PAPER

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

Simon Geirnaert[1,2,3], Ruochen Ding[1,3] and Alexander Bertrand[1,3,*]

[1]Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, 3001 Leuven, Belgium
[2]Department of Neurosciences, Research Group ExpORL, KU Leuven, 3000 Leuven, Belgium
[3]Leuven.AI - KU Leuven institute for AI, 3000 Leuven, Belgium
*Author to whom any correspondence should be addressed.

E-mail: alexander.bertrand@kuleuven.be

Keywords: selective auditory attention, wireless EEG sensor network

Abstract

Objective: Selective auditory attention decoding (sAAD) enables neuro-steered hearing devices through identification of the attended speaker in a multi-speaker environment, using neural activity recorded with electroencephalography (EEG). Despite substantial progress in decoding algorithms, 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 miniaturized, galvanically isolated EEG sensor nodes. *Approach:* We investigate sAAD using a wireless EEG sensor network (WESN) setup consisting of two synchronized, compact around-ear EEG sensor nodes worn bilaterally. The nodes are fully galvanically isolated (i.e., there is no wire between them), each providing 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. Using a newly recorded dataset acquired with this setup, we evaluate correlation-based stimulus decoding and assess the effects of miniaturization and galvanic isolation between WESN nodes. *Main results:* Using a standard correlation-based stimulus decoding approach, an average sAAD accuracy of 69.24% is achieved on 60 s decision windows, comparable to wired around-ear EEG systems that measure long-distance scalp potentials. Hidden Markov model–based post-processing further improves performance to a steady-state accuracy of 77.17% with an average attention switch detection time of 32.79 s. Combining sensor nodes at both ears outperforms single-ear configurations, primarily through increased robustness via redundancy 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. *Significance:* These results demonstrate the practical feasibility of sAAD using a fully wireless, galvanically isolated around-ear WESN. By establishing a realistic performance benchmark under practical hardware constrains, this work represents a considerate step towards deployable neuro-steered hearing devices.

1 Introduction

Selective auditory attention to one out of multiple competing speech streams can be decoded from neural responses, for example recorded using electroencephalography (EEG) (Ding and Simon, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2014). 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 (Van Eyndhoven et al., 2017; Geirnaert et al., 2021c). 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 using deep neural networks (Nguyen

et al., 2025; Sridhar et al., 2025), by introducing advanced post-processing strategies for correlation-based neural tracking methods (Miran et al., 2018; Heintz et al., 2025), or by developing unsupervised variants (Geirnaert et al., 2021b, 2022). In parallel, the integration of sAAD algorithms with speech separation and enhancement pipelines has been investigated (e.g., Van Eyndhoven et al., 2017; Ceolini et al., 2020), as well as validation in more realistic acoustic environments (e.g., Choudhari et al., 2024; Van de Ryck et al., 2025) and in hearing-impaired populations (e.g., Fuglsang et al., 2020). 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 the latter 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 (Nicolas-Alonso and Gomez-Gil, 2012). 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 (Bertrand, 2015). 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 (Geirnaert et al., 2025). 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 (Bertrand, 2015; Ding et al., 2025b).

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 (Mirkovic et al., 2015; Mundanad Narayanan and Bertrand, 2020). Mundanad Narayanan and Bertrand (2020) 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 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. (2021) 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 cm, with optimal sensor locations situated above the temporal lobe, near the auditory cortex, as expected for auditory decoding. Additionally, Mundanad Narayanan and Bertrand (2020) 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) (Mirkovic et al., 2016; Holtze et al., 2022; Nogueira et al., 2019; Zhu et al., 2025; Geirnaert et al., 2025) and customized in-ear EEG devices (Fiedler et al., 2017; Geirnaert et al., 2025). Using correlation-based stimulus decoding based on neural tracking, typical decoding accuracies of approximately 70% for wet around-ear EEG (Mirkovic et al., 2016; Holtze et al., 2022; Geirnaert et al., 2025) and 60% for dry in-ear EEG have been reported on 60 s decision windows (Geirnaert et al., 2025). Furthermore, Geirnaert et al. (2025) 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. (2025b), using two compact, synchronized wireless EEG sensor nodes with a printed circuit board (PCB)-size of 2×3 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. (2025a) 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 by four-channel 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, given that alternative direct classification approaches that decode, for instance, the spatial focus of attention (Geirnaert et al., 2021a; Vandecappelle et al., 2021), despite achieving higher accuracies at shorter decision windows, are also more susceptible to overfitting, shortcut learning, and overly optimistic performance estimates (Puffay et al., 2023; Rotaru et al., 2024; Ivucic et al., 2024; Yan et al., 2025). 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 (Heintz et al., 2025), thereby establishing a state-of-the-art benchmark for sAAD performance under realistic hardware constraints for neuro-steered hearing devices.

In Section 2, we describe the employed sensor nodes and the EEG-Linx platform, and the recorded dataset. The sAAD algorithm, preprocessing, and evaluation procedure are detailed in Section 3. The results are reported in Section 4 and discussed in Section 5.

2 Methods: Data Collection

This section describes the around-ear WESN platform employed for EEG data collection in Section 2.1 and the recorded dataset in Section 2.2.

2.1 Wireless EEG Sensor Nodes for the WESN

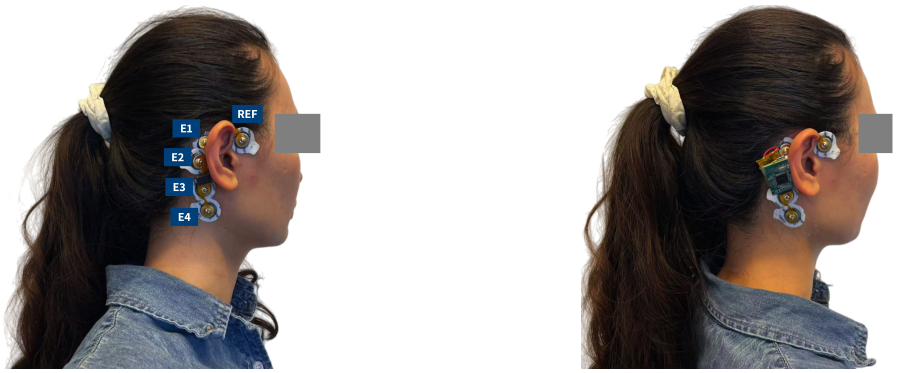
EEG was recorded using two battery-powered around-ear sensor nodes worn bilaterally (one per ear; see Figure 1), which streamed data wirelessly to a hub node that was connected to an acquisition PC via USB. Each node recorded four EEG channels in a (local) common-reference montage, with four around-ear 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 node (Ding et al., 2025b).

2.1.1 Wireless EEG sensor nodes The wearable EEG sensor node, originally introduced in (Ding et al., 2025b), is realized on a compact double-sided, six-layer PCB (approximately 2×3 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 Hz and streamed wirelessly to the hub node during the experiments. Each node was powered by a 3.7 V, 250 mAh Li-Po battery.

2.1.2 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 (Bleichner and Debener, 2017). 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 24 mm Arbo electrodes H124SG) with a diameter of 24 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



(a) (b)

Figure 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.

(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.

2.1.3 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 (Ding et al., 2025b). 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 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 (Ding et al., 2025b).

2.2 Dataset

This section describes a newly recorded sAAD dataset acquired using the wireless two-node around-ear EEG setup, as described in Section 2.1. The experimental protocol was approved by the KU Leuven Social and Societal Ethics Committee. The complete dataset is made publicly available (Ding et al., 2026).

2.2.1 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 (yet non-native except for one) English speakers and provided written informed consent prior to the experiment.

2.2.2 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 narrated by English-speaking male speakers (taken from `https://openslr.org`) and were presented simultaneously through two loudspeakers positioned at approximately $\pm 30°$ azimuth and at a distance of one meter 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 min each, with a maximum total duration of 50 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.

Due to various reasons (depleted battery, premature ending of the experiment on request of the participant, (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 min trials across the 16 participants were completed in full, resulting in an average recording duration of 42 min per participant (standard deviation: 6.7 min; minimum duration: 25 min). All trials, including incomplete ones, were retained and included in the analysis. All participants had a sufficient number of trials to be included in the analysis.

3 Methods: Data Analysis

Section 3.1 presents the correlation-based linear stimulus decoding algorithm for sAAD, complemented by a hidden Markov Model (HMM) post-processor (Heintz et al., 2025). Preprocessing steps for the EEG and speech signals are detailed in Section 3.2, and the evaluation procedure is described in Section 3.3.

3.1 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 3.1.1).

2. An HMM-based post-processor that exploits temporal regularities in attentional processes by integrating neural tracking correlations across successive decision windows (Section 3.1.2).

3.1.1 Correlation-based stimulus decoding We employ a linear stimulus decoding approach for sAAD (O'Sullivan et al., 2014; Biesmans et al., 2017; Geirnaert et al., 2021c), which is widely regarded as one of the most robust methods and has been successfully applied in combination with around-ear EEG (Mirkovic et al., 2016; Holtze et al., 2022; Geirnaert et al., 2025). The method relies on neural tracking, whereby neural responses synchronize more strongly over time with the attended speech signal than with the unattended one (Ding and Simon, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2014). 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 speech representation, the popular acoustic envelope is used (O'Sullivan et al., 2014; Biesmans et al., 2017; Geirnaert et al., 2021c). Below, a brief summary of the decoding approach is provided; extensive explanations are available in literature (O'Sullivan et al., 2014; Biesmans et al., 2017; Geirnaert et al., 2021c, 2022).

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

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

Here, $\mathbf{X}$ is a block Hankel matrix containing lagged EEG samples per channel. The time lags are chosen acausally, spanning the interval $[0, 400]$ms to capture neural responses following the stimulus (Geirnaert et al., 2022). This least-squares formulation is equivalent to maximizing the Pearson correlation between the reconstructed and attended speech envelopes (Biesmans et al., 2017). 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 (Ledoit and Wolf, 2004), which has been shown to perform well for this decoder, including in the context of around-ear EEG (Geirnaert et al., 2025). 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 (Geirnaert et al., 2021b, 2022), 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 $\hat{\mathbf{y}}^{(\text{test})} = \mathbf{X}^{(\text{test})}\hat{\mathbf{d}}$:

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

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 3.1.2).

3.1.2 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. (2025), 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 be able to fully evaluate the HMM post-processing performance (see Section 3.3).

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 $\mathbf{r}(t) = [r_1(t), r_2(t)]$ 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 here explicitly choose the causal version of the post-processor, to realistically simulate the online and causal processing of attentional scores. Emission probabilities $\mathbb{P}\left(\mathbf{r}(t) \mid \hat{S}(t) = S_j\right)$ are modeled as Gaussian distributions conditioned on the attentional state (Lopez-Gordo et al., 2025; Heintz et al., 2025).

The same hyperparameter settings are chosen as in Heintz et al. (2025): an attention switch probability of 0.001 and a decision window length of 1 s to generate the observed correlation scores. 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 average difference of attended and unattended correlations of 0.1 before normalization (Lopez-Gordo et al., 2025). For each decision window, the estimated attentional state corresponds to the state with the highest posterior probability.

3.2 Preprocessing

3.2.1 EEG preprocessing EEG recordings from both sensor nodes and the speech stimuli are processed on a per-trial basis. EEG preprocessing is deliberately kept minimal and can be performed fully independently per node. EEG signals are bandpass-filtered between 1–9 Hz using a zero-phase fourth-order Butterworth filter and resampled to 20 Hz. This frequency band is well established for correlation-based stimulus decoding (O'Sullivan et al., 2014; Geirnaert et al., 2021c, 2025). Finally, per-trial normalization is applied by scaling the Frobenius norm of the EEG channels to one. In case of channel selection (see Section 4.2), this normalization is performed after selection to avoid implicit overfitting.

3.2.2 Speech preprocessing Speech envelopes are extracted following the procedure described by Biesmans et al. (2017). Speech signals are first passed through a gammatone filterbank with 19 bands spanning 50–5000 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 Hz using the same Butterworth filter as for EEG preprocessing, and resampled to 20 Hz. Finally, envelopes are z-scored on a per-trial basis.

3.3 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. This strict trial-based CV scheme is necessary to avoid shortcut learning and overfitting in sAAD (Puffay et al., 2023; Rotaru et al., 2024; Ivucic et al., 2024; Yan et al., 2025), although correlation-based stimulus decoding paradigms are less susceptible to such shortcut learning (Geirnaert et al., 2024). Decision window lengths are varied across $T_{\text{test}} \in \{60, 30, 20, 10, 5, 2, 1\}$s. Raw

correlation-based accuracies are primarily analyzed on 60 s windows, while 1 s windows are used for HMM post-processing.

For HMM evaluation, two metrics are considered: steady-state accuracy and switch detection time. To enable their computation, attention switches are simulated on the correlation scores that are provided as an input to the HMM. First, all 1 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 min, with a minimum inter-switch interval of 2 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. 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. This switch simulation procedure is performed 100 times to reliably estimate the average steady-state accuracy and switch detection time.

4 Results

This section presents the experimental results obtained with the proposed wireless EEG sensor network. We first compare performance when combining sensor nodes from both ears with single-ear configurations (Section 4.1). Next, we analyze bipolar channel selection strategies to identify practically optimal electrode configurations within the WESN setup (Section 4.2). Finally, we evaluate the impact of HMM post-processing (Section 4.3).

4.1 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 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. 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 $p = 0.0122$) and the right ear only (uncorrected $p = 0.0386$, corrected $p = 0.0386$).

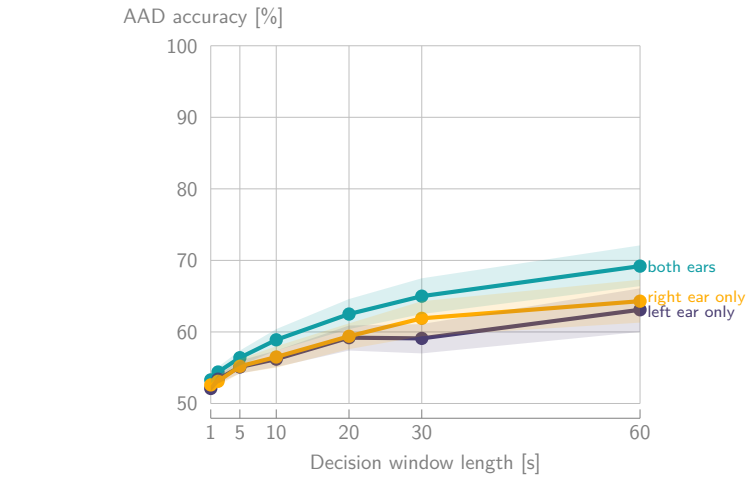
To investigate how combining both ears compares to single-ear configurations across varying number of channels, Figure 3 illustrates the effect of limiting the number of channels per ear on the sAAD accuracy based on the raw correlations in Eq. (2). Across all channel combinations (taken from the original 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 both ears, selected channels are always matched across both ears.

4.2 Channel selection using bipolar configurations

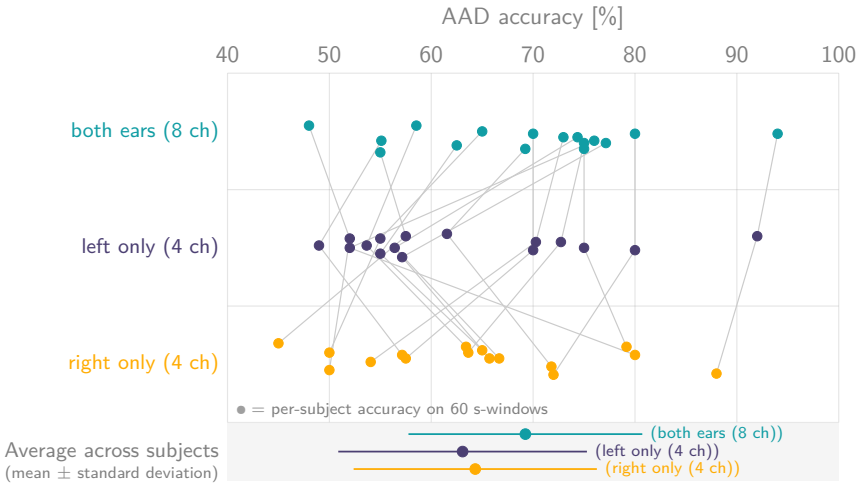
Each sensor node provides five electrodes: four around-ear electrodes (E1-4) and one reference electrode (REF) (see Figure 1a). In the previous analyses, four monopolar channels per ear were used by referencing each around-ear electrode to the local reference electrode. However, alternative bipolar configurations 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.



(a)



(b)

Figure 2: **(a)** The average sAAD performance curves (mean ± standard error across the 16 participants) 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.

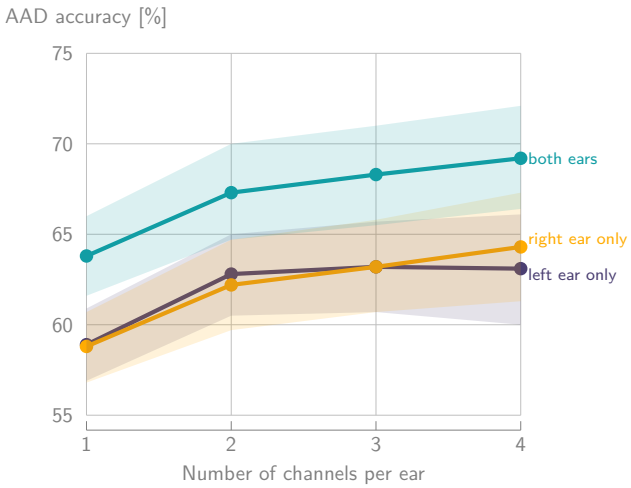
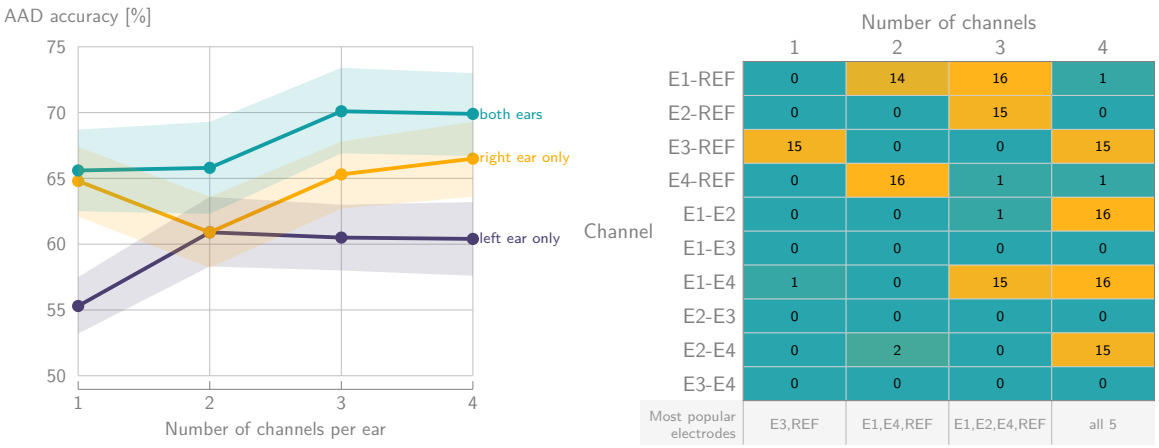


Figure 3: The average sAAD accuracy (mean ± standard error across the 16 participants and all channel combinations) for 60 s decision windows as a function of the number of channels per ear shows a consistent 5 percentage point difference in accuracy between combining both ears over single-ear configurations.



(a) (b)

Figure 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 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.

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

4.3 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 3.1.2. Attention switches were simulated as described in Section 3.3. Figure 5 shows per-participant steady-state accuracies and average switch detection times for the configuration of combining both ears.

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, 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 using three channels per ear and four electrodes of Section 4.2 reaches after HMM post-processing 77.08% accuracy with an average switch detection time of 33.40 s.

5 Discussion

Based on the results presented in Section 4, three main discussion points can be identified and are elaborated below.

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

First, the successful 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 (Ding et al., 2025b), including its synchronization protocol (Section 2.1.3). 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 achieving approximately 70% decoding accuracy on 60 s decision windows, which typically relied on a single physically connected EEG grid with a shared reference electrode (Mirkovic et al., 2016; Holtze et al., 2022; Geirnaert et al., 2025). In the present work, galvanic isolation between the two nodes prevents the measurement of longer-distance cross-ear potential differences. Nevertheless, this constraint does not

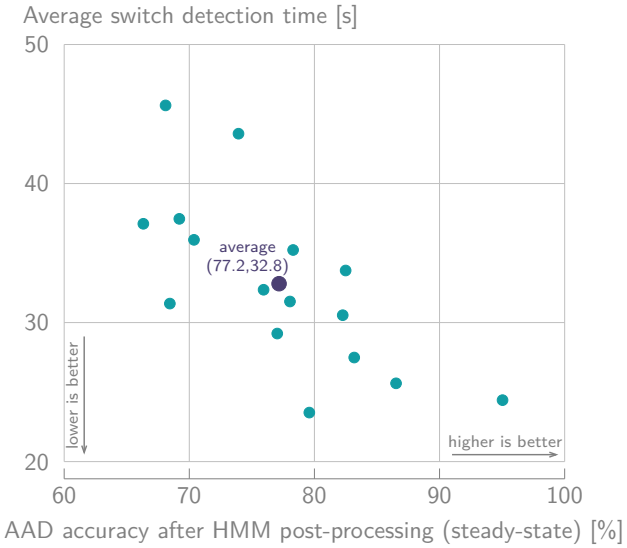


Figure 5: The per-participant steady-state sAAD accuracies and average switch detection times after HMM post-processing for the combined-ear configuration show a clear improvement compared to the raw sAAD accuracies shown in Figure 2b.

appear to degrade sAAD performance. This observation is in line with earlier findings by Geirnaert et al. (2025), 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 that most information is rather decoded from same-ear potentials.

Third, the HMM-based post-processing results reported in Section 4.3 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 switch detection time of 32.79 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 (Heintz et al., 2025). 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 (Hjortkjær et al., 2025), also enabled by, for example, grouping speech streams in conversations (Van de Ryck et al., 2025). 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.

5.2 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. 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 best-performing single-ear configuration for each participant across all trials, an average accuracy of 68.65% is obtained on 60 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.

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

The channel selection analysis presented in Section 4.2 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 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.

6 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 s decision windows, which is in line with performances reported in the literature using 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, primarily 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 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) was 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. Furthermore, the authors declare that no competing interests exist.

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

Funding

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.

Author contributions

Simon Geirnaert: Conceptualization, Data curation, Formal analysis, Funding Acquisition, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. *Ruochen Ding*: Conceptualization, Data curation, Investigation, Resources, Writing - original draft, Writing - review & editing. *Alexander Bertrand*: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing.

Data availability

The full dataset will be made publicly available upon acceptance.

References

Bertrand, A. (2015). Distributed Signal Processing for Wireless EEG Sensor Networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(6):923–935.

Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2017). Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(5):402–412.

Bleichner, M. G. and Debener, S. (2017). Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. *Frontiers in Human Neuroscience*, 11.

Ceolini, E., Hjortkjær, J., Wong, D. D., O'Sullivan, J. A., Raghavan, V. S., Herrero, J., Mehta, A. D., Liu, S. C., and Mesgarani, N. (2020). Brain-informed speech separation (BISS) for enhancement of target speaker in multitalker speech perception. *NeuroImage*, 223(117282).

Choudhari, V., Han, C., Bickel, S., Mehta, A. D., Schevon, C., McKhann, G. M., and Mesgarani, N. (2024). Brain-Controlled Augmented Hearing for Spatially Moving Conversations in Multi-Talker Environments. *Advanced Science*, 11(41):2401379.

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

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

Ding, R., Geirnaert, S., Hovine, C., and Bertrand, A. (2025a). 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)*, pages 1–4.

Ding, R., Hovine, C., Callemeyn, P., Kraft, M., and Bertrand, A. (2025b). A Wireless, Scalable, and Modular EEG Sensor Network Platform for Unobtrusive Brain Recordings. *IEEE Sensors Journal*, 25(12):22580–22590.

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., and Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, 14(3):036020.

Fuglsang, S. A., Märcher-Rørsted, J., Dau, T., and Hjortkjær, J. (2020). Effects of Sensorineural Hearing Loss on Cortical Synchronization to Competing Speech during Selective Attention. *Journal of Neuroscience*, 40(12):2562–2572.

Geirnaert, S., Francart, T., and Bertrand, A. (2021a). Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns. *IEEE Transactions on Biomedical Engineering*, 68(5):1557–1568.

Geirnaert, S., Francart, T., and Bertrand, A. (2021b). Unsupervised Self-Adaptive Auditory Attention Decoding. *IEEE Journal on Biomedical and Health Informatics*, 25(10):3955–3966.

Geirnaert, S., Francart, T., and Bertrand, A. (2022). Time-adaptive Unsupervised Auditory Attention Decoding Using EEG-based Stimulus Reconstruction. *IEEE Journal of Biomedical and Health Informatics*, 26(8):3767–3778.

Geirnaert, S., Kappel, S. L., and Kidmose, P. (2025). A direct comparison of simultaneously recorded scalp, around-ear and in-ear EEG for neural selective auditory attention decoding to speech. *Scientific Reports*, 15(41655).

Geirnaert, S., Rotaru, I., Francart, T., and Bertrand, A. (2024). Linear stimulus reconstruction works on the KU Leuven audiovisual, gaze-controlled auditory attention decoding dataset. *arRxiv*, (2412.01401).

Geirnaert, S., Vandecappelle, S., Alickovic, E., de Cheveigne, A., Lalor, E. C., Meyer, B. T., Miran, S., Francart, T., and Bertrand, A. (2021c). Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. *IEEE Signal Processing Magazine*, 38(4):89–102.

Heintz, N., Francart, T., and Bertrand, A. (2025). Post-processing of EEG-based Auditory Attention Decoding Decisions via Hidden Markov Models. *arXiv preprint*, (arXiv:2506.24024).

Hjortkjær, J., Wong, D. D., Catania, A., Märcher-Rørsted, J., Ceolini, E., Fuglsang, S. A., Kiselev, I., Di Liberto, G., Liu, S.-C., Dau, T., et al. (2025). Real-time control of a hearing instrument with EEG-based attention decoding. *Journal of Neural Engineering*, 22(1):016027.

Holtze, B., Rosenkranz, M., Jaeger, M., Debener, S., and Mirkovic, B. (2022). Ear-EEG Measures of Auditory Attention to Continuous Speech. *Frontiers in Neuroscience*, 16.

Ivucic, G., Pahuja, S., Putze, F., Cai, S., Li, H., and Schultz, T. (2024). 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)*, pages 1–4.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Lopez-Gordo, M. A., Geirnaert, S., and Bertrand, A. (2025). Unsupervised Accuracy Estimation for Brain-Computer Interfaces based on Selective Auditory Attention Decoding. *IEEE Transactions on Biomedical Engineering*, 72(8):2388–2399.

Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485:233–236.

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., and Babadi, B. (2018). Real-time tracking of selective auditory attention from M/EEG: A Bayesian filtering approach. *Frontiers in Neuroscience*, 12:262.

Mirkovic, B., Bleichner, M. G., De Vos, M., and Debener, S. (2016). Target Speaker Detection with Concealed EEG Around the Ear. *Frontiers in Neuroscience*, 10(349).

Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of Neural Engineering*, 12(4):46007.

Mundanad Narayanan, A. and Bertrand, A. (2020). Analysis of Miniaturization Effects and Channel Selection Strategies for EEG Sensor Networks with Application to Auditory Attention Detection. *IEEE Transactions on Biomedical Engineering*, 67(1):234–244.

Mundanad Narayanan, A., Zink, R., and Bertrand, A. (2021). EEG miniaturization limits for stimulus decoding with EEG sensor networks. *Journal of Neural Engineering*, 18(5):056042.

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

Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors*, 12(2):1211–1279.

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

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2014). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, 25(7):1697–1706.

Puffay, C., Accou, B., Bollens, L., Monesi, M. J., Vanthornhout, J., Van hamme, H., and Francart, T. (2023). Relating EEG to continuous speech using deep neural networks: a review. *Journal of Neural Engineering*, 20(4):041003.

Rotaru, I., Geirnaert, S., Heintz, N., Van de Ryck, I., Bertrand, A., and Francart, T. (2024). What are we really decoding? Unveiling biases in EEG-based decoding of the spatial focus of auditory attention. *Journal of Neural Engineering*, 21(1):016017.

Sridhar, G., Boselli, S., Skoglund, M. A., Bernhardsson, B., and Alickovic, E. (2025). Improving auditory attention decoding in noisy environments for listeners with hearing impairment through contrastive learning. *Journal of Neural Engineering*, 22(3):036041.

Van de Ryck, I., Heintz, N., Rotaru, I., Geirnaert, S., Bertrand, A., and Francart, T. (2025). EEG-based Decoding of Auditory Attention to Conversations with Turn-taking Speakers. *Hearing Research*, 471(109539).

Van Eyndhoven, S., Francart, T., and Bertrand, A. (2017). EEG-Informed Attended Speaker Extraction from Recorded Speech Mixtures with Application in Neuro-Steered Hearing Prostheses. *IEEE Transactions on Biomedical Engineering*, 64(5):1045–1056.

Vandecappelle, S., Deckers, L., Das, N., Ansari, A. H., Bertrand, A., and Francart, T. (2021). EEG-based detection of the locus of auditory attention with convolutional neural networks. *eLife*, 10(e56481).

Yan, Y., Xu, X., Zhu, H., Li, S., Wang, B., Wu, X., and Chen, J. (2025). Overestimated performance of auditory attention decoding caused by experimental design in EEG recordings. In *Interspeech 2025*, pages 1053–1057.

Zhu, H., Yan, Y., Xu, X., Ge, Z., Tian, P., Wu, X., and Chen, J. (2025). 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)*, pages 1–5.