Selective Auditory Attention Decoding with a Two-Node Wireless EEG Sensor Network

Simon Geirnaert, Ruochen Ding, and Alexander Bertrand, *Senior Member, IEEE*

***Abstract*— Selective auditory attention decoding (sAAD) enables neuro-steered hearing devices by identifying the attended speaker in a multi-speaker environment from neural activity recorded with electroencephalography (EEG). Despite algorithmic progress, practical deployment remains constrained by a lack of wearable, unobtrusive, and fully wireless EEG acquisition solutions. Therefore, this work aims to evaluate whether reliable sAAD can be achieved under realistic hardware constraints imposed by using a wireless EEG sensor network (WESN) consisting of miniaturized, galvanically isolated EEG sensor nodes. Here, we use such a WESN consisting of two synchronized, compact around-ear EEG sensor nodes worn bilaterally. Each node provides four local EEG channels derived from five pre-gelled electrodes, including a local reference. Sample-wise wireless synchronization of data from both nodes enables joint processing as an eight-channel EEG. On a newly recorded dataset acquired with this setup, correlation-based stimulus decoding achieves an average sAAD accuracy of 69.24% on 60s decision windows, comparable to wired around-ear EEG systems that measure long-distance scalp potentials. Hidden Markov model-based post-processing further improves to a steady-state accuracy of 77.17% with an average simulated attention switch detection time of 32.79s. Combining sensor nodes at both ears outperforms single-ear configurations, primarily by providing redundancy that increases robustness rather than by exploiting complementary spatial information. Finally, we show that a fixed bipolar configuration using four electrodes per ear, yielding three channels, suffices to maintain performance. These results demonstrate the practical feasibility of sAAD using a fully wireless, galvanically isolated around-ear WESN and establish a realistic performance benchmark under practical hardware constraints.**

***Index Terms*— selective auditory attention, wireless EEG sensor network**

This research is funded by the Research Foundation - Flanders (FWO) project No G081722N, junior postdoctoral fellowship fundamental research of the FWO (for S. Geirnaert, No. 1242524N), the European Research Council (ERC) under the European Union's research and innovation program (grant agreement No 802895 and No 101138304), Internal Funds KU Leuven (project IDN/23/006), and the Flemish Government (AI Research Program). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the granting authorities. Neither the European Union nor the granting authorities can be held responsible for them.

S. Geirnaert, R. Ding, and A. Bertrand are with KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics and with Leuven.AI - KU Leuven institute for AI, Kasteelpark Arenberg 10, B–3001 Leuven, Belgium (e-mail: simon.geirnaert@kuleuven.be, alexander.bertrand@kuleuven.be).

S. Geirnaert is also with KU Leuven, Department of Neurosciences, Research Group ExpORL, Herestraat 49 box 721, B–3000 Leuven, Belgium.

I. Introduction

SELECTIVE auditory attention to one out of multiple competing speech streams can be decoded from neural responses, for example, recorded using electroencephalography (EEG) [1]–[3]. This capability enables the development of neuro-steered hearing devices, such as hearing aids and cochlear implants, in which the user can steer the device towards the speech stream of interest through attention-driven speech enhancement [4], [5]. Such neuro-steered hearing devices therefore hold significant potential to assist hearing-impaired listeners in challenging cocktail party scenarios. However, realizing this potential in practice requires balancing decoding performance with strict constraints on wearability, comfort, and long-term usability imposed by hearing device form factors.

In recent years, substantial progress has been made in multiple components required for neuro-steered hearing devices. From a methodological perspective, selective auditory attention decoding (sAAD) algorithms have been further developed, for example, by employing deep neural networks [6], [7], by introducing advanced post-processing strategies for correlation-based neural tracking methods [8], [9], or by developing unsupervised variants [10], [11]. In parallel, the integration of sAAD algorithms with speech separation and enhancement pipelines has been investigated [4], [12], as well as validation in more realistic acoustic environments [13], [14] and in hearing-impaired populations [15]. Despite these advances, a major - if not the most important one - remaining obstacle to practical deployment is the integration of wearable, unobtrusive neuromonitoring sensors with sAAD algorithms. In this work, we focus on this bottleneck.

EEG is commonly considered the preferred neuromonitoring modality for neuro-steered hearing devices because of its excellent temporal resolution, which is essential for neural tracking-based sAAD, and its non-invasive, relatively low-cost, and wearable nature [16]. These characteristics make EEG particularly attractive for integration into everyday hearing devices. However, the vast majority of sAAD algorithms have been developed and validated using high-density, full-scalp wet EEG systems, which are impractical for long-term, real-world use due to their bulkiness and obtrusiveness. Enabling sAAD in hearing devices therefore necessitates wearable, unobtrusive, and ideally dry EEG solutions.

One promising approach towards unobtrusive EEG monitoring is the use of a *wireless EEG sensor network* (WESN), consisting of multiple miniaturized EEG sensor nodes (mini-

EEGs) that record EEG locally at different scalp locations [17]. A defining characteristic of such WESNs is the absence of a shared electrical reference through a wired connection between sensor nodes, as the mini-EEG nodes are galvanically isolated from each other by design. As a direct consequence, each node can only record short-distance, within-node scalp potentials, in contrast to conventional EEG systems that measure long-distance potentials referenced across the scalp. This limitation introduced by miniaturization can be mitigated by deploying multiple mini-EEGs in a network, thereby providing a spatially diversified representation of neural activity that is required for effective sAAD [18]. At the same time, the absence of wired connections reduces susceptibility to electromagnetic interference, minimizes motion-induced wire artifacts, and increases robustness to individual node failures [17], [19].

Assuming idealized mini-EEGs that can be flexibly placed on the scalp, the optimal sensor locations for sAAD have been investigated in a data-driven manner for correlation-based stimulus decoding algorithms [20], [21]. Mundanad Narayanan and Bertrand [21] demonstrated that, starting from a full-scalp EEG configuration, the number of channels could be reduced from 64 to 10 using a greedy utility-based selection algorithm without loss in sAAD performance, achieving approximately 85–90% accuracy on $60\,\mathrm{s}$ decision windows. However, this analysis did not account for the constraint of short inter-electrode distances inherent to mini-EEGs. To address this, Mundanad Narayanan et al. [22] emulated mini-EEGs using ultra-high-density 255-channel EEG recordings and showed that sAAD performance remained stable for inter-electrode distances down to $3\,\mathrm{cm}$, with optimal sensor locations situated above the temporal lobe, near the auditory cortex, as expected for auditory decoding. Additionally, Mundanad Narayanan and Bertrand [21] demonstrated through emulation the potential benefit of using two-channel rather than single-channel mini-EEG sensor nodes. The addition of a second channel allows each node to record EEG potentials in two orthogonal directions, thereby allowing to capture source dipoles in any direction.

Importantly, the aforementioned studies assume idealized sensor placement. In practice, sensor locations are constrained by form-factor considerations, user comfort, and presence or absence of hair. Consequently, practical mini-EEG implementations for sAAD have primarily focused on flex-printed arrays placed around and behind the ear (e.g., cEEGrid) [18], [23]–[26] and customized in-ear EEG devices [18], [27]. Using correlation-based stimulus decoding, which relies on neural tracking, typical decoding accuracies of approximately 70% for wet around-ear EEG [18], [23], [24] and 60% for dry in-ear EEG have been reported on $60\,\mathrm{s}$ decision windows [18]. Furthermore, Geirnaert et al. [18] emulated galvanic isolation between the ears by applying same-ear rereferencing, demonstrating no performance loss for around-ear EEG, but a 4.1 percentage point decrease in accuracy for in-ear EEG. They further showed that combining ear-centered mini-EEGs with a small number of scalp electrodes can yield near full-scalp performance, for example by adding only three scalp electrodes to in-ear EEG.

Nevertheless, in all prior work on mini-EEGs and WESNs for sAAD, the sensor nodes were still physically connected by wires, and the effects of miniaturization resulting in true galvanic isolation and exclusive reliance on short-distance potentials have only been studied through emulation. In this work, we address this gap by employing a practical realization of a WESN for sAAD using two fully galvanically isolated around-ear sensor nodes. Specifically, we build on the EEG-Linx platform introduced by Ding et al. [19], using two compact, synchronized wireless EEG sensor nodes with a printed circuit board (PCB)-size of $2 \times 3\,\mathrm{cm}$, each equipped with an embedded processor and wireless radio and each providing four EEG channels. Crucially, no wired connection is present between the two ears, as each node operates independently. Ding et al. [28] previously demonstrated the feasibility of this setup for decoding auditory steady-state responses (ASSRs), showing improved robustness. Here, we extend this platform to sAAD, providing a proof of concept of a two-node, four-channel-per-node WESN for practical neuro-steered hearing devices and enabling an empirical investigation of the effects of miniaturization. We use a correlation-based stimulus decoding algorithm rather than alternative direct classification approaches that decode, for instance, the spatial focus of attention [29], [30]. Although such approaches can achieve higher accuracies at shorter decision windows, they are also much more susceptible to overfitting, shortcut learning, and overly optimistic performance estimates [31]–[34]. Moreover, we show that the accuracy can be substantially improved by incorporating the recently proposed hidden Markov model post-processing approach for correlation-based stimulus decoding [9], thereby establishing a state-of-the-art sAAD performance benchmark under the realistic hardware constraints of neuro-steered hearing devices.

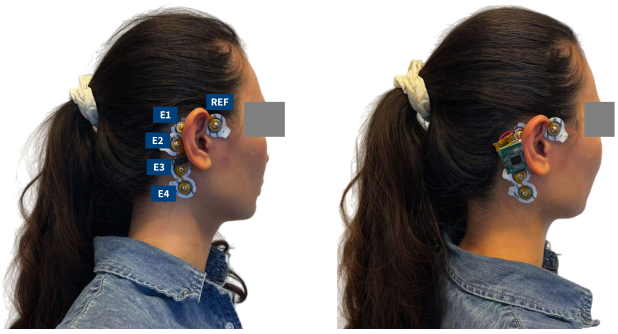
In Section II, we describe the employed sensor nodes, the EEG-Linx platform, and the recorded dataset. The sAAD algorithm, preprocessing, and evaluation procedure are detailed in Section III. The results are reported in Section IV and discussed in Section V.

II. Methods: Data Collection

This section describes the around-ear WESN platform employed for EEG data collection in Section II-A and the recorded dataset in Section II-B.

A. Wireless EEG Sensor Nodes for the WESN

EEG was recorded using two battery-powered around-ear sensor nodes worn bilaterally (one per ear; see Figure 1), which streamed data wirelessly to a hub node that was connected to an acquisition PC via USB. Each node recorded four EEG channels in a (local) common-reference montage, with four around-ear electrodes referenced to a single local electrode on the same side. This resulted in eight EEG channels in total across both ears. Since the two ear-worn nodes were galvanically isolated, referencing was necessarily performed independently per ear, restricting measurements to short-distance, same-ear potentials. The system operated without an extra driven right leg (DRL) or ground electrode, as input biasing was provided by the analog front-end circuitry on each



(a) (b)

Fig. 1: **(a)** The reusable C-shaped flex-PCB carrier with disposable snap-type pre-gelled ECG electrodes. Recording electrodes are labeled E1–E4, and the local reference electrode (REF) is placed in the preauricular region anterior to the tragus at the temporal hairline. **(b)** An example of the EEG-Linx sensor node worn behind the ear.

node [19]. The same WESN setup has been successfully used to record ASSRs in Ding et al. [28], providing initial evidence for the validity of the setup.

1) Wireless EEG sensor nodes: The wearable EEG sensor node, originally introduced and validated in Ding et al. [19], is realized on a compact double-sided, six-layer PCB (approximately $2\times3\,\mathrm{cm}$). Each node integrates (i) an EEG analog front-end, (ii) an embedded processing and wireless communication unit, and (iii) power-management circuitry. The analog front-end provides the interface to the around-ear electrodes and applies on-node biasing to set the input operating point during recordings. EEG signals were sampled at $250\,\mathrm{Hz}$ and streamed wirelessly to the hub node during the experiments. Each node was powered by a $3.7\,\mathrm{V}$, $250\,\mathrm{mAh}$ Li-Po battery.

2) Electrode interface and around-ear electrode layout: To facilitate rapid preparation and ensure hygiene across recording sessions, the electrode interface was implemented using disposable snap-type ECG electrodes mounted on a reusable flexible PCB carrier. The carrier consisted of a C-shaped flex PCB with a grid layout following the peri-auricular contour, similar to the cEEGrid configuration [35]. Integrated snap studs provided both the mechanical and electrical interface between the flex PCB and the electrode pads, allowing electrodes to be attached and replaced without tools.

Commercial foam ECG electrodes (Kendall Arbo electrodes H124SG) with a diameter of $24\,\mathrm{mm}$ and pre-applied conductive hydrogel were used as disposable skin-contact elements. No additional gel application was required. Prior to electrode placement, the skin was cleaned using a dedicated skin-preparation gel. The disposable electrodes were then snapped onto the flex PCB carrier, after which the carrier was positioned around the ear according to the intended montage.

Electrode positions along the carrier were labeled E1–E4 following their order from superior-anterior to inferior-posterior around the pinna, as shown in Figure 1a. The local reference electrode (REF) was placed in the preauricular region, anterior to the tragus at the temporal hairline. After each recording session, the disposable ECG electrodes were detached from the snap studs and discarded, while the flex PCB carrier was retained for reuse after routine cleaning.

3) Wireless acquisition, synchronization, and logging with EEG-Linx: During the experiments, the ear-worn sensor nodes transmitted EEG data wirelessly to a hub node connected to the acquisition computer via USB [19]. Since each node samples EEG using its own local oscillator, the data streams exhibit relative clock offsets and drift. EEG-Linx compensates for these effects using packet timestamps and clock-drift compensation to align the streams in time according to the clock of the hub node, which acts as the universal time for the WESN nodes. In parallel, the stimulus presentation computer transmitted trigger events to the hub node during audio playback. These triggers were recorded and time-stamped by the hub node, allowing to align the EEG data with the stimulus timeline and to segment trials in the subsequent analysis.

EEG-Linx was used to configure the sensor nodes, monitor signal quality during experimental setup, and log the synchronized EEG streams together with corresponding stimulus trigger events [19].

B. Dataset

This section describes a newly recorded sAAD dataset acquired using the wireless two-node around-ear EEG setup, as described in Section II-A. The experimental protocol was approved by the KU Leuven Social and Societal Ethics Committee. The complete dataset is made publicly available [36].

1) Participants: Sixteen adults (seven women, nine men) between 20 and 51 years of age (mean: 29.8, standard deviation: 8.4), all with self-reported normal hearing, participated in the study. All participants were proficient English speakers (non-native, except for one) and provided written informed consent prior to the experiment.

2) Protocol and stimuli: During the selective auditory attention experiment, participants were instructed to attend to one of two competing speech streams while seated in a regular room without electrical shielding. The two speech streams consisted of stories read by English-speaking male narrators (taken from `https://openslr.org`) and were presented simultaneously through two loudspeakers positioned at approximately $\pm30^\circ$ azimuth and at a distance of $1\,\mathrm{m}$ from the participant. The sound level was adjusted individually to a comfortable listening level. No additional manipulations were applied to the auditory stimuli.

The stories were divided into ten parts, resulting in an experimental protocol comprising ten trials of $5\,\mathrm{min}$ each, with a maximum total duration of $50\,\mathrm{min}$. Prior to each trial, the experimenter instructed the participant to attend to either the left or right speaker, with the attended side alternating between successive trials. Participants were allowed to take breaks between trials at their own discretion. No explicit behavioral attention assessment (e.g., comprehension questions) was performed.

For various reasons (depleted battery, premature ending of the experiment at the participant's request, (human) errors in the configuration before each trial, and other technical issues), the number of trials and duration of individual trials varied slightly across participants, sometimes leading to incomplete or missing trials. Approximately 80% of all $5\,\mathrm{min}$ trials across

the 16 participants were completed in full, resulting in an average recording duration of $42\,\mathrm{min}$ per participant (standard deviation: $6.7\,\mathrm{min}$; minimum duration: $25\,\mathrm{min}$). All trials, including incomplete ones, were retained and included in the analysis as all participants had a sufficient number of trials.

III. Methods: Data Analysis

Section III-A presents the correlation-based linear stimulus decoding algorithm for sAAD, complemented by a hidden Markov model (HMM) post-processor [9]. Preprocessing steps for the EEG and speech signals are detailed in Section III-B, and the evaluation procedure is described in Section III-C.

A. Selective auditory attention decoding algorithm

The employed sAAD framework consists of two components:

1) A correlation-based neural tracking algorithm using linear stimulus decoding to generate attentional scores for each competing speech stream (Section III-A.1).
2) An HMM-based post-processor that exploits temporal regularities in attentional processes by integrating neural tracking correlations across successive decision windows (Section III-A.2).

1) Correlation-based stimulus decoding: We employ a linear stimulus decoding approach for sAAD [3], [5], [37], which is widely regarded as one of the most robust methods and has been successfully applied in combination with around-ear EEG [18], [23], [24]. The method relies on neural tracking, whereby neural responses synchronize more strongly over time with the attended speech signal than with the unattended one [1]–[3]. A linear decoder is trained to reconstruct a representation of the attended speech signal from the EEG responses, after which the reconstructed signal is correlated with candidate speech representations to infer the attended speaker.

As a speech representation, the popular acoustic envelope is used [3], [5], [37]. Below, a brief summary of the decoding approach is provided; extensive explanations are available in the literature [3], [5], [11], [37].

Given a training set of T samples of time-lagged EEG data $\mathbf{X} \in \mathbb{R}^{T\times CL}$ (where C denotes the number of EEG channels and L the number of time lags) and the attended speech envelope $\mathbf{y}_{\text{att}} \in \mathbb{R}^{T}$, a spatio-temporal linear decoder $\mathbf{d} \in \mathbb{R}^{CL}$ is learned to reconstruct the attended envelope $\hat{\mathbf{y}} = \mathbf{X}\mathbf{d}$ by solving

$$\begin{aligned}\hat{\mathbf{d}} &= \underset{\mathbf{d}}{\operatorname{argmin}}\ ||\mathbf{y}_{\text{att}} - \hat{\mathbf{y}}||_2^2 + \lambda\,||\mathbf{d}||_2^2 \\ &= \underset{\mathbf{d}}{\operatorname{argmin}}\ ||\mathbf{y}_{\text{att}} - \mathbf{X}\mathbf{d}||_2^2 + \lambda\,||\mathbf{d}||_2^2\,.\end{aligned} \tag{1}$$

Here, $\mathbf{X}$ is a block Hankel matrix containing lagged EEG samples per channel. The time lags are chosen acausally, spanning the interval $[0, 400]\mathrm{ms}$ to capture neural responses following the stimulus [11]. This least-squares formulation is equivalent to maximizing the Pearson correlation between the reconstructed and attended speech envelopes [37]. The closed-form solution of Eq. (1) is given by

$$\hat{\mathbf{d}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}_{\text{att}}.$$

The L_2-norm regularization parameter λ is determined using the analytical estimator proposed by Ledoit and Wolf [38], which has been shown to perform well for this decoder, including in the context of around-ear EEG [18]. Training is supervised, as the attended speech envelope $\mathbf{y}_{\text{att}}$ is assumed known. Although unsupervised training strategies have been proposed that do not require attention labels [10], [11], they are beyond the scope of this work.

At test time, given the learned decoder $\hat{\mathbf{d}}$, test EEG data $\mathbf{X}^{(\text{test})} \in \mathbb{R}^{T_{\text{test}}\times CL}$ and competing speech envelopes $\mathbf{y}_1^{(\text{test})}, \mathbf{y}_2^{(\text{test})} \in \mathbb{R}^{T_{\text{test}}}$, with T_{test} the decision window length, attentional scores r_1, r_2 are computed as the Pearson correlation coefficients with the reconstructed envelope $\hat{\mathbf{y}}^{(\text{test})} = \mathbf{X}^{(\text{test})}\hat{\mathbf{d}}$:

$$r_1 = \mathrm{corr}\left(\hat{\mathbf{y}}^{(\text{test})}, \mathbf{y}_1^{(\text{test})}\right), r_2 = \mathrm{corr}\left(\hat{\mathbf{y}}^{(\text{test})}, \mathbf{y}_2^{(\text{test})}\right). \tag{2}$$

Traditionally, the speaker associated with the highest correlation is selected as the attended one. In this work, the correlation scores r_1, r_2 are additionally used as observations for HMM-based post-processing (Section III-A.2).

2) Hidden Markov model-based post-processing: Auditory attention exhibits temporal structure, as listeners typically maintain attention to the same speaker across successive time windows, with switches occurring relatively infrequently. Smart post-processors can exploit this temporal structure to compute a more reliable estimate of the attentional state based on the 'observed' attentional scores. We apply an HMM-based post-processor as proposed by Heintz et al. [9], which has been shown to substantially improve steady-state decoding accuracy and reduce switch detection latency. Note that while no switches within a trial are present in the recorded experiment, attention switches will be simulated to enable a full evaluation of the HMM post-processing performance (see Section III-C). While this simulation does not fully capture real attention switch performance or behavior, it nonetheless allows a partial evaluation of switching performance when no real switches are available. Importantly, in Heintz et al. [9], the HMM-based post-processor has already been extensively validated with real attention switches.

We employ a two-state HMM, where each state corresponds to attention to one of the two speakers, denoted by S_1 and S_2. The observations consist of normalized correlation score vectors $\mathbf{r}(t) = \left[r_1(t), r_2(t)\right]$ computed for each decision window indexed by t (see Eq. (2)). State probabilities $\mathbb{P}(\hat{S}(t) = S_j \mid \mathbf{r}_{0:t})$ are estimated using the computationally efficient forward algorithm. Note that we explicitly choose the causal version of the post-processor here, to realistically simulate the online and causal processing of attentional scores. Emission probabilities $\mathbb{P}\Big(\mathbf{r}(t) \mid \hat{S}(t) = S_j\Big)$ are modeled as Gaussian distributions conditioned on the attentional state [9], [39].

The same hyperparameter settings are chosen as in [9]: an attention switch probability of 0.001 and a decision window length of $1\,\mathrm{s}$ to generate the observed correlation scores. Note that these hyperparameters are explicitly taken from Heintz et al. [9] to guarantee honest external validation without fine-tuning them to this specific experiment and simulated switch settings. All correlation scores are normalized by subtracting the global mean and dividing by the global standard deviation

σ_{global} across all (attended and unattended) correlations prior to HMM inference. Given this normalization, Gaussian emission probability distributions are chosen with standard deviations 1 and means $\frac{0.05}{\sigma_{\text{global}}}$ (attended) and $\frac{-0.05}{\sigma_{\text{global}}}$ (unattended), assuming a realistic, yet again preset and untuned, average difference of 0.1 between attended and unattended correlations before normalization [39]. For each decision window, the estimated attentional state corresponds to the state with the highest posterior probability.

B. Preprocessing

1) EEG preprocessing: EEG recordings from both sensor nodes and the speech stimuli are processed on a per-trial basis. Separate EEG preprocessing is deliberately kept minimal and can be performed fully independently per node. However, it is important to remark that the data-driven stimulus decoding in Section III-A.1 also implicitly performs and thus handles noise reduction. EEG signals are bandpass-filtered between 1–$9\,\text{Hz}$ using a zero-phase fourth-order Butterworth filter and resampled to $20\,\text{Hz}$. This frequency band is well established for correlation-based stimulus decoding [3], [5], [18]. Finally, per-trial normalization is applied by scaling the Frobenius norm of the EEG channels to one. In case of channel selection (see Section IV-B), this normalization is performed after selection to avoid implicit overfitting.

2) Speech preprocessing: Speech envelopes are extracted following the procedure described by Biesmans et al. [37]. Speech signals are first passed through a gammatone filterbank with 19 bands spanning 50–$5000\,\text{Hz}$, mimicking cochlear filtering. For each band, the envelope is obtained by applying a power-law compression with exponent 0.6 to the absolute value of the filtered signal. Subband envelopes are then summed, bandpass-filtered between 1–$9\,\text{Hz}$ using the same Butterworth filter as for EEG preprocessing, and resampled to $20\,\text{Hz}$. Finally, envelopes are z-scored on a per-trial basis.

C. Evaluation procedure

A leave-one-trial-out cross-validation (CV) scheme is used in combination with participant-specific decoders. For each participant, decoders are trained on all but one trial and tested on the held-out trial. During training, all training trials are concatenated, where the incomplete ones contribute fewer samples due to their shorter length. Although correlation-based stimulus decoding paradigms are less susceptible to such shortcut learning [40], we adhere to this strict trial-based CV scheme as a best practice to avoid shortcut learning and overfitting in sAAD [31]–[34]. Decision window lengths are varied across $T_{\text{test}} \in \{60, 30, 20, 10, 5, 2, 1\}\text{s}$. Raw correlation-based accuracies are primarily analyzed on $60\,\text{s}$ windows, while $1\,\text{s}$ windows are used for HMM post-processing.

The significance level for AAD accuracy, at α-level $= 0.05$, is computed using a random label-permutation approach. Per participant and for each window length, the attention labels are $10\,000$ times randomly flipped with 50% probability to build a null distribution of AAD accuracies under chance. The significance level for the mean is then defined as the 95%-percentile of the distribution of averaged accuracies across participants under randomly flipped labels, while the per-participant significance levels are based on the individual null distributions. This approach automatically accounts for differences in the number of decision windows caused by incomplete trials. At the individual participant level, the consequence is that there is no global significance level; instead, each AAD accuracy is compared with the significance level for that specific participant.

For HMM evaluation, two metrics are considered: steady-state accuracy and simulated switch detection time. To enable their computation, attention switches are simulated on the correlation scores that are provided as input to the HMM. First, all $1\,\text{s}$ correlation scores are concatenated across trials and reordered such that the first score always corresponds to the attended speaker. Attention switches are then generated randomly across the entire recording at an average rate of 2.5 switches per $10\,\text{min}$, with a minimum inter-switch interval of $2\,\text{min}$. A switch is implemented by swapping the correlation scores, effectively reversing the attention labels. Switch detection time is defined as the latency required for the posterior probability of the newly attended speaker to exceed that of the previously attended speaker. Given the two-speaker setup, in practice this boils down to the time needed for the newly attended speaker to reach $\geq 50\%$ attended probability. Steady-state accuracy is computed based on all posterior probabilities across all decision windows, excluding those within the switch transition periods that determine the switch detection time to reduce cross-contamination between both metrics. This switch simulation procedure is performed 100 times to reliably estimate the average steady-state accuracy and switch detection time.

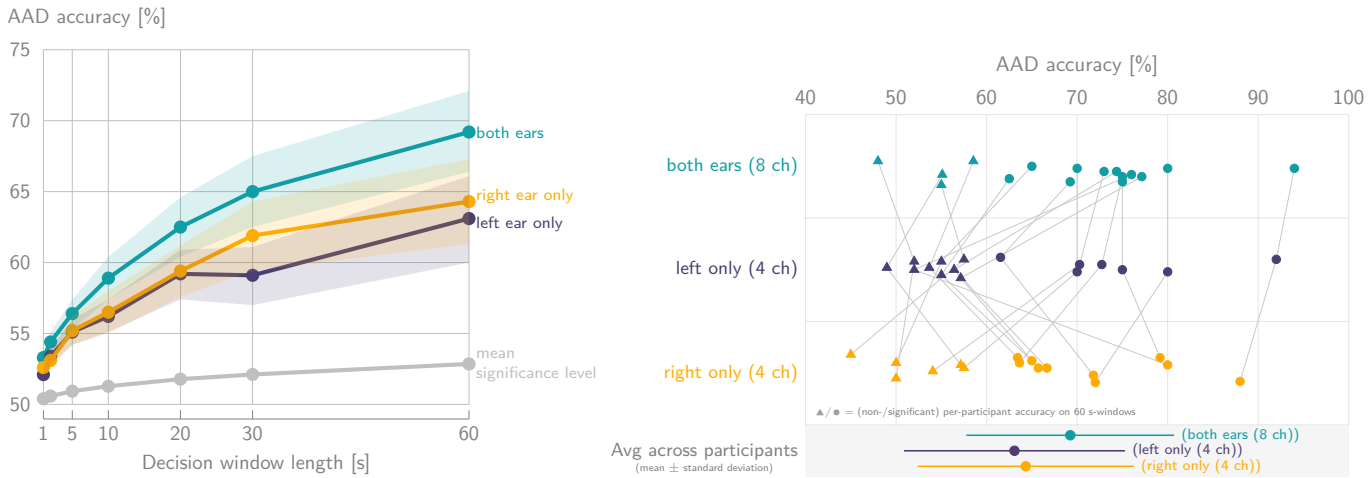
IV. Results

This section presents the experimental results obtained with the proposed wireless EEG sensor network. We first compare the performance of combining sensor nodes from both ears with single-ear configurations (Section IV-A). Next, we analyze bipolar channel selection strategies to identify practically optimal electrode configurations within the WESN setup (Section IV-B). Finally, we evaluate the impact of HMM post-processing (Section IV-C).

A. Comparison both ears combined versus single-ear configurations

We first investigate whether combining sensor nodes from both ears improves sAAD performance compared to single-ear configurations. Figure 2a shows the average sAAD accuracy as a function of decision window length, based on the raw neural tracking correlations in Eq. (2). Figure 2b reports per-participant accuracies for $60\,\text{s}$ decision windows.

When using all available channels, combining both ears yields an average accuracy of 69.24%, compared to 63.07% for the left ear only and 64.32% for the right ear only ($60\,\text{s}$ windows). All three settings exhibit significant decoding accuracy on average, as seen in Figure 2a. At the individual level for $60\,\text{s}$ windows, $12/16$ participants exhibit significant AAD accuracy for both ears, $7/16$ for the left ear only, and



(a) (b)

Fig. 2: **(a)** The average sAAD performance curves (mean ± standard error across the 16 participants; mean significance level computed based on random label-permutation, seeSection III-C) based on raw neural tracking correlations show that using wireless EEG sensor nodes at both ears simultaneously substantially outperforms single-ear configurations. **(b)** The per-participant sAAD accuracies for 60 s decision windows. Gray lines connect results from the same participant; participants with non-significant AAD accuracies (based on per-participant significance levels using random label-permutation, see Section III-C) are indicated using a triangle.

$10/16$ for the right ear only (Figure 2b). Using a paired Wilcoxon signed-rank test with Bonferroni–Holm correction ($\alpha = 0.05$, two comparisons), performance with both ears (all channels) is significantly higher than with the left ear only (uncorrected $p = 0.0061$, corrected $p = 0.0122$; $W = 8$, rank-biserial $r = 0.727$, $N = 16$) and the right ear only (uncorrected $p = 0.0386$, corrected $p = 0.0386$; $W = 28$, rank-biserial $r = 0.517$, $N = 16$).

To investigate how combining both ears compares to single-ear configurations across a varying number of channels, Figure 3 illustrates the effect of limiting the number of channels per ear on the sAAD accuracy based on the raw correlations in Eq. (2). Across all channel combinations (taken from the original set of four channels per ear) and participants, combining both ears consistently outperforms single-ear configurations by approximately 5 percentage points, regardless of the number of channels per ear. For the combined-ear configuration, selected channels are always matched across both ears.

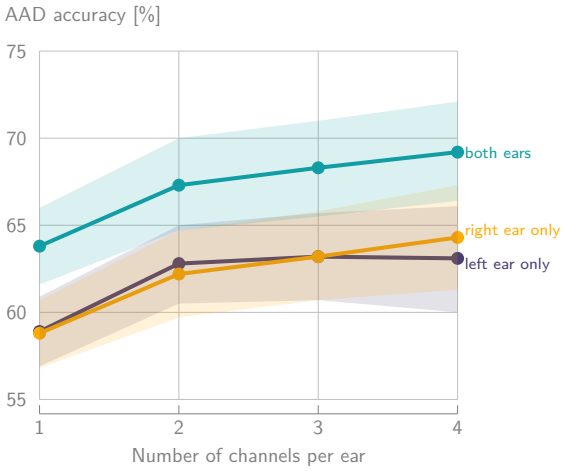


Fig. 3: The average sAAD accuracy (mean ± standard error across the 16 participants and all channel combinations) for 60 s decision windows as a function of the number of channels per ear shows a consistent 5 percentage point difference in accuracy between combining both ears and single-ear configurations.

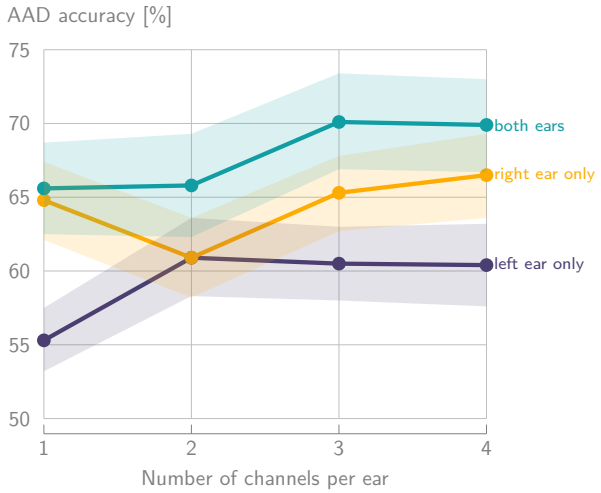
B. Channel selection using bipolar configurations

Each sensor node provides five electrodes: four around-ear electrodes (E1-4) and one reference electrode (REF) (see Figure 1a). In the previous analyses, four monopolar channels per ear were used by referencing each around-ear electrode to the local reference electrode. However, alternative bipolar configurations may yield more informative dipoles. We therefore perform a channel selection analysis considering all possible bipolar channels.

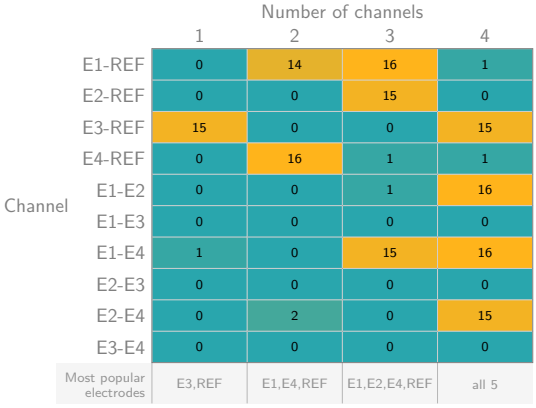
Per ear, the five electrodes allow for ten possible bipolar channels. For a given number of channels per ear (ranging from one to four), channel selection is performed using a leave-one-participant-out procedure combined with participant-specific decoding as before. Specifically, for each left-out participant, optimal channel combinations are selected based on the average performance across the remaining participants using leave-one-trial-out CV. These selected channels are then evaluated on the left-out participant, again using participant-specific leave-one-trial-out CV. When combining both ears, channel selection is constrained to use identical channel configurations for the left and right ears. Additionally, channel combinations are restricted to use at most one additional electrode beyond the number of channels, reflecting practical hardware constraints.

Figure 4a shows the resulting average sAAD accuracies for 60 s decision windows. Bipolar channel selection mostly yields (slightly) higher accuracies compared to the original monopolar configurations. Figure 4b summarizes the frequency with which specific bipolar channels are selected across participants when combining both ears.

Based on the most frequently selected channels in Figure 4b,



(a)



(b)

Fig. 4: **(a)** The average sAAD accuracy (mean ± standard error across the 16 participants) on 60 s decision windows using a bipolar channel selection as a function of the number of channels per ear shows that three channels using four electrodes per ear are sufficient to reach around 70%. **(b)** The heatmap of selected bipolar channels across participants for the combined-ear configuration indicates the most popular channels and required electrodes to be included in a practical WESN configuration.

a uniform bipolar configuration using three channels per ear and four electrodes (channels E1–REF, E2–REF, and E1–E4) achieves an average accuracy of 70.63% prior to HMM post-processing.

C. HMM post-processing results

Finally, we evaluate the impact of HMM-based post-processing, which exploits the temporal structure of attentional processes as explained in Section III-A.2. Attention switches were simulated as described in Section III-C. Figure 5 shows per-participant steady-state accuracies and average switch detection times for the combined-ear configuration, clearly showing how better-performing participants yield both improved steady-state accuracy and lower average switch detection times.

HMM post-processing increases the average sAAD accuracy using all four channels from both ears from 69.24% (60 s decision windows) to 77.17%, while achieving a mean switch detection time of 32.79 s. For single-ear configurations,

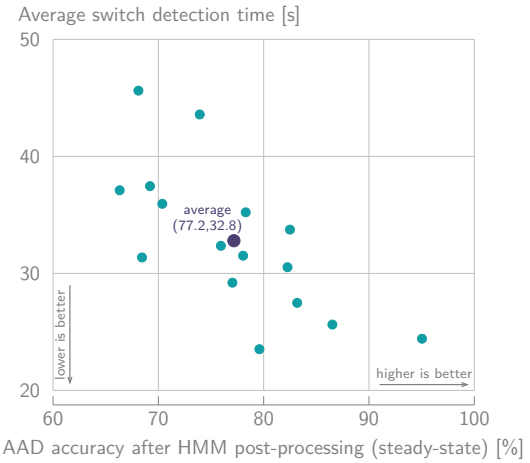


Fig. 5: The per-participant steady-state sAAD accuracies and average simulated switch detection times after HMM post-processing for the combined-ear configuration show a clear improvement compared to the raw sAAD accuracies shown in Figure 2b.

HMM post-processing similarly improves performance, yielding average accuracies of 74.10% (left ear, 35.41 s switch detection time) and 75.03% (right ear, 32.19 s switch detection time). The trade-off between steady-state accuracy and switch detection time can be further tuned by adjusting the HMM transition probabilities, yet this is beyond the scope of this paper. Lastly, the optimal bipolar configuration from Section IV-B, using three channels per ear and four electrodes, reaches an average accuracy of 77.08% after HMM post-processing, with an average switch detection time of 33.40 s.

V. Discussion

Based on the results presented in Section IV, three main discussion points can be identified and are elaborated below.

A. Selective auditory attention can be decoded using synchronized and galvanically separated EEG sensor nodes around the ear

First, the successful and significant sAAD performance obtained when combining two wireless EEG sensor nodes positioned around both ears (see, e.g., Figure 2a) demonstrates the practical viability of the developed WESN platform [19], including its synchronization protocol (Section II-A.3). These results confirm that synchronized, battery-powered, and galvanically separated miniature EEG sensor nodes can be reliably employed in a practical neuro-steered hearing device setting.

Second, the decoding performance obtained using the miniature around-ear WESN setup is highly comparable to the performance reported in previous around-ear-based studies, which typically relied on a single physically connected EEG grid with a shared reference electrode and achieved approximately 70% decoding accuracy on 60 s decision windows [18], [23], [24]. In the present work, galvanic isolation between the two nodes prevents the measurement of long-distance cross-ear potential differences. Nevertheless, this constraint does not appear to degrade sAAD performance. This observation is

in line with earlier findings by Geirnaert et al. [18], where galvanic isolation was emulated in around-ear EEG and no performance loss was observed. Taken together, these results suggest that little additional neural information for effective sAAD can be captured from cross-ear potential differences, but rather that most information is decoded from same-ear potentials.

Third, the HMM-based post-processing results reported in Section IV-C provide a realistic state-of-the-art performance benchmark for sAAD using a practical WESN. A steady-state decoding accuracy of 77.17% was achieved, with an average simulated switch detection time of $32.79\,\mathrm{s}$. While acceptable performance thresholds for real-world hearing device users remain unclear, these values are likely insufficient for a fully satisfactory user experience aimed at substantially improving speech intelligibility. Importantly, when HMM output probabilities are directly mapped to per-speaker gain control, the reported switch detection time corresponds to the point in time at which the relative playback levels of both speakers become equal following an attention switch. However, the perceptual gain adaptation towards the newly attended speaker begins much earlier, as the HMM relative output probabilities evolve smoothly over time [9]. As a result, the listener may experience a gradual enhancement of the attended speech well before the formal switch detection point. Moreover, while the obtained performance metrics may be insufficient for highly dynamic listening scenarios, many everyday listening scenarios are characterized by more slowly varying attentional dynamics [41], also enabled by, for example, grouping speech streams in conversations [14]. Within this context, the presented results provide a realistic baseline for what can currently be achieved if implemented in neuro-steered hearing devices, against which future methodological and hardware improvements can be evaluated.

B. Combining sensor nodes around both ears outperforms single-ear configurations primarily through increased robustness

The results consistently demonstrate a performance advantage when combining EEG sensor nodes at both ears compared to using single-ear configurations (Figures 2 and 3), with statistically significant differences observed at the group level and more participants exhibiting significant accuracies at the individual level. Beyond confirming the successful synchronization of the two wireless nodes, these findings indicate that a binaural EEG configuration is beneficial for achieving higher and more reliable sAAD performance.

Importantly, further analysis suggests that this benefit is driven primarily by increased robustness rather than by consistently exploitable spatially complementary neural information across ears. Specifically, when considering the post-hoc best-performing single-ear configuration for each participant across all trials, an average accuracy of 68.65% is obtained on $60\,\mathrm{s}$ decision windows, which is not significantly different from the 69.24% achieved by the combined-ear configuration (paired Wilcoxon signed-rank test; $p = 0.8911$). Moreover, for approximately half of the participants, the left ear yielded the best single-ear performance, while for the other half, the right ear performed best (Figure 2b). This inter-individual variability indicates that combining both ears mainly provides redundancy that mitigates ear-specific performance differences, leading to a more robust decoding performance across users. While complementary information across ears cannot be entirely ruled out, the present results suggest that robustness is the dominant factor underlying the observed performance gains. An important question for future research is whether, within individual participants, the better-performing ear remains consistent across recording sessions or instead reflects coincidental, session-specific factors, such as the specific experimental setup and conditions, electrode placement, or electrode impedances. Nonetheless, in practice, the best-performing single-ear configuration is in any case not known beforehand, warranting the use of the more robust both-ear configuration.

C. Three channels using four electrodes per ear are sufficient for robust sAAD

The channel selection analysis presented in Section IV-B demonstrates that comparable sAAD performance can be achieved using three bipolar channels derived from four electrodes per ear, rather than the original four channels using five electrodes per ear. Using a fixed bipolar configuration consisting of channels E1–REF, E2–REF, and E1–E4 (across both ears), and thus relying on electrodes E1, E2, E4, and REF only (see Figure 1a), an average decoding accuracy of 70.63% before and 77.08% after HMM post-processing (with an average switch detection time of $33.40\,\mathrm{s}$) was obtained.

These results indicate that a reduced and participant-independent electrode layout can be employed without sacrificing decoding performance, which is highly relevant for the design of wearable, concealable, and user-friendly neuro-steered hearing devices. In contrast, further reductions in the number of electrodes result in noticeable performance degradation (Figure 4a). Overall, this analysis provides clear design guidelines for integrating around-ear EEG into a practical WESN-based neuro-steered hearing device, balancing decoding performance against wearability and system complexity. In future work, additional mini-EEGs that do not compromise wearability and unobtrusiveness, such as in EEG glasses, could be seamlessly added through this WESN platform to further improve decoding performance.

VI. Conclusion

In this work, we established a proof of concept of a wireless EEG sensor network (WESN) consisting of two synchronized and galvanically separated around-ear EEG sensor nodes for selective auditory attention decoding in neuro-steered hearing devices. Using a traditional correlation-based stimulus decoding algorithm, an average decoding accuracy of 69.24% was achieved on $60\,\mathrm{s}$ decision windows, which is in line with the performance reported in the literature for physically connected around-ear EEG setups with a shared reference.

We further demonstrated that combining sensor nodes at both ears consistently outperforms single-ear configurations,

and hypothesized that this is primarily driven by providing redundancy that increases robustness across users rather than by exploiting additional cross-ear neural information. Using HMM-based post-processing, a realistic performance benchmark for sAAD with a practical WESN was established, yielding a steady-state accuracy of 77.17% with an average simulated switch detection time of 32.79 s. This benchmark provides a meaningful reference point for future practical implementations of neuro-steered hearing devices based on wireless EEG. Finally, through a bipolar channel selection analysis, we showed that the number of electrodes per ear can be reduced from five to four (from four to three channels) using a fixed and participant-independent configuration, without degrading decoding performance.

Overall, this work demonstrates that selective auditory attention decoding is feasible using a fully wireless, synchronized, and galvanically isolated EEG sensor network with unobtrusive around-ear EEG, representing a considerable step towards practical and wearable neuro-steered hearing devices.

Acknowledgments

ChatGPT (version 5.2; OpenAI) and Claude Sonnet (version 4.6; Anthropic) were used as a writing support for language editing and grammar checking. Claude Sonnet (version 4; Anthropic) was used as a coding assistant for software development and implementation support. All algorithms, analyses, and interpretations were designed, verified, and validated by the authors, who take full responsibility for the results. All authors have reviewed and approved the final content of the manuscript and take full responsibility for the paper's scientific integrity.

The authors thank Joachim Rivera Cea and Albrecht Willemans for providing the trigger implementation.

References

[1] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 29, pp. 11 854–11 859, 2012.

[2] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, pp. 233–236, 2012.

[3] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG," *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, 2014.

[4] 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, 2017.

[5] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigne, E. C. Lalor, B. T. Meyer, S. Miran, T. Francart, and A. Bertrand, "Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 89–102, 2021.

[6] N. D. T. Nguyen, H. Phan, S. Geirnaert, K. Mikkelsen, and P. Kidmose, "AADNet: An End-to-End Deep Learning Model for Auditory Attention Decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 2695–2706, 2025.

[7] G. Sridhar, S. Boselli, M. A. Skoglund, B. Bernhardsson, and E. Alickovic, "Improving auditory attention decoding in noisy environments for listeners with hearing impairment through contrastive learning," *Journal of Neural Engineering*, vol. 22, no. 3, p. 036041, 2025.

[8] 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.

[9] N. Heintz, T. Francart, and A. Bertrand, "Post-processing of EEG-based Auditory Attention Decoding Decisions via Hidden Markov Models," *arXiv preprint*, no. arXiv:2506.24024, 2025.

[10] S. Geirnaert, T. Francart, and A. Bertrand, "Unsupervised Self-Adaptive Auditory Attention Decoding," *IEEE Journal on Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3955–3966, 2021.

[11] ——, "Time-adaptive Unsupervised Auditory Attention Decoding Using EEG-based Stimulus Reconstruction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3767–3778, 2022.

[12] E. Ceolini, J. Hjortkjær, D. D. Wong, J. A. O'Sullivan, V. S. Raghavan, J. Herrero, A. D. Mehta, S. C. Liu, and N. Mesgarani, "Brain-informed speech separation (BISS) for enhancement of target speaker in multitalker speech perception," *NeuroImage*, vol. 223, no. 117282, 2020.

[13] V. Choudhari, C. Han, S. Bickel, A. D. Mehta, C. Schevon, G. M. McKhann, and N. Mesgarani, "Brain-Controlled Augmented Hearing for Spatially Moving Conversations in Multi-Talker Environments," *Advanced Science*, vol. 11, no. 41, p. 2401379, 2024.

[14] I. Van de Ryck, N. Heintz, I. Rotaru, S. Geirnaert, A. Bertrand, and T. Francart, "EEG-based Decoding of Auditory Attention to Conversations with Turn-taking Speakers," *Hearing Research*, vol. 471, no. 109539, 2025.

[15] S. A. Fuglsang, J. Märcher-Rørsted, T. Dau, and J. Hjortkjær, "Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention," *Journal of Neuroscience*, vol. 40, no. 12, pp. 2562–2572, 2020.

[16] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, a Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[17] 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.

[18] S. Geirnaert, S. L. Kappel, and P. Kidmose, "A direct comparison of simultaneously recorded scalp, around-ear and in-ear EEG for neural selective auditory attention decoding to speech," *Scientific Reports*, vol. 15, no. 41655, 2025.

[19] R. Ding, C. Hovine, P. Callemeyn, M. Kraft, and A. Bertrand, "A Wireless, Scalable, and Modular EEG Sensor Network Platform for Unobtrusive Brain Recordings," *IEEE Sensors Journal*, vol. 25, no. 12, pp. 22 580–22 590, 2025.

[20] 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. 46007, 2015.

[21] 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, 2020.

[22] A. Mundanad Narayanan, R. Zink, and A. Bertrand, "EEG miniaturization limits for stimulus decoding with EEG sensor networks," *Journal of Neural Engineering*, vol. 18, no. 5, p. 056042, 2021.

[23] 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, no. 349, 2016.

[24] B. Holtze, M. Rosenkranz, M. Jaeger, S. Debener, and B. Mirkovic, "Ear-EEG Measures of Auditory Attention to Continuous Speech," *Frontiers in Neuroscience*, vol. 16, 2022.

[25] W. Nogueira, H. Dolhopiatenko, I. Schierholz, A. Büchner, B. Mirkovic, M. G. Bleichner, and S. Debener, "Decoding Selective Attention in Normal Hearing Listeners and Bilateral Cochlear Implant Users With Concealed Ear EEG," *Frontiers in Neuroscience*, vol. 13, 2019.

[26] H. Zhu, Y. Yan, X. Xu, Z. Ge, P. Tian, X. Wu, and J. Chen, "Using Ear-EEG to Decode Auditory Attention in Multiple-speaker Environment," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[27] L. Fiedler, M. Wöstmann, C. Graversen, A. Brandmeyer, T. Lunner, and J. Obleser, "Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech," *Journal of Neural Engineering*, vol. 14, no. 3, p. 036020, 2017.

[28] R. Ding, S. Geirnaert, C. Hovine, and A. Bertrand, "Synchronized eeg with two galvanically-separated miniature wireless behind-the-ear

eeg sensors," in *7th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2025, pp. 1–4.

[29] S. Geirnaert, T. Francart, and A. Bertrand, "Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1557–1568, 2021.

[30] S. Vandecappelle, L. Deckers, N. Das, A. H. Ansari, A. Bertrand, and T. Francart, "EEG-based detection of the locus of auditory attention with convolutional neural networks," *eLife*, vol. 10, no. e56481, 2021.

[31] C. Puffay, B. Accou, L. Bollens, M. J. Monesi, J. Vanthornhout, H. Van hamme, and T. Francart, "Relating EEG to continuous speech using deep neural networks: a review," *Journal of Neural Engineering*, vol. 20, no. 4, p. 041003, aug 2023.

[32] I. Rotaru, S. Geirnaert, N. Heintz, I. Van de Ryck, A. Bertrand, and T. Francart, "What are we really decoding? Unveiling biases in EEG-based decoding of the spatial focus of auditory attention," *Journal of Neural Engineering*, vol. 21, no. 1, p. 016017, 2024.

[33] G. Ivucic, S. Pahuja, F. Putze, S. Cai, H. Li, and T. Schultz, "The Impact of Cross-Validation Schemes for EEG-Based Auditory Attention Detection with Deep Neural Networks," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.

[34] Y. Yan, X. Xu, H. Zhu, S. Li, B. Wang, X. Wu, and J. Chen, "Overestimated performance of auditory attention decoding caused by experimental design in EEG recordings," in *Interspeech 2025*, 2025, pp. 1053–1057.

[35] M. G. Bleichner and S. Debener, "Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG," *Frontiers in Human Neuroscience*, vol. 11, 2017.

[36] R. Ding, S. Geirnaert, and A. Bertrand, "Wireless around-ear SAAD dataset [Data set]," Zenodo [WILL BE PUSBLISHED UPON ACCEPTANCE], 2026.

[37] 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 Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 5, pp. 402–412, 2017.

[38] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[39] M. A. Lopez-Gordo, S. Geirnaert, and A. Bertrand, "Unsupervised Accuracy Estimation for Brain-Computer Interfaces based on Selective Auditory Attention Decoding," *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 8, pp. 2388–2399, 2025.

[40] S. Geirnaert, I. Rotaru, T. Francart, and A. Bertrand, "Linear stimulus reconstruction works on the KU Leuven audiovisual, gaze-controlled auditory attention decoding dataset," *arRxiv*, no. 2412.01401, 2024.

[41] J. Hjortkjær, D. D. Wong, A. Catania, J. Märcher-Rørsted, E. Ceolini, S. A. Fuglsang, I. Kiselev, G. Di Liberto, S.-C. Liu, T. Dau *et al.*, "Real-time control of a hearing instrument with EEG-based attention decoding," *Journal of Neural Engineering*, vol. 22, no. 1, p. 016027, 2025.