced number of electrodes, but also in terms of reduced inter-electrode distances, i.e. the distance between the reference and the recording electrode. While electrode selection is often studied in the literature, miniaturization effects are often not accounted for or only evaluated on an ad hoc basis for a specific device at a pre-defined position [12], [20].

In this paper, we investigate the impact of reducing the inter-electrode distance in a common neural decoding task, namely stimulus reconstruction, in which the time course of a sensory stimulus is decoded from multi-channel EEG data. A relevant application in this context is auditory attention decoding (AAD) [1], [2], [21], [22] where the goal is to detect the attended speaker in a multi-talker environment. By reconstructing the envelope of the attended speech from the EEG data, and comparing this decoded envelope to the actual envelopes of all speech signals in the environment, an AAD algorithm is able to identify which speaker is the attended one. This has applications in, e.g. cognitive control of hearing aids [3] to inform the hearing aid's noise reduction algorithm which speaker should be enhanced in a multi-speaker scenario. Such AAD algorithms are typically developed and tested using standard cap-EEG recordings, either in laboratory settings [1], [21]–[24] or real-life settings [4]. However, using this in everyday life, e.g. in the hearing aid context, requires unobtrusive continuous neuromonitoring which is enabled by mini-EEGs. Unfortunately, decoding performances generally decreases when using such mini-EEGs instead of standard EEG caps [2], [20], [25], due to the reduction in spatial coverage and channel count. However, in [26] we suggested that the performance of mini-EEGs can be improved by combining a multitude of them. This concept of simultaneous use of multiple mini-EEGs, connected wirelessly in a sensor network-like architecture, is referred to as a wireless EEG sensor network (WESN) [26]–[28]. WESNs mitigate drawbacks of miniaturization like lower number of channels and limited spatial coverage while preserving the advantages of mini-EEGs in terms of miniaturization, absence of wires, and flexible positioning, thereby allowing for unobtrusive neuromonitoring in a real-life setting. One of the first hardware prototypes of this WESN concept appeared in [17], where it was referred to as 'EEG dust', in which multiple stand-alone EEG sensors wirelessly communicate with each other.

Even though a WESN can aggregate EEG measured at different scalp locations by the constituent mini-EEGs, it can only rely on EEG potentials between electrodes *within* a mini-EEG node, but not on potentials between two electrodes of *different* nodes. This is because of the lack of a common galvanic reference point to link them, i.e., the EEG signals of each mini-EEG device float relative to each other based on an unknown reference signal. Traditional cap-EEG measurements, on the other hand, generally have such

a common reference[1], such that the scalp potentials across *any* pair of electrodes can be deduced by means of re-referencing. The latter simply consists of subtracting the two potentials recorded by each electrode, thereby removing their common galvanic reference point. The signal recorded by a mini-EEG node can then be emulated as the subtraction of the signals from two neighboring electrodes in a cap-EEG recording [26], [29]. This implies that we assume that a mini-EEG node consists of two regular disc electrodes where one of the two acts as a local reference. An alternative approach to extract local EEG recordings from cap-EEG consists of computing a Laplacian weighted linear combination across a cluster of nearby electrodes [30], [31], which is claimed to reduce volume conduction effects, thereby improving the spatial resolution in case local sources are to be extracted. However, in our study we assume a data-driven decoder design, which implies that any prior linear transformation on the data (including a Laplacian weighting) becomes obsolete due to the decoder optimization, which automatically selects optimal channel weights for stimulus reconstruction. Furthermore, applying a Laplacian transformation requires at least 5 electrodes within a single mini-EEG to record one channel, which may be undesirable (although the transform can also be mimicked directly in hardware through the use of tailored electrode topologies such as concentric ring electrodes [31]).

In [26] to study miniaturization effects in WESNs were studied via emulation of mini-EEGs by re-referencing each electrode in a 64-electrode cap to its neighboring electrodes. This resulted in mini-EEGs with an average inter-electrode distance of 3.7 cm. It was found that this miniaturization does not significantly affect AAD performance compared to an equal amount of EEG channels without miniaturization constraints, if optimal electrode locations are selected. However, the limited electrode density prohibited reducing the average inter-electrode distance of the emulated mini-EEGs below 3.5 cm, such that the actual miniaturization limits remain unknown.

In this paper, we present a new data set in which an AAD task is performed while recording with ultra-high density (255-channel) EEG, which allows emulation of mini-EEGs with inter-electrode distances down to 1cm. We use the utility-based greedy method [32] to select the optimal placement and orientation of a pre-defined number of emulated mini-EEGs. We then analyse the decoding performance as a function of the inter-electrode dis-

[1]This common reference can be a single (reference) electrode or a virtual reference generated in hardware or software.

tance, where the high spatial resolution then allows to identify the break point where miniaturization starts to significantly affect AAD performance. We found that an inter-electrode distance of < 3 cm leads to performance drops which are significantly worse than with an equal number of standard (long-distance) EEG channels.

To support these findings, we used the "Bayesian Estimation Supersedes t-test" (BEST) framework [33] to analyse our results, instead of the traditional $p-$value analysis in previous studies [26]. The use of BEST allows to estimate parameters like the group means together with probabilities for these estimates. These probabilities can be directly compared to confidence levels and thereby allow to make inferences, not only on the presence, but also on the *absence* of effects. Moreover, since BEST estimates are entirely data dependent, no corrections for multiple comparisons are required [34], which is an important advantage due to the many variables in our analysis (different number of nodes/channels and different inter-electrode distances).

Summary of Contributions: We collected a new ultra-high density 255-channel EEG-AAD dataset with the objective of emulating mini-EEGs with inter-electrode distances as small as 1 cm. The dataset is publicly available at `https://zenodo.org/record/4518754` [35] which can be useful for research requiring high spatial resolution in EEG. We investigated AAD performance as a function of distance between electrodes, which is a first such study up to our knowledge. We found a significant drop in performance when using mini-EEGs with an inter-electrode distance < 3 cm placed at optimal locations on the scalp.

2 Methods

2.1 EEG data collection

In order to study miniaturization effects, we collected 255-channel EEG data during an AAD task. The EEG was recorded using a SynAmps RT device (Compumedics, Australia) with active Ag/Cl electrodes at a sampling rate of 1 kHz. The electrodes were placed on the head according to the international 10-5 (5%) system. The Cz electrode was used as the reference. 30 normal hearing male subjects between 22 and 35 years old participated in the experiment. All of them signed an informed consent form approved by the KU Leuven ethical committee.

During the AAD experiment, two Dutch stories narrated by different male speakers were each divided into two parts of 6 minutes each. In each trial, two parts (one of both stories) were simultaneously presented to the subject through insert phones (Etymotic ER3A) at 60dBA, and the subject was asked to attend to only one of them while ignoring the other. The goal of AAD is then to decode from the EEG to which of both speakers the user is attending (see Section 2.2). It is noted that both speakers are audible in both ears, yet the speech signals were filtered using a head-related transfer function (HRTF) such that the stories seemed to arrive from two distinct spatial locations, namely left and right with respect to the subject with 180 degrees separation. Per subject, four experiment trials of 6 minutes each were carried out, in which each story part is used twice (once as attended story, and once as unattended story). The order of presentations was randomized and balanced over different subjects. One such presentation is illustrated in Fig. 1. For the sake of reproducibility, we have made the data set publicly available online [35]. In all the analyses in this paper, we only consider 27 of the total 30 subjects. The AAD performance of the 3 excluded subjects, namely S11, S12 and S8 in the online data set, exceeded the outlier threshold set as $> 1.5 \times$ interquartile range(IQR) above the 75-th percentile. The AAD performance metric used here is the minimal expected switching duration (MESD), described in Section 2.6, using EEG data from all the 254 Cz-referenced channels (higher the value of MESD, poorer the performance).

We implement a completely data-driven approach to study the impact of miniaturization and intend to avoid any ad-hoc heuristic which may bias our results. Therefore, no pre-processing steps like artifact removal or elimination of noisy channels were carried out here. In both LS-based neural decoding and UB-G channel selection method, the data-driven methods were observed to be robust against low-quality channels, i.e., they will receive a low weight in the LS-based decoder, and they will be eliminated in the UB-G channel selection (assuming they are a minority).

2.2 Stimulus reconstruction

A multitude of AAD algorithms have been developed to decode auditory attention from EEG [21], [23], [36]–[38]. The most widely adopted technique currently is to train a linear spatio-temporal decoder that aims to reconstruct the attended speech envelope from the EEG data in a least-squares (LS) sense, which was originally proposed in [21]. The envelopes of the different

Attention: L	Attention: R	Attention: L	Attention: R
Trial 1	Trial 2	Trial 3	Trial 4
Left: Story 1-Part 1 Right: Story 2-Part 1	Left: Story 1-Part 2 Right: Story 2-Part 2	Left: Story 2-Part 3 Right: Story 1-Part 3	Left: Story 2-Part 4 Right: Story 1-Part 4

Figure 1: Experiment protocol: Schematic showing one presentation of the AAD experiment. The experiment consisted of 4 trials with two stories presented through insert ear-phones, such that the sound of stories seemed to arrive from either left or right. During each trial the subject is asked to attend to one direction, left or right.

competing speakers are then correlated with this reconstructed envelope, and the one showing the highest correlation is then selected as the attended speaker. In this paper, we adopt this approach and combine it with a channel selection method for optimal node placement/selection (see Section 2.4)

LS-based stimulus reconstruction aims to design a spatio-temporal EEG decoder $\hat{\mathbf{w}}$ which minimizes the squared error between the decoder output and T samples of the stimulus $\mathbf{s}_a \in \mathbb{R}^T$ that elicited the relevant neural response (in our case $\mathbf{s}_a$ corresponds to the attended speech envelope, which is assumed to be known during a training phase):

$$\hat{\mathbf{w}} = arg\, min_{\mathbf{w}} ||\mathbf{A}\mathbf{w} - \mathbf{s}_a||^2 \tag{1}$$

where $\mathbf{A}$ is a $T \times QC$ matrix containing T time samples of the C EEG channels and $Q-1$ non-causal time-lagged copies of each channel in its columns. The time-lagged copies are non-causal with respect to the time axis of the stimulus, because the EEG response comes *after* the stimulus. Therefore, reconstructing the value of the stimulus at sample index t requires EEG samples at sample indices $\{t, t+1, \ldots, t+Q-1\}$.

Each channel and its $Q-1$ time-lagged copies are assumed to be grouped in adjacent columns in the matrix $\mathbf{A}$. The solution of (1) is given by:

$$\hat{\mathbf{w}} = \mathbf{R}^{-1}\mathbf{r} \tag{2}$$

where $\mathbf{R} = \mathbf{A}^T\mathbf{A}$ and $\mathbf{r} = \mathbf{A}^T\mathbf{s}_a$, corresponding to an autocorrelation matrix and a cross-correlation vector, respectively. Although the addition of a regularizer to $\mathbf{R}$ in (2) is suggested in some previous works [21], [24], this is not necessary if sufficient training data is available to construct $\mathbf{R}$ such that it accurately represents the true underlying EEG data covariance matrix [23].

In the validation experiments reported in this paper, the EEG data of each subject is divided into $\kappa = 1, \ldots, K$ epochs of t samples and a leave-one-epoch-out cross-validation procedure is carried out. That is, each epoch is used once as a test epoch while all the other epochs are used to train the decoder $\hat{\mathbf{w}}$ using (2). As proposed in [23], we compute $\mathbf{R}$ and $\mathbf{r}$ over $T - t$ data points, thereby omitting the need for regularization (as opposed to [21] where (2) is computed per t-sample epoch and then averaged over all $K - 1$ epochs).

For each epoch κ, the stimulus is reconstructed using

$$\hat{\mathbf{s}}_a = \mathbf{A}_\kappa \hat{\mathbf{w}}_\kappa, \tag{3}$$

where $\mathbf{A}_\kappa$ contains the rows of $\mathbf{A}$ corresponding to epoch κ and where $\hat{\mathbf{w}}_k$ is computed from (2) after removing the data from epoch κ. The Pearson correlation is then computed between the reconstructed stimulus $\hat{\mathbf{s}}_a$ and the true speech envelope of both speakers, and the one with the highest correlation is selected as the attended speaker. The AAD accuracy is then defined as the percentage of correct decisions across all epochs.

2.3 WESN Emulation

Similar to [26], we re-reference the EEG signals collected with a regular EEG cap to nearby electrodes in order to emulate mini-EEGs, from hereon referred to as a 'node' of the WESN. In [26], each of the electrodes were paired once with all other electrodes that are at a distance of at most $r = 5$ cm, leading to an average inter-electrode distance of 3.7 cm. In this paper, we aim to better control the inter-electrode distance, in order to analyse how the neural decoding performance is influenced by it. However, due to the high electrode density in our data, varying only the maximum inter-electrode distance r will result in a pool of electrode pairs with highly varying inter-electrode distances, in particular for larger values of r. For example, a pool of electrode pairs with $r = 5$ cm will also contain approximately 300 pairs with inter-electrode distance between 1 cm and 1.5 cm and 600 pairs with inter-electrode distance between 3 cm and 3.5 cm, which are smaller than the intended distance of $r = 5$ cm. To avoid this, we define ring-neighborhoods instead, which is explained below and formalized in Algorithm 1 in Appendix B.

For each electrode, we define a candidate neighborhood with the electrode at its center. The neighborhood is designed to be in the form of a ring around

the center electrode, with an inner radius of L_{inner} cm and outer radius [2] of L_{outer} cm as illustrated in Fig. 2. The electrodes which fall within its candidate neighborhood are paired with the center electrode to form a set $\mathcal{P}_k$ of two-electrode single channel candidate nodes where k refers to the index of the center electrode. Each electrode pair in $\mathcal{P}_k$ represents a single-channel mini-EEG sensor with a particular position and orientation. In the case where electrode k has no neighbors within a range of L_{outer} cm, leading to an empty candidate-node set $\mathcal{P}_k$, small increments each equal to ϵ cm are added to L_{outer} until $\mathcal{P}_k$ contains at least one candidate node. The sets of candidate nodes of all electrodes $k = 1, \ldots, 255$ are then pooled into a single set $\mathcal{P}$ (after removing all duplicate electrode pairs).

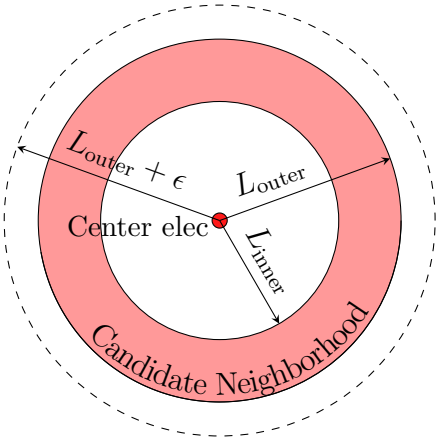


Figure 2: Candidate-node set creation: Defining neighborhoods in the form of rings to represent a range of inter-electrode distances corresponding to a set of candidate mini-EEG nodes.

We set $\epsilon = 0.1$ and used the values $L_{\text{inner}} = 1$ cm, 1.5 cm, $\ldots, 3.5$ cm and $L_{\text{outer}} = L_{\text{inner}} + 0.5$ cm leading to the creation of six candidate-node sets $\mathcal{P}$ with different average inter-electrode distances as listed in Table 1.

It can be observed in the table that the total number of candidate nodes is substantially larger for the last two sets with rings (3 cm-3.5 cm) and (3.5 cm-4 cm). As a result, larger candidate-node sets will introduce more degrees of freedom in the node selection procedure (see Section 2.4), thereby possibly

[2]Distances were calculated in 3D space using the standard spherical coordinate of electrodes on the EEG cap. However, since L_{outer} is at most 4 cm, the influence of the curvature of the head is considered to be negligible.

Label	$(L_{\text{inner}}, L_{\text{outer}})$	Avg. distance	# Nodes	Corrected # nodes
$\mathcal{S}_1$	(1 cm, 1.5 cm)	1.47 cm	311	NA
$\mathcal{S}_2$	(1.5 cm, 2 cm)	1.78 cm	381	NA
$\mathcal{S}_3$	(2 cm, 2.5 cm)	2.21 cm	432	NA
$\mathcal{S}_4$	(2.5 cm, 3 cm)	2.94 cm	377	NA
$\mathcal{S}_5$	(3 cm, 3.5 cm)	3.27 cm	633	383
$\mathcal{S}_6$	(3.5 cm, 4 cm)	3.8 cm	561	367

Table 1: Details of candidate-node sets: All the (L_{inner}, L_{outer}) pairs, average inter-electrode distances, and total number of candidate nodes created. For the largest two distances, a reduction of candidate nodes was carried out to create a similar number of nodes as in the sets with shorter distances. A label is also assigned for each candidate-node sets for reference in the paper.

leading to better decoding performance. To properly assess the influence of the average inter-electrode distance, this confound should be removed, i.e., the total number of candidate nodes of all the sets should be similar so that the only variable between the sets is the inter-electrode distance. To this end, we used the following procedure to reduce the total number of candidate nodes of the larger sets to numbers comparable to the smaller sets which is formalized in Algorithm 2 in Appendix B.

Assume $\mathcal{P}$ and $\mathcal{Q}$ are two sets of candidate nodes with $|\mathcal{P}| \ll |\mathcal{Q}|$. The goal is to create a new candidate-node set $\hat{\mathcal{Q}} \subset \mathcal{Q}$ such that $|\hat{\mathcal{Q}}| \approx |\mathcal{P}|$. Furthermore, we aim to preserve those electrode pairs from $\mathcal{P}$ that have a similar orientation as the nodes in $\mathcal{Q}$, such that they are sensitive to the same set of dipoles. To this end, for each node p_n in $\mathcal{P}$ we select the node from $\mathcal{Q}$ which shares an electrode with p_n, and which has the closest angular orientation to p_n. If, eventually, not all the electrodes from the 255-electrode cap are present in nodes of $\hat{\mathcal{Q}}$ based on this criterion, we add one extra random node per missing electrode. While this second step might result in a set $\hat{\mathcal{Q}}$ with slightly more nodes than $\mathcal{P}$, it ensures complete scalp coverage. For the reduction of the two sets mentioned earlier, we set $\mathcal{Q}$ equal to the set denoted as $\mathcal{S}_4$ in Table 1.

2.4 Utility-based greedy channel selection

In the remaining of this paper, we will use the term 'channel selection' to refer to selecting channels of an EEG cap as well as for the selection of nodes of a WESN, since each node consists of 2 electrodes, which collect a single EEG channel. We introduce the parameter C to denote the total number of available channels (either in the EEG cap or the WESN).

To select the best N out of C channels, we use the least-squares utility-based greedy (UB-G) method from [32], [39], which was demonstrated to outperform several other commonly used EEG channel selection methods in neural decoding tasks [26], [32]. In short, the method starts from the decoder that uses all C channels, and iteratively removes the channel with the lowest 'utility' until N channels remain. The utility of a channel is defined as the increase in the squared error loss if the channel were to be removed, and the decoder would be re-optimized for the remaining channels. As the re-optimization of the decoder for each potential channel removal in each iteration is computationally infeasible for large values of C, the key ingredient of the method is a recursive algebraic trick which allows to compute these channel utilities with a sufficiently low computational cost [39]. The UB-G method is briefly summarized in Appendix A.

It is important to note here that the pool of candidate nodes in a WESN (see Table 1) is typically larger than the original number of EEG channels ($C = 255$). The larger number of candidate nodes implies that there will be redundancy in the data matrix $\mathbf{A}$ as a linear dependence relationship exists between the EEG signals corresponding to different candidate nodes. Thereby, the columns of $\mathbf{A}$ are not linearly independent. To elaborate, as the signal recorded by electrodes E_1-E_2 can be obtained by subtracting the signal E_2-E_3 from the signal E_1-E_3, the columns of $\mathbf{A}$ are not linearly independent. As a result, the inverse of the autocorrelation matrix computed over all EEG channels ($\mathbf{R}^{-1}$ in (6)) does not exist and the decoders to be estimated ($\mathbf{w}$ and $\mathbf{w}_{-k}$ in (5) in Appendix A) are ill-defined. This requires a modification to the utility metric, which includes a minimum-norm constraint. For more details on this generalized utility metric, we refer to [39]. In the experiments in this paper, we used the UB-G toolbox [40], which also includes this generalized utility metric for linearly dependent variables.

2.5 Parameter settings

We filtered the EEG data using a bandpass filter between $1-9$Hz followed by downsampling to 20Hz. The speech envelopes were computed as in [23], i.e., we first pass the speech signals through a gammatone filterbank. Each of the corresponding subband signals is rectified and compressed with a power-law operation with exponent 0.6, filtered between $1-9$ Hz and downsampled to 20Hz, after which all resulting subband envelopes are summed across all subbands. For each subject, we estimated a subject-dependent linear decoder $\hat{\mathbf{w}}$ based on (2), where we used $Q = 6$ time lags corresponding to all sample delays up to 250 ms post stimulus. It has been shown that the time delays up to 250 ms are the most effective for reconstructing envelopes using EEG for the sake of attention decoding [21], [23]. Within these delays, the delays between 140 ms and 200 ms have been shown to be the most discriminative to decode auditory attention to speech [21], [25].

We implement a completely data-driven approach to study the impact of miniaturization and intend to avoid any ad-hoc heuristic which may bias our results. Therefore, no pre-processing steps like artifact removal or elimination of noisy channels were carried out here. In both LS-based neural decoding and UB-G channel selection method, the data-driven methods were observed to be robust against low-quality channels, i.e., they will receive a low weight in the LS-based decoder, and they will be eliminated in the UB-G channel selection (assuming they are a minority).

Using the procedure for WESN emulation described in Section 2.3, we created 6 sets $\{\mathcal{S}_1, \ldots, \mathcal{S}_6\}$ of single-channel (two-electrode) candidate nodes, each representing a different level of miniaturization. The details of each set are listed in Table 1. The best $N = 2, 3, \ldots, 15$ nodes were selected to emulate WESNs for each of the 6 sets using the UB-G channel selection method described in Section 2.4. Epochs of durations $60, 30, 20, 10, 5, 2$ and 1 seconds were used in the leave-one-epoch-out cross-validation.

2.6 Performance Evaluation

The AAD accuracy is defined as the percentage of epochs that show a stronger correlation coefficient with the attended speech envelope than with the unattended one (see Section 2.2). However, it is well-known that this accuracy strongly depends on the epoch (decision window) length over which the correlation coefficient is computed, as longer windows yield more accurate cor-

relation estimates. This leads to a fundamental trade-off between the AAD accuracy and the detection delay, where the latter increases with the decision window length [22]. For example, high AAD accuracies of more than 90% can easily be achieved when using long decision windows of 30-60 sec but at the cost of impractical detection delays in the target application of a neuro-steered hearing aid. In order to resolve this trade-off, we used the performance metric of minimal expected switching duration (MESD), as proposed in [22], which automatically identifies the optimal trade-off point between AAD accuracy and fast decisions from the perspective of a practical neuro-steered gain control system. This metric quantifies the expected time to perform a stable [3] gain switch after the user switches attention between two speakers(more details in [22]). Therefore, a lower MESD corresponds to a faster (yet stable) AAD-based gain control system, which is the goal of a neuro-steered hearing aid. The MESD metric has the major advantage that it captures the full accuracy-vs-decision time trade-off in a single-number metric, which facilitates statistical testing. We computed the MESD using the MESD toolbox [41] and its default configurations based on AAD accuracies at epochs of $60, 30, 20, 10, 5, 2$ and 1 seconds.

Since the AAD accuracy at a specific window length is still a widely used metric to evaluate AAD algorithms [21], [23], we also report the AAD accuracy using the median window length of 10 seconds among the evaluated epochs.

3 Results

3.1 Baseline performance on standard cap-EEG data

Before investigating the EEG miniaturization effects we first analyse the baseline performance on the original 255-channel EEG data described in Section 2.1. This will inform us of the best achievable AAD performance in this data set, while it also gives us insight into how the performance changes as a function of the number of channels used. We will also use this benchmark analysis to familiarize the readers with the BEST analysis framework [33], as an alternative for traditional null hypothesis significance testing (NHST). As the original data was referenced to the Cz electrode, we will refer to these

[3] Stability in this context is defined as maintaining a tolerable gain level for at least 80% of the time in the presence of AAD errors. [22]

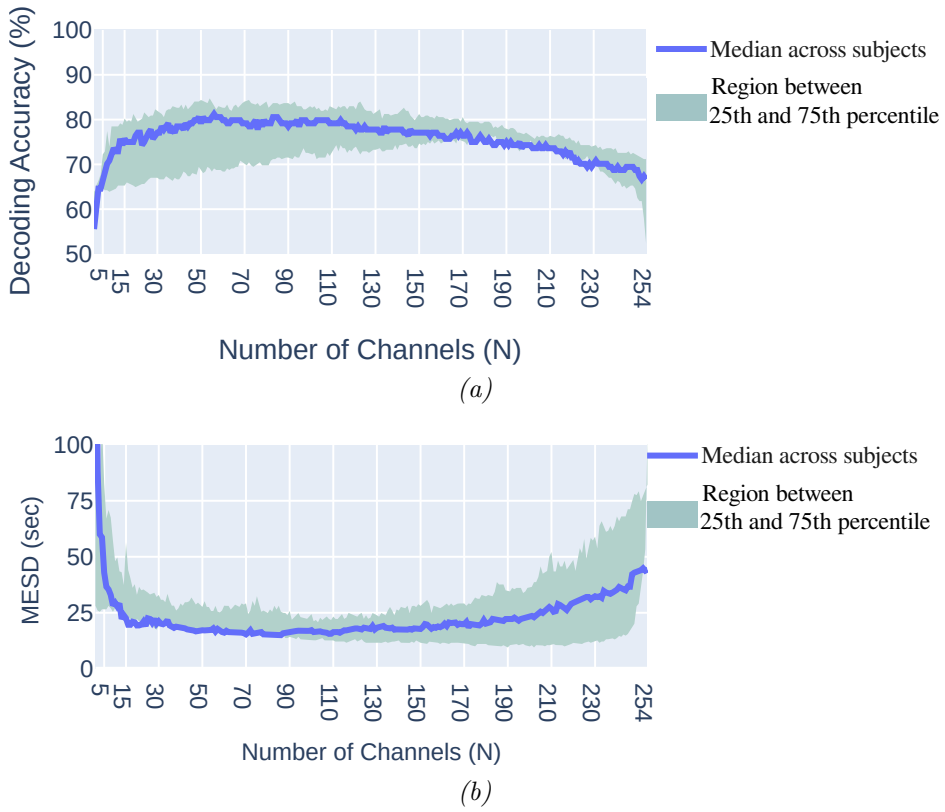


Figure 3: AAD performance across different subjects with respect to the best N Cz-ref channels $N = 1, 2 \ldots, 254$. (a) AAD accuracy (%) using 10 sec decision windows (b) MESD.

255 EEG channels as the 'Cz-ref' channels.

The AAD accuracies obtained at 10 sec decision windows and the MESDs as a function of the number of channels N (selected by the UB-G method) are shown in Fig. 3a and Fig. 3b respectively. These figures show that the AAD accuracy increases and the MESD decreases, i.e. the AAD performance improves, as the number of channels increases. A knee point in AAD performance is reached around $N = 15$ for both AAD accuracy and MESD with negligible improvement for $N > 15$. Interestingly, Fig. 3b shows that the MESD using the best $N > 210$ Cz-ref channels seems to gradually increase as the number of channels increase. This decrease in performance is also reflected in the AAD accuracy plot in Fig. 3a. We believe this is due to overfitting effects (caused by an ill-conditioning of the matrix $\mathbf{R}$ in (2)) as

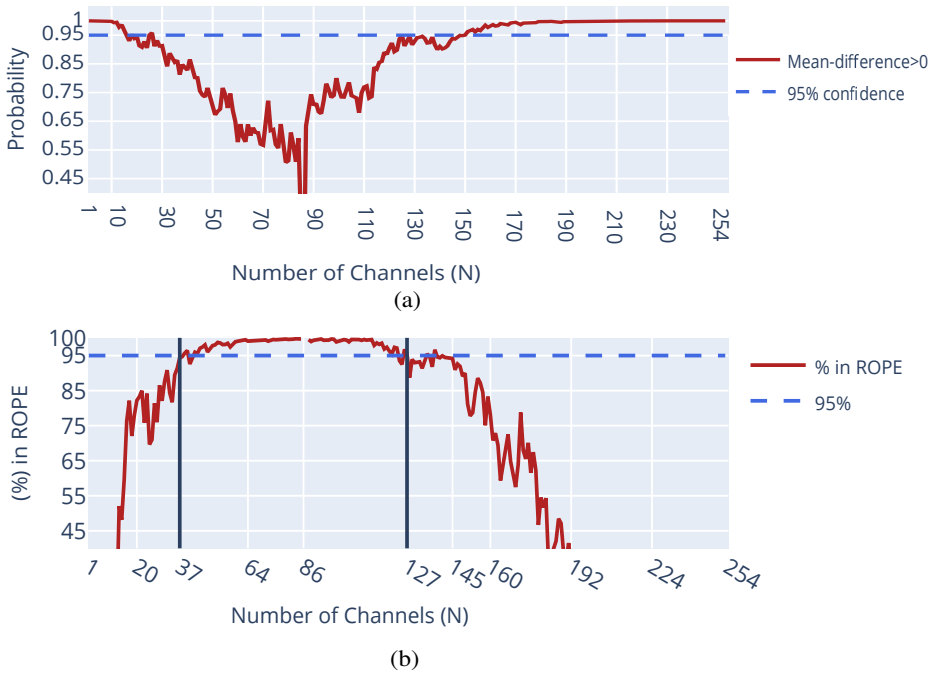


Figure 4: Analysis using BEST: (a) Probability of $\mu_N - \mu_{86} > 0$ *(b) Probability that* $\mu_N - \mu_{86}$ *falls within the ROPE. Here,* μ_N *is the mean MESD using* $N = 1, 2, \ldots, 254$ *Cz-ref channels selected using the UB-G method. The ROPE is defined as* $(-0.1 * \mu_{86}^{mean}, 0.1 * \mu_{86}^{mean})$*, where* μ_{86}^{mean} *is the baseline mean MESD using* $N = N_{best} = 86$*.*

the degrees of freedom in the decoder and the size of the matrix $\mathbf{R}$ in (2) increase linearly with the number of channels.

For the subsequent statistical analysis, we will be using only MESD as this AAD performance metric better evaluates real-world implications of an AAD algorithm. Moreover, it does not require an a-priori (arbitrary) choice for the decision window length since the MESD is computed using AAD accuracies at different window lengths including the 10 sec windows used for computing the reported AAD accuracies.

In what follows, we aim to identify the values for N at which the AAD performance is significantly lower than the optimal performance. First, we identified the number of channels N_{best} that leads to the *best* AAD performance. We defined the 'best' MESD as $\text{MESD}_{N_{\text{best}}}$, where $\text{MESD}_{N_{\text{best}}} \leq \text{MESD}_N \ \ \forall N \neq N_{\text{best}}$. We found that $N_{\text{best}} = 86$ with $\text{MESD}_{N_{\text{best}}} = 15.02$

sec. Then, we compared the MESDs at $N_{\text{best}} = 86$ with the MESDs at all other values of N to find the number of channels N for which there is a statistically significant difference with $\text{MESD}_{N_{\text{best}}}$.

To this end, the BEST analysis can be used as an alternative to NHST to compare two groups of data, by estimating the probability distributions for parameters like group means, standard deviations, normality of data, etc. These distributions can be used to obtain the probability of every possible value of these parameters [33] or to compute the probability that the (difference between) these parameters falls within a predefined interval. By computing these probabilities explicitly, BEST can be used to perform statistical tests similar to NHST but leading to more intuitive results, which -unlike traditional NHST- could even lead to accepting the null hypothesis. Moreover, since BEST estimates are entirely data dependent, no corrections for multiple comparisons are required [34]. For all the analyses in this paper, we used the BEST toolbox version 0.5.2 [42] built for software R (version 3.6.0) along with JAGS (version 4.3.0).

In Fig. 4 (a), we have plotted the probabilities of $\mu_N > \mu_{86}$ where μ_N denotes the mean MESD using the best N Cz-ref channels as selected using the UB-G method. Since BEST is a Bayesian analysis, multiple BEST results can be compared directly unlike p-values of frequentist analyses. Assuming a significance level corresponding to a probability of 95%, the plot indicates that the MESDs using the best $N < 15$ Cz-ref channels are significantly worse than using $N_{\text{best}} = 86$ channels. For values of $N > 15$, no significant difference between MESDs can be established. Nevertheless, the probabilities can be compared directly with each other in order to identify the trends. The plot shows a steady decrease in the probability of $\mu_N > \mu_{86}$ as N increases until $N = N_{\text{best}} = 86$, after which the probabilities start to increase again until $N = 150$. For $N > 150$, the MESDs obtained are again significantly different from the MESDs at $N_{\text{best}} = 86$, i.e. above the 95% significance threshold.

To prove the equivalence of two groups we establish a region of practical equivalence (ROPE) around the null value that expresses a range of parameter values that are equivalent to the null value for all practical purposes [43]. If the probability mass that is captured within the ROPE is at least 95% then we decide to accept that the two groups are equivalent for practical purposes, and otherwise we remain undecided. Here, we define this ROPE as $(-0.1 \times \mu_{86}^{\text{mean}}, 0.1 \times \mu_{86}^{\text{mean}})$, i.e. we accept a relative difference of 10% in MESD to be practically equivalent. Fig. 4 (b) shows the probability mass within

the ROPE for all values of N. The plot shows that for $N = 37, 38, \ldots, 127$, more than 95% of the mass indeed falls within the ROPE. Hence, the MESDs obtained using these values of N are practically equivalent to MESDs using $N_{\text{best}} = 86$ channels. For the range $N = 15, \ldots, 37$, the test is undecisive, i.e., there is no significant difference with $MESD_{N_{\text{best}}}$ (Fig. 4 (a)), but there is also not sufficient statistical evidence that these are within the ROPE (Fig. 4 (b)). Nevertheless, we see in Fig. 4 (b) that the probability mass is still relatively high in this range ($> 75\%$), whereas it drops very quickly for $N < 15$.

3.2 AAD Performance vs inter-electrode distance

We used the UB-G channel selection method to select the best $N = 2, \ldots, 15$ subject-dependent short-distance single-channel nodes from each of the candidate node sets listed in Table 1 and compare their performance with the best Cz-referenced channels (each time selected using the UB-G method).

The AAD performances, namely AAD accuracy using 10 sec epochs and the MESD, across subjects using the best $N = 2, \ldots, 15$ nodes selected from the six node sets are plotted in Fig. 5 and Fig. 6 respectively. The performance using a corresponding number of best Cz-ref channels is also plotted. In both Fig. 5 and Fig. 6, nodes with larger inter-electrode distances generally achieve a higher AAD performance, i.e. , higher decoding accuracies and lower MESDs.

Let μ_N^i and $\mu_N^{\text{Cz-ref}}$ be the mean MESDs across subjects using the best N nodes from the set with miniaturized nodes $\mathcal{S}_i$ and the best N Cz-ref channels respectively. BEST was used to estimate the probability of $(\mu_N^i - \mu_N^{\text{Cz-ref}}) > 0$, i.e., whether the mean MESD is larger for the miniaturized set $\mathcal{S}_i$ than for the Cz-ref set. The results are shown in Fig. 7 (a). It can be observed from the figure that for all node sets with average inter-electrode distance < 3 cm, i.e $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ the probability that μ_N^i is larger than $\mu_N^{\text{Cz-ref}}$ is larger than 95% (hence significant). Moreover, for node sets with average inter-electrode distance close to or larger than 3 cm, i.e. $\mathcal{S}_4$, $\mathcal{S}_5$ and $\mathcal{S}_6$, this probability is smaller than 95% for almost all N.

To investigate the (practical) equivalence of AAD performances obtained using the nodes from the candidate-node sets and Cz-ref channels, we first have to define a ROPE like in Section 3.1. To this end, a ROPE was defined at each value of N as $\left(0, 0.1 \times \mu_N^{\text{Cz-ref}}\right)$, i.e., we tolerate a 10% increase in MESD. The results are plotted in Fig. 7 (b), which shows that none of the

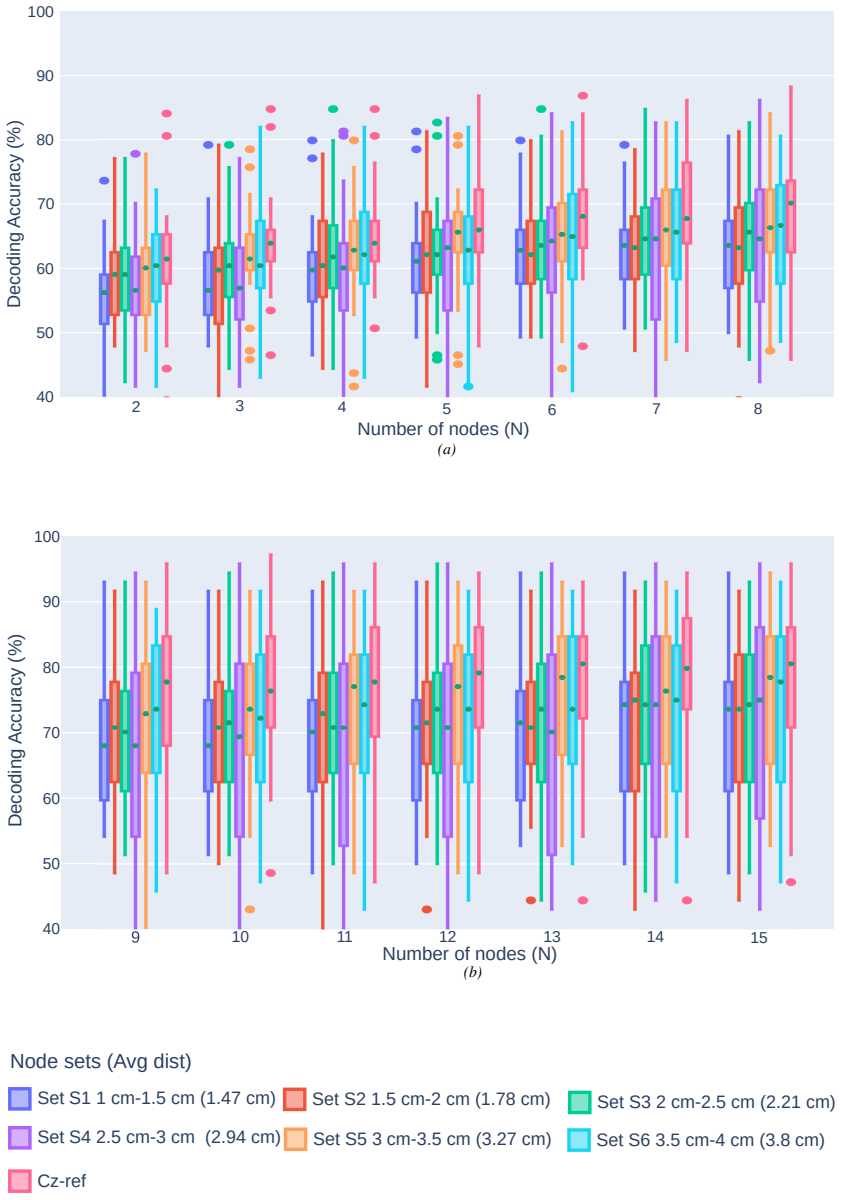


Figure 5: AAD accuracy using 10 *sec windows using the best nodes of candidate-node sets with different inter-electrode distances. (a)* $N = 2, \ldots, 8$ *nodes (b)* $N = 9, \ldots, 15$ *nodes. The colored circles represent individual subject performances which are outliers, more than* $1.5 \times IQR$ *away from the edges of the box.*

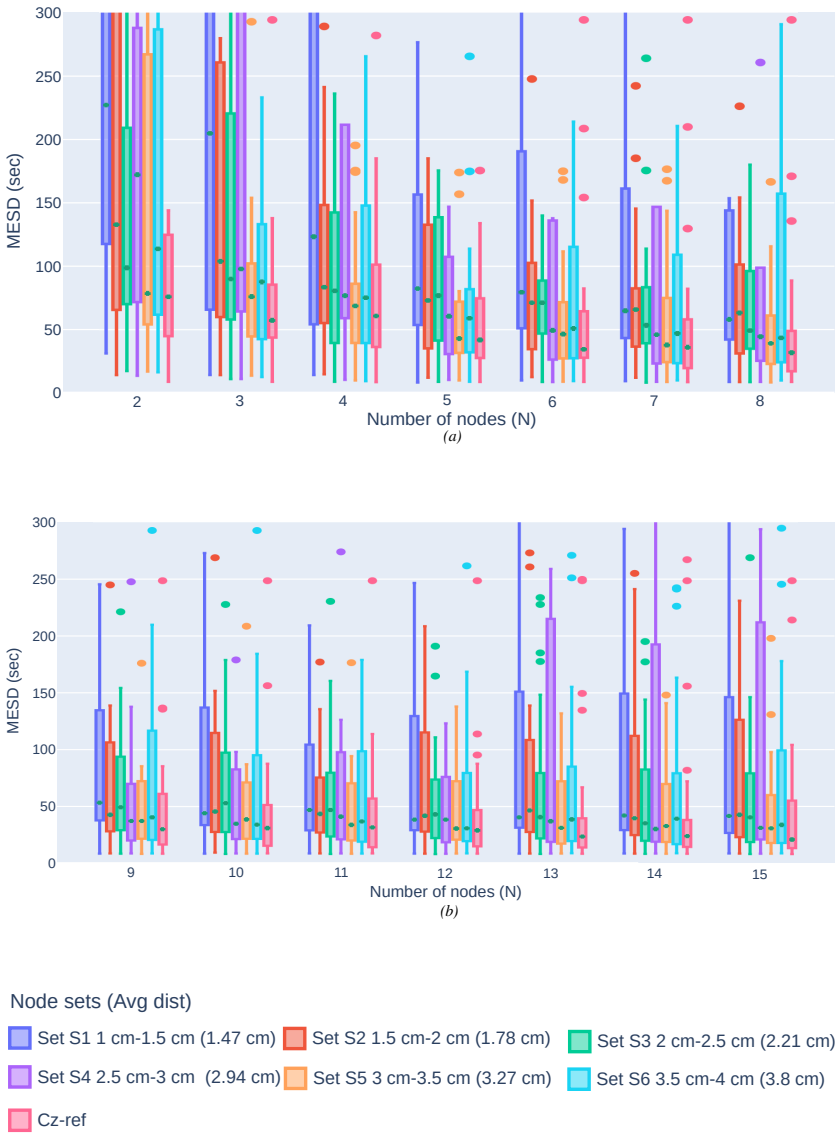


Figure 6: MESD with respect to the best nodes of candidate-node sets with different inter-electrode distances. (a) $N = 2, \ldots, 8$ nodes (b) $N = 9, \ldots, 15$ nodes. The colored circles represent individual subject performances which are outliers, more than $1.5 \times IQR$ away from the edges of the box.

probabilities reach the threshold of 95%, i.e. there is no sufficient evidence to conclude that any of the candidate-node sets achieve the same performance as the Cz-ref channels. It is to be noted that in Fig. 7 (b), for many instances of N nodes selected from the node sets with inter-electrode distances < 3 cm ($\mathcal{S}_1$, $\mathcal{S}_2, \mathcal{S}_3$), the probabilities of difference in performances falling within the ROPE is $< 5\%$. In other words, the probabilities are $> 95\%$ for these differences falling outside ROPE. We can conclude from Fig. 7 (a) and Fig. 7 (b) that the performance is significantly worse for inter-electrode distances < 3 cm for almost any N, while for larger distances of 3 cm or more ($\mathcal{S}_4, \mathcal{S}_5$ and $\mathcal{S}_6$) the (practical) equivalence or difference with Cz-ref is inconclusive. Nevertheless, all plots show a clear breakpoint at an inter-electrode distance of 3 cm, i.e., the trends clearly change above versus below this point.

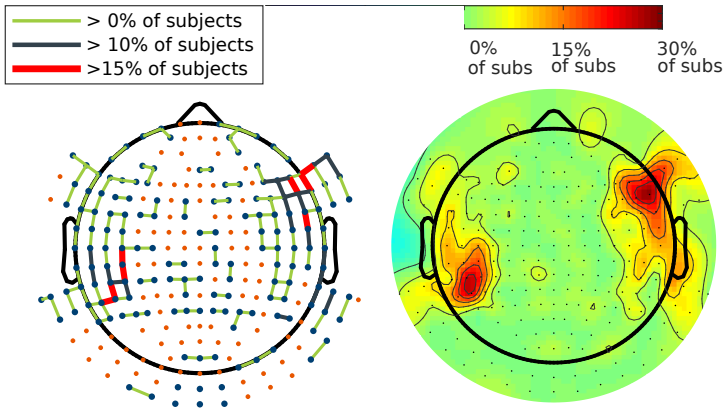
To illustrate the emulated WESN node orientations and their locations on the scalp, in Fig. 8, the distribution of nodes selected from three candidate sets, namely $\mathcal{S}_1$ (1 cm $- 1.5$ cm, Avg $= 1.47$ cm), $\mathcal{S}_3$(2 cm $- 2.5$ cm, Avg $=$ 2.21 cm) and $\mathcal{S}_5$(3 cm $- 3.5$ cm, Avg $= 3.27$ cm) is shown. We show the best $N = 10$ nodes selected from the different candidate-node sets. There are two plots per candidate-node set in the figure. In the first plot the electrode pairs representing single-channel nodes of a WESN are shown as lines between the corresponding two electrodes. There are three colors for these lines, each indicating the percentage of subjects for which this node was within the best $N = 10$ nodes; green for $\leq 10\%$ subjects, blue for $> 10\%$ & $\leq 15\%$ of subjects and red for $> 15\%$ of subjects. The second plot aggregates the node selection by regions on the scalp in the form of a heat map. To generate this heat map, for each electrode we count the number of subjects that have at least one node containing that electrode. The second plot supports the claim that for the majority of the subjects, nodes are selected that are located near the ears.

4 Discussion and Conclusion

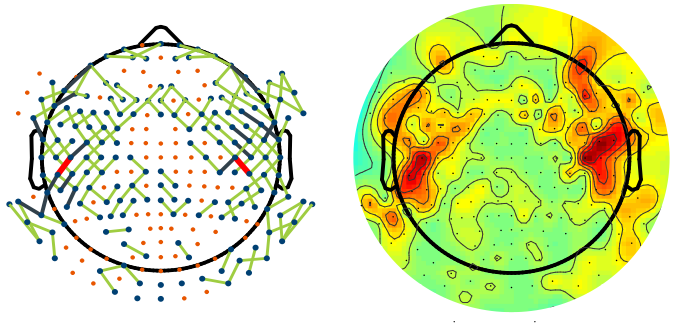
The goal of this paper was to analyse the EEG miniaturization effects on the decoding performance when using mini-EEGs. By collecting a new 255-channel EEG dataset, we were able to emulate mini-EEG nodes with inter-electrode distances down to 1 cm. To this end, we have proposed an algorithm to create different sets of candidate nodes for which the average inter-electrode distances can be controlled, leading to 6 sets with average



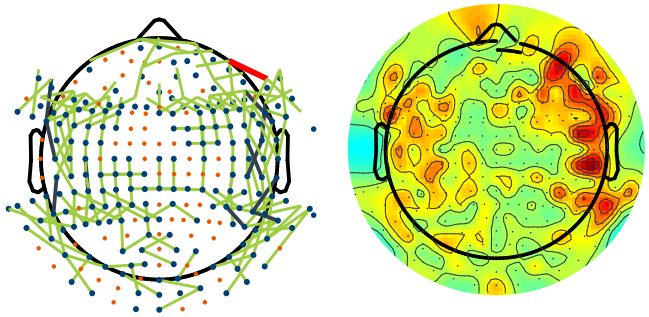
Figure 7: Analysis using BEST: (a) $P\left(\mu_N^i - \mu_N^{Cz\text{-}ref} > 0\right)$ *(b) Probability of* $\mu_N^i - \mu_N^{Cz\text{-}ref}$ *being in the ROPE.* μ_N^i *is the mean MESD using the best* N *nodes from the candidate-node set* $\mathcal{S}_i$ *and* $\mu_N^{Cz\text{-}ref}$ *is the mean MESD using the best* N *Cz-ref channels, both selected using the UB-G method. The ROPE was defined at each value of* N *as* $\left(-0.1 \times \mu_N^{Cz\text{-}ref}, 0.1 \times \mu_N^{Cz\text{-}ref}\right)$.



(a) Set S1 1 cm-1.5 cm (Avg= 1.47 cm)



(b) Set S3 2 cm-2.5 cm (Avg = 2.21 cm)



(c) Set S5 3 cm-3.5 cm (Avg = 3.27 cm)

Figure 8: Distribution of the best $N = 10$ nodes selected across all subjects, illustrated for three candidate-node sets (a) $\mathcal{S}_1$ $(1\ cm{-}1.5\ cm, Avg = 1.47\ cm)$, (b) $\mathcal{S}_3(2\ cm - 2.5\ cm, Avg = 2.21\ cm)$ and (c) $\mathcal{S}_5(3\ cm - 3.5\ cm, Avg = 3.27\ cm)$. The left plot shows the actual node links (electrode pairs), whereas the right plot shows a heat map of the selections per electrode.

distances ranging from 1.4 cm to 3.8 cm, as shown in Table 1. We have evaluated the neural decoding performance in the context of an AAD task [22]. Moreover, this 255-channel EEG dataset has been made publicly available as a supplement to this paper [35].

In a first analysis, we established the best achievable AAD performance in our data set as a baseline, and analysed how the performance is affected by the number of channels N in Cz-ref data. To this end, we used a Bayesian approach for statistical analysis, namely BEST. The BEST analysis confirmed a significant increase in MESD (decrease in AAD performance) for $N < 15$ compared to the best achievable MESD using $N = 86$ channels (see Fig. 4 (a)). In [26], [32], a similar significant drop in AAD performance was observed for $N < 10$ channels. The discrepancy between the position of the turnover point can probably be explained by the different performance metrics that were used. We have used the MESD metric, whereas [26], [32] only reports AAD accuracies for a 60 sec decision window. It is noted that such a long decision window is not representative for AAD performance in practical neuro-steered hearing aids. The MESDs seem to approach the baseline MESD at $N_{\text{best}} = 86$ when $N \geq 15$ channels are used. Furthermore, Fig. 4 (a) indicates that the MESD again increases significantly when using $N > 150$ channels. As explained in Section 3.1, we hypothesize that this is due to overfitting effects. A similar performance decrease for large channel counts was observed in [44] for the case of neural tracking of speech using a similar LS decoder (yet in a single-speaker setting), where it was also attributed to overfitting effects.

Furthermore, a region of practical equivalence (ROPE) was defined around the null (zero) mean-difference between MESDs obtained using N Cz-ref channels and $N_{\text{best}} = 86$ channels. With the ROPE defined to include all mean-differences within 10% of the mean MESD using $N = N_{\text{best}} = 86$ Cz-ref channels, we found that the MESDs using $37 \leq N \leq 127$ and $N_{\text{best}} = 86$ are practically equivalent. Note that this is not contradictive with the previous claim that a significant difference is only obtained for $N < 15$. To claim equivalence, the evidence requires to be stronger which is found only using a larger value of N. For values $16 \leq N \leq 36$, the tests are inconclusive, yet probability in Fig. 4 (b) remains high for $N > 15$ and drops very quickly for $N < 15$ indicating that the value $N = 15$ can be seen as a turn-over point.

Using the candidate mini-EEG node sets listed in Table 1, $N = 2, 3, \ldots 15$ nodes were selected from each set to emulate WESNs. We have observed that for inter-electrode distances $\sim$ 3 cm or larger, the MESDs start to

approach the performance when using an equal number of long distance, Cz-ref channels as observed in the boxplots Fig. 5 and Fig. 6. The plots show that using nodes from the $\mathcal{S}_1(1\text{ cm} - 1.5\text{ cm}, \text{Avg} = 1.47\text{ cm}), \mathcal{S}_2(1.5\text{ cm} - 2\text{ cm}, \text{Avg} = 1.78\text{ cm})$, and $\mathcal{S}_3(2\text{ cm} - 2.5\text{ cm}, \text{Avg} = 2.21\text{ cm})$ candidate-node sets leads to MESDs poorer than using a corresponding number of Cz-ref channels. Results of the BEST statistical analyses as shown in Fig. 7 (a) confirms that these performances are indeed significantly worse. However, in the case of node sets with inter-electrode distances $\geq$ 3 cm, the BEST framework's capacity to test for practical equivalence could not be leveraged in order to confirm an absence of an effect. The relative low probabilities ($< 50\%$) in Fig. 7 (b) show that it is more probable that there is a difference of more than 10% with the Cz-ref set. Nevertheless, all plots show a clear turn-over point at an inter-electrode distance of 3 cm, i.e., the trends clearly change towards a substantially lower AAD performance when the inter-electrode distance drops below 3 cm.

Another EEG miniaturization study was reported in [26], where an NHST analysis showed that the AAD performance did not significantly decrease for mini-EEGs with an average inter-electrode distance of 3.7 cm. This is indeed confirmed by our findings as well. Furthermore, our new high-density data set allowed investigating even smaller inter-electrode distances in order to find the limits of miniaturization. Our results clearly show a turn-over point at an inter-electrode distance of 3 cm, which can be viewed as a miniaturization limit beyond which the performance becomes largely affected. One hypothesis which could explain this limit is the drop in signal amplitude at these short inter-electrode distances. Such weak bio-potentials may then be highly affected by amplifier noise. In this case the turn-over point at 3 cm could potentially be shifted depending on the amplifier quality. Another reason why short inter-electrode distances may lead to a poorer performance could be the limited sensitivity to deeper sources. Indeed, compared to superficial sources, the equipotential contours related to deeper sources are more spread out on the scalp, thereby becoming unobservable when measuring over shorter distances. Nevertheless, we hope that this can be used as a guideline in the design of mini-EEG devices for neural decoding, and in particular for AAD-related applications such as neuro-steered hearing aids.

The locations of the selected nodes, shown in Fig. 8, is consistent across different candidate-node sets, such that we can infer with high probability that the performance difference observed between nodes selected from different candidate-node sets are due to the differences in inter-electrode distances.

Furthermore, most nodes are indeed located on the scalp regions near the auditory cortex. This selection makes sense physiologically as the task performed by the subjects during the recordings was an auditory task. Here, we would like to note that, the channels in the occipital region were found to be generally more susceptible to noise and artefacts due to poor electrode impedances during the recording. This is because it was sometimes hard to properly fit the cap on some subjects in this occipital region. The high density of electrodes made the cap more rigid in this region, creating difficulties to make it tightly connect to the scalp in the back of the head. Hence, it is possible that there is a slight bias towards non-occipital channels. However, similar locations as those found in our study were also found to be optimal in other studies with (lower density) EEG datasets, e.g., in [26], where the average inter-electrode distance of the selected nodes was ~ 3.5 cm. Interestingly, we can observe both in [26] and in Fig. 8, that predominantly the selected nodes are located on the right side of the scalp compared to the left side. We cannot explain this observation fully, further research is required to investigate if any significant difference between left and right scalp locations exist in an AAD task. However, these results motivate the use of mini-EEGs with form factors like behind-the-ear-EEGs [7], [14], [20], [25], [45] or in-ear EEG [10]–[13] in WESNs for AAD-based applications like neuro-steered hearing aids.

5 Acknowledgement

We would like to thank Prof. Marc van Hulle and the people at the Laboratory for Neuro- and Psychophysiology, Department of Neurosciences, KU Leuven for their help and support during EEG recording experiments.

References

[1] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *Neuroimage*, vol. 156, pp. 435–444, 2017.

[2] W. Nogueira, H. Dolhopiatenko, I. Schierholz, *et al.*, "Decoding selective attention in normal hearing listeners and bilateral cochlear implant users with concealed ear EEG," *Frontiers in neuroscience*, vol. 13, p. 720, 2019.

[3] S. Van Eyndhoven, T. Francart, and A. Bertrand, "EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1045–1056, 2016.

[4] R. Zink, S. Proesmans, A. Bertrand, S. Van Huffel, and M. De Vos, "Online detection of auditory attention with mobile EEG: Closing the loop with neurofeedback," *BioRxiv*, p. 218 727, 2017. DOI: 10.1101/218727.

[5] C. Zich, S. Debener, C. Schweinitz, A. Sterr, J. Meekes, and C. Kranczioch, "High-intensity chronic stroke motor imagery neurofeedback training at home: Three case reports," *Clinical EEG and neuroscience*, vol. 48, no. 6, pp. 403–412, 2017.

[6] J. Zhang, Z. Jadavji, E. Zewdie, and A. Kirton, "Evaluating if children can use simple brain computer interfaces," *Frontiers in human neuroscience*, vol. 13, p. 24, 2019.

[7] M. G. Bleichner and R. Emkes, "Building an ear-EEG system by hacking a commercial neck speaker and a commercial EEG amplifier to record brain activity beyond the lab," *Journal of Open Hardware*, vol. 4, no. 1, 2020.

[8] J. Dan, B. Vandendriessche, W. V. Paesschen, D. Weckhuysen, and A. Bertrand, "Computationally-efficient algorithm for real-time absence seizure detection in wearable electroencephalography," *International Journal of Neural Systems*, vol. 30, no. 11, p. 2 050 035, 2020.

[9] S. Ladouce, D. I. Donaldson, P. A. Dudchenko, and M. Ietswaart, "Understanding minds in real-world environments: Toward a mobile cognition approach," *Frontiers in human neuroscience*, vol. 10, p. 694, 2017.

[10] D. Looney, P. Kidmose, C. Park, *et al.*, "The in-the-ear recording concept: User-centered and wearable brain monitoring," *IEEE pulse*, vol. 3, no. 6, pp. 32–42, 2012.

[11] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, "In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 271–277, 2015.

[12] S. L. Kappel, M. L. Rank, H. O. Toft, M. Andersen, and P. Kidmose, "Dry-contact electrode ear-EEG," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 150–158, 2018.

[13] K. B. Mikkelsen, Y. R. Tabar, S. L. Kappel, *et al.*, "Accurate whole-night sleep monitoring with dry-contact ear-EEG," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[14] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear," *Scientific reports*, vol. 5, no. 1, pp. 1–11, 2015.

[15] S. Blum, R. Emkes, F. Minow, J. Anlauff, A. Finke, and S. Debener, "Flex-printed forehead EEG sensors (fEEGrid) for long-term EEG acquisition," *Journal of neural engineering*, 2020.

[16] D. Hoelle, J. Meekes, and M. G. Bleichner, "Mobile ear-EEG to study auditory attention in everyday life," *Behavior Research Methods*, 2021. DOI: \url{doi.org/10.3758/s13428-021-01538-0}.

[17] T. Tang, L. Yan, J. H. Park, *et al.*, "EEG dust: A BCC-based wireless concurrent recording/transmitting concentric electrode," in *2020 IEEE International Solid-State Circuits Conference-(ISSCC)*, IEEE, 2020, pp. 516–518.

[18] L. M. Ferrari, U. Ismailov, J.-M. Badier, F. Greco, and E. Ismailova, "Conducting polymer tattoo electrodes in clinical electro-and magneto-encephalography," *npj Flexible Electronics*, vol. 4, no. 1, pp. 1–9, 2020.

[19] W.-H. Yeo, Y.-S. Kim, J. Lee, *et al.*, "Multifunctional epidermal electronics printed directly onto the skin," *Advanced materials*, vol. 25, no. 20, pp. 2773–2778, 2013.

[20] M. G. Bleichner, B. Mirkovic, and S. Debener, "Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison," *Journal of neural engineering*, vol. 13, no. 6, p. 066 004, 2016.

[21] J. A. O'Sullivan, A. J. Power, N. Mesgarani, *et al.*, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, 2014.

[22] S. Geirnaert, T. Francart, and A. Bertrand, "An interpretable performance metric for auditory attention decoding algorithms in a context of neuro-steered gain control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307–317, 2019.

[23] W. Biesmans, N. Das, T. Francart, and A. Bertrand, "Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario," *IEEE Trans. on Neural Systems and Rehabilitation Engg*, vol. 25, no. 5, pp. 402–412, 2017.

[24] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos, "Decoding the attended speech stream with multi-channel EEG: Implications for online, daily-life applications," *Journal of neural engineering*, vol. 12, no. 4, p. 046 007, 2015.

[25] B. Mirkovic, M. G. Bleichner, M. De Vos, and S. Debener, "Target speaker detection with concealed EEG around the ear," *Frontiers in neuroscience*, vol. 10, p. 349, 2016.

[26] A. Mundanad Narayanan and A. Bertrand, "Analysis of miniaturization effects and channel selection strategies for EEG sensor networks with application to auditory attention detection," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 234–244, Jan. 2020, ISSN: 1558-2531. DOI: 10.1109/TBME.2019.2911728.

[27] A. Bertrand, "Distributed signal processing for wireless EEG sensor networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 923–935, 2015.

[28] B. Somers and A. Bertrand, "Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis," *Journal of neural engineering*, vol. 13, no. 6, p. 066 008, 2016.

[29] C. B. Christensen, J. M. Harte, T. Lunner, and P. Kidmose, "Ear-EEG-based objective hearing threshold estimation evaluated on normal hearing subjects," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1026–1034, 2018.

[30] B. Hjorth, "An on-line transformation of EEG scalp potentials into orthogonal source derivations," *Electroencephalography and clinical neurophysiology*, vol. 39, no. 5, pp. 526–530, 1975.

[31] G. Besio, K. Koka, R. Aakula, and W. Dai, "Tri-polar concentric ring electrode development for Laplacian electroencephalography," *IEEE transactions on biomedical engineering*, vol. 53, no. 5, pp. 926–933, 2006.

[32] A. Mundanad Narayanan, P. Patrinos, and A. Bertrand, "Optimal versus approximate channel selection methods for EEG decoding with application to topology-constrained neuro-sensor networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 92–102, 2020. DOI: 10.1109/TNSRE.2020.3035499.

[33] J. K. Kruschke, "Bayesian estimation supersedes the t test.," *Journal of Experimental Psychology: General*, vol. 142, no. 2, p. 573, 2013.

[34] A. Gelman, J. Hill, and M. Yajima, "Why we (usually) do not have to worry about multiple comparisons," *Journal of Research on Educational Effectiveness*, vol. 5, no. 2, pp. 189–211, 2012.

[35] A. Mundanad Narayanan, R. Zink, and A. Bertrand, *Ultra high-density 255-channel EEG-AAD dataset*, Accessed: 8 Feb, 2021, Feb. 2021. DOI: 10.5281/zenodo.4518754.

[36] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-time tracking of selective auditory attention from M/EEG: A Bayesian filtering approach," *Frontiers in neuroscience*, vol. 12, p. 262, 2018.

[37] A. de Cheveigne, D. D. Wong, G. M. Di Liberto, J. Hjortkjaer, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[38] G. Ciccarelli, M. Nolan, J. Perricone, *et al.*, "Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[39] A. Bertrand, "Utility metrics for assessing and selecting input variables in linear estimation algorithms," *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 93–99, Nov. 2018.

[40] A. M. Narayanan, *Channel selection in a least-squares problem*, https://github.com/AlexanderBertrandLab/channel-select, [Online; accessed 19-March-2021], 2019.

[41] S. Geirnaert, *MESD toolbox*, https://github.com/exporl/mesd-toolbox, [Online; accessed 6-October-2020], 2019.

[42] M. Meredith and J. Kruschke, *BEST package: Bayesian Estimation Supersedes the t-Test*, `https://cran.r-project.org/web/packages/BEST/vignettes/BEST.pdf`, [Online; accessed 10-November-2020], 2019.

[43] J. K. Kruschke and T. M. Liddell, "The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective," *Psychonomic Bulletin & Review*, vol. 25, no. 1, pp. 178–206, 2018.

[44] J. Montoya-Martinez, J. Vanthornhout, A. Bertrand, and T. Francart, "Effect of number and placement of EEG electrodes on measurement of neural tracking of speech," *Plos one*, vol. 16, no. 2, e0246769, 2021.

[45] M. G. Bleichner and S. Debener, "Concealed, unobtrusive ear-centered EEG acquisition: cEEGrids for transparent EEG," *Frontiers in human neuroscience*, vol. 11, p. 163, 2017.

A The UB-G method

Assuming the Q time-lagged copies of each EEG channel are in adjacent columns in the matrix $\mathbf{A}$, we can define the following partitioning for the spatio-temporal decoder $\hat{\mathbf{w}}$:

$$\hat{\mathbf{w}} = \begin{bmatrix} \hat{\mathbf{w}}_1 \\ \hat{\mathbf{w}}_2 \\ \vdots \\ \hat{\mathbf{w}}_C \end{bmatrix} \tag{4}$$

with the subvectors $\hat{\mathbf{w}}_k \in \mathbb{R}^Q \; \forall \; k \in \{1, \ldots, C\}$ capturing the decoder coefficients corresponding to the k-th channel and its Q copies.

The utility of channel k (represented by a group of columns of $\mathbf{A}$) is defined as:

$$U_k = \min_{\mathbf{w}_{-k}} ||\mathbf{A}_{-k}\mathbf{w}_{-k} - \mathbf{d}||^2 - \min_{\mathbf{w}} ||\mathbf{A}\mathbf{w} - \mathbf{d}||^2 \tag{5}$$

where $\mathbf{A}_{-k}$ is the matrix $\mathbf{A}$ with the Q columns corresponding to channel k removed. This implies that the utility is defined as the difference between the least squared error with or without the use of channel k (where the decoder $\mathbf{w}$ is re-optimized for each case).

It has been shown that this group-utility, can be computed efficiently based on $\hat{\mathbf{w}}$ without having to compute the new optimal decoder $\hat{\mathbf{w}}_{-k}$ for each channel k [39], which would quickly become computationally infeasible for large values of C. To this end, assume without loss of generality (w.l.o.g.) that the channel k and its time-lagged copies for which we compute the group-utility corresponds to the last Q columns of $\mathbf{A}$. Defining the block partitioning of $\mathbf{R}^{-1}$ in (2) as:

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y}^T & \mathbf{Z} \end{bmatrix} \tag{6}$$

where $\mathbf{Z}$ is a $Q \times Q$ matrix corresponding to the Q time lags associated with the target channel. The group-utility of channel k can be efficiently computed as [26], [39]:

$$U_k = \hat{\mathbf{w}}_k^T \mathbf{Z}^{-1} \hat{\mathbf{w}}_k \tag{7}$$

where $\hat{\mathbf{w}}_k$ contains the last Q entries of $\hat{\mathbf{w}}$. It can be shown that (7) leads to the exact same quantity as defined in (5) [39] without the need to recompute (2), which would involve a large matrix inversion for each candidate channel removal [39].

To select N (out of C) channels of EEG data used in (1), first, the utility of each of the C channels is computed using (7) followed by the removal of the channel with the least utility. After this removal, $\hat{\mathbf{w}}$ is recomputed using (2) but now with the $(C-1)$ remaining channels. The new utilities of each channel in the new $(C-1)$ channel set are re-computed from (7), again followed by removal of the channel with the least utility. The procedure is repeated until only N channels remain.

B Node creation and reduction algorithms

Algorithm 1: Candidate-node set creation

e_i, represents a cap-EEG electrode where $i = 1, \ldots, 255$;
$p_{ij} = p_{ji} = e_i, e_j$ represents a node formed by electrode pair ;
$D(p_{ij})$ represents the inter-electrode distance of node p_{ij} ;
for $k = 1 : 255$ **do**
    Let $\mathcal{P}_k = \{p_{ki} = e_k, e_i : \ L_{\text{inner}} \leq D(p_{ki}) \leq L_{\text{outer}}\}$;
    while $|\mathcal{P}_k| < 1$ **do**
        Set $L_{\text{outer}} \leftarrow L_{\text{outer}} + \epsilon$ and re-evaluate $\mathcal{P}_k$;
    end
end
$\mathcal{P} = \cup_{k=1}^{255} \mathcal{P}_k$

Algorithm 2: Candidate-node set reduction

$\mathcal{P}$ and $\mathcal{Q}$ are two sets of candidate nodes with $|\mathcal{P}| \ll |\mathcal{Q}|$;
Initialize $\hat{\mathcal{Q}} = \varnothing$, $n = 1, m = 1$;
for $n = 1 : |\mathcal{P}|$ **do**
    Take the $n-$th node p_n of $\mathcal{P}$; Let $\mathcal{Q}^n$=$\{q$: $q \in \mathcal{Q}$ **and** q and p_n share an electrode$\}$;
    Find $\hat{q} = arg\min_q \angle(q, p_n)$ s.t $q \in \mathcal{Q}^n$ **and** $q \notin \hat{\mathcal{Q}}$ where $\angle(q, p_n)$ denotes the angle (in 3D space) between the vector that connects the two electrode positions of the electrode pair q, and the vector that does the same for the pair p_n;
    Add $\hat{q}$ to $\hat{\mathcal{Q}}$;
end
while ***not*** *255 electrodes present in* $\hat{\mathcal{Q}}$ **do**
    Select a random $p \in \mathcal{Q}$ which contains one of the missing electrodes and add it to $\hat{\mathcal{Q}}$;
end