PROBABILISTIC GAIN CONTROL IN A MULTI-SPEAKER SETTING USING EEG-BASED AUDITORY ATTENTION DECODING

Nicolas Heintz⋆† *Simon Geirnaert*⋆† *Iris Van de Ryck*† *Tom Francart*† *Alexander Bertrand*⋆

⋆KU Leuven, Department of Electrical Engineering (ESAT),
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Belgium
†KU Leuven, Department of Neurosciences, ExpORL, Belgium

ABSTRACT

Current hearing aids have difficulties focusing on the correct speaker in complex environments where multiple people are talking simultaneously. This is because hearing aids do not know to which speaker the user aims to listen. A promising solution is to use so-called auditory attention decoding (AAD) algorithms, which infer the attended speaker based on brain activity recorded with, e.g., electroencephalography (EEG). AAD models decode on a window-to-window basis to which person a subject wishes to listen. However, only limited research has been done on how these AAD decisions can be converted into a gain control system that controls the volume of each competing speaker in a hearing aid. Existing gain control systems are either difficult to tune, unstable and/or not designed for use in situations with more than two competing speakers. We therefore propose a novel general purpose gain control system that can be easily used on any AAD model and in scenarios with an arbitrary number of speakers. We demonstrate that the gain control system is stable, even for AAD algorithms with very low accuracies, and even in scenarios with more than 2 speakers.

***Index Terms*—** Auditory attention decoding, unsupervised learning, electroencephalography, neuro-steered hearing devices

1. INTRODUCTION

Despite recent advancements in beamforming and source separation techniques, hearing aids still struggle in so-called cocktail party environments where multiple speakers are talking simultaneously. A main culprit is that it is difficult to accurately detect to which speaker a person wishes to listen, and thus on which speaker the hearing aid should focus [1].

A promising solution is to directly decode from electroencephalography (EEG) signals who the attended speaker is. This is called (selective) Auditory Attention Decoding (AAD). Most AAD algorithms use the fact that the brain tracks the envelope of attended speech more than the envelope of unattended speech to decode who the attended speaker is [2–8]. Other AAD algorithms focus on, e.g. decoding the location of the attended speaker [9, 10], or use end-to-end decoding with neural networks [11].

In general, AAD algorithms cut the recorded EEG signal in small windows and predict per window who the attended speaker is. However, only a limited amount of research has focused on how these decisions should be incorporated into a gain control system that controls the volume (gain) of each speaker. A Bayesian state-space model was proposed in [12] and adopted in [5, 13, 14]. While this model achieves promising results, the vast number of unintuitive hyperparameters makes it difficult to use in different AAD models or listening contexts. The Markov chain proposed in [15] is, in contrast, designed to be plug-and-play and interpretable. It only contains a single hyperparameter that directly controls the balance between stability and speed. However, it is unable to take the decision certainty of AAD models into account, making it unstable when the AAD model is uncertain who the attended speaker is for a prolonged amount of time. Furthermore, it is, to our knowledge, either non-trivial [5, 12–14] or impossible [15] to directly use these aforementioned gain control systems in situations with more than two competing speakers.

We, therefore, propose a new gain control system that combines the advantages of the two previous models and also works in situations with an arbitrary number of speakers. Similar to [15], it is designed to be 'plug-and-play' such that it can be used in combination with any AAD model and it only requires a single hyperparameter that directly controls the stability and speed of the gain control system. Similar to [5, 12, 13], it is able to take the decision certainty of the AAD model into account, ensuring that the speaker gains remain stable when the AAD model is uncertain who the attended speaker is. The model is also kept as simple as possible, to ensure its easy usability. The gain control system is designed in Section 2 and evaluated in Section 3. We evaluate the gain control on data with both two and three competing speakers.

2. GAIN CONTROL SYSTEM

The gain control system regulates the volume of each competing speaker in a cocktail party. To achieve this, the gain control system changes the volume of each speaker based on the probability that this speaker is attended. However, most AAD algorithms do not directly estimate this probability, but simply give a score to each speaker, based on which binary decision can be made. Therefore, we explain in Section 2.1 how the attention probability can be estimated for each speaker. The actual gain control system is derived in Section 2.2.

This research is funded by Aspirant Grant 1S31522N (for N. Heintz) and 1S34821N (for I. Van de Ryck) from the Research Foundation - Flanders (FWO), junior postdoctoral fellowship fundamental research 1242524N from the FWO (for S. Geirnaert), the Research Foundation - Flanders (FWO) project No G0A4918N and G081722N, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 802895 and grant agreement No 637424), and the Flemish Government (AI Research Program). N. Heintz, S. Geirnaert and A. Bertrand are also affiliated with Leuven.AI - KU Leuven institute for AI, Belgium. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or granting authorities. Neither the European Union nor the granting authority can be held responsible for them.

2.1. Attention probability estimation

AAD algorithms decode to which of S speakers a subject is attending from that subject's EEG. In general, these algorithms cut the EEG in non-overlapping windows and then compute for each window i a set of scores $\mathbf{x}(i) \in \mathbb{R}^{S\times 1}$, which are related to the probability of each of the S speakers being attended or not.

We wish to compute the probability $P(s = a|\mathbf{x}(i))$ that speaker s is attended given the scores $\mathbf{x}(i)$. To do this, we assume that the scores are independently drawn from either the attended normal distribution $\mathcal{N}(\mu_a, \sigma_a)$ if the corresponding speaker is attended or the unattended normal distribution $\mathcal{N}(\mu_u, \sigma_u)$ otherwise. $\mu_{a/u}$ and $\sigma_{a/u}$ are, respectively, the mean and the standard deviation of each of the distributions. They can be estimated from the same training data that is used to train the AAD decoders. This estimation can even be unsupervised, as shown in [16] (e.g. for the AAD algorithm used in Section 3.2.2). It is sufficient to determine these values in a 2-speaker paradigm (independent of S), as they are speaker-independent parameters describing the attended class or unattended class. Furthermore, we assume that each speaker is a priori equally likely to be attended (i.e. $P(s = a) = 1/S$), and that only one speaker can be attended at any point in time.

To shorten notation, we will further denote $P(s = a|\mathbf{x}(i))$ as $p_s(i)$. Using Bayes theorem, we can compute $p_s(i)$:

$$p_s(i) = \frac{P(\mathbf{x}(i)|s = a)P(s = a)}{P(\mathbf{x}(i))} \tag{1}$$

$$= \frac{P(x_s(i)|\mu_a, \sigma_a)\prod_{j\neq s} P(x_j(i)|\mu_u, \sigma_u)}{\sum_{k=1}^{S} P(x_k(i)|\mu_a, \sigma_a)\prod_{j\neq k} P(x_j(i)|\mu_u, \sigma_u)}, \tag{2}$$

with

$$P(x_j(i)|\mu_{a/u}, \sigma_{a/u}) = \frac{1}{\sigma_{a/u}\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(\frac{x - \mu_{a/u}}{\sigma_{a/u}}\right)^2\right).$$

Note that the estimation of $p_s(i)$ in (2) can be replaced by any other, more accurate estimation if available. This is especially relevant for AAD algorithms that inherently estimate the probability of attention, such as Bayesian models.

2.2. Gain control system design

The gain control system regulates the volume gain of each speaker within a fixed interval $[g_{min}, g_{max}]$ based on the set of probabilities $p_s(i),\ s \in [1, \ldots, S]$. For ease of notation, we map the gain interval onto the interval $[0, 1]$, noting that a zero-gain does not necessarily imply that the speaker becomes fully silent. At each time instance i, the gain $g_s(i)$ of speaker s is updated with a step $\delta_s(i)$ as follows:

$$g_s(i) = \begin{cases} 0 & \text{if } g_s(i-1) + \delta_s(i) \leq 0 \\ 1 & \text{if } g_s(i-1) + \delta_s(i) \geq 1 \\ g_s(i-1) + \delta_s(i) & \text{otherwise.} \end{cases} \tag{3}$$

The step $\delta_s(i)$ is determined according to the following set of rules:

- **Proportional**: The step $\delta_s(i)$ must be directly and linearly proportional to the probability of attention $p_s(i)$.
- **Unbiased**: The linear relation between $\delta_s(i)$ and $p_s(i)$ must be equal for all speakers s.
- **Balanced**: The sum of all steps $\sum_s \delta_s(i) = 0$ at each time instance i.

The last rule ensures that the gain control system can only amplify one speaker by suppressing another speaker (until the gain reaches the maximal or minimal level). To satisfy the first and second rules, the steps must be of the form

$$\delta_s(i) = \frac{1}{N}p_s(i) + b, \tag{4}$$

with $N > 0$ and b equal for all speakers. To ensure that the third rule is also satisfied, and because $\sum_{s=1}^{S} p_s(i) = 1$, we get:

$$\sum_{s=1}^{S} \delta_s(i) = 0 \tag{5}$$

$$\Leftrightarrow \frac{1}{N} + Sb = 0 \tag{6}$$

$$\Leftrightarrow b = -\frac{1}{SN}. \tag{7}$$

By combining (4) and (7), we get:

$$\delta_s(i) = \frac{1}{N}(p_s(i) - \frac{1}{S}). \tag{8}$$

The hyperparameter N directly controls the trade-off between the speed and stability of the gain control system. A larger N creates a more stable, but slower system and vice versa. The exact influence of N on the gain trajectory is further experimentally analysed in Section 3.

Equation (8) is intuitive in the sense that the gain $g_s(i)$ of speaker s only increases in volume when $p_s(i) > 1/S$, i.e. when the probability that this speaker is attended is larger than chance. On average, the gain of the attended speaker increases with $\frac{1}{N}(\bar{p}_a - \frac{1}{S})$, with $\bar{p}_a$ the average estimated probability that the attended speaker is attended. In other words, as long as the AAD algorithm performs better than chance, the gain of the attended speaker will, on average, increase. This is true irrespective of the number of speakers present.

When the gain system is unsure whether a certain speaker is attended or unattended, their gain will hardly change. This avoids any sudden (random) changes in the gain during periods of high uncertainty, e.g. when the recorded subject has a lapse of attention or the EEG is heavily contaminated with artifacts. This is in contrast to the gain control systems proposed in [5, 14, 15].

3. EXPERIMENTS

To assess the behaviour of the gain control system, it is applied on the output of the least-squares (LS) AAD algorithm proposed in [2, 3]. This framework was tested on two datasets: the publicly available 2-speaker KUL dataset [17] and on pilot data of a novel dataset where three conversations between two speakers are being held simultaneously (further called the conversations dataset).

3.1. Datasets

3.1.1. 2-speaker KUL dataset

In this public dataset, a 64-channel EEG signal of 16 normal hearing Flemish speaking subjects is measured with a BioSemi ActiveTwo system. The data consists of 8 trials of 6 minutes and 12 trials of 2 minutes each. In each trial, the subject must listen to one of two competing speakers. The competing speakers are located at the left and right side of the subject, and the side of the attended speaker changes across trials. All speakers are male and narrate Flemish stories. For more details, we refer to [3, 17].

3.1.2. Conversation dataset

This is a pilot dataset for a different study, containing 64-channel EEG data from 13 normal hearing Flemish speaking subjects. The data is collected with a BioSemi ActiveTwo system and a 10-20 electrode lay-out. Per subject, 6 trials of 10 minutes each are recorded. In each trial, the subject is listening to one of three conversations, located left, right and in front of the subject. Each conversation is held between two Flemish speakers in the form of a podcast. Each speaker is represented by a loudspeaker, so there are six loudspeakers in total.

The 6 trials are divided into 3 trials where the subject sustains attention to a single conversation (once to the left, once to the front and once to the right conversation) and 3 trials where the attention switches after 5 minutes. The possible switches are from right to front, from left to front, from right to left, and vice versa.

Important to note is that in this dataset, we decode the attended conversation, rather than the attended speaker. There are thus three classes to select from, not six.

3.2. Experiment protocol

3.2.1. Preprocessing

We extracted the envelopes of all speech signals using the gammatone and power law framework from [3]. The EEG signals and the envelopes are then bandpass filtered between 1 Hz and 32 Hz and resampled to a 64 Hz sample frequency. The EEG and envelopes are then all cut in non-overlapping windows of 1 s (unless mentioned otherwise) and classified by the LS AAD algorithm outlined in Section 3.2.2.

3.2.2. Least-squares decoding

The LS AAD algorithm is exactly recreated from [3]. For each window, the attended envelope $e_a(t)$ is reconstructed from the EEG signal $\mathbf{M} \in \mathbb{R}^{T\times C}$ using a linear spatio-temporal LS filter $\mathbf{D} \in \mathbb{R}^{N_l\times C}$, with T the length of a window, N_l the number of time lags and C the number of channels:

$$\tilde{e}_a(t) = \sum_{n=0}^{N_l-1}\sum_{c=1}^{C}\mathbf{D}(n,c)\mathbf{M}(t+n,c). \tag{9}$$

The LS filter $\mathbf{D}$ is trained to minimise the square error

$$\mathbb{E}\left[(\tilde{e}_a(t) - e_a(t))^2\right]$$

between the reconstructed envelope $\tilde{e}_a(t)$ and the attended envelope $e_a(t)$. This training is done on all other trials using leave-one-trial-out cross-validation.

The reconstructed envelope $\tilde{e}_a(t)$ is then correlated with the envelope of each speaker (or conversation for the conversation dataset), and the per-speaker correlation coefficients are used as the entries of $\mathbf{x}(i)$ in the gain control system from Section 2. The correlations are expected to be higher for an attended speaker than for unattended speakers.

3.2.3. Gain control protocol

The gain control system explained in Section 2 is used to transform the set of correlations $\mathbf{x}(i)$ per window i into a sequence of relative gains per speaker $g_s(i)$. To ensure that the results are comparable between subjects and datasets, N is automatically tuned for each subject independently, such that the gain control system is equally stable for each subject (unless mentioned otherwise).

Similar to [15], the hyperparameter N in (8) is tuned by enforcing that the gain of the attended speaker is higher than a comfort level $c = 0.65$ for $P_0 = 80\%$ of the time when the gain control system is in steady state (i.e. when there was no recent switch in attention). The N that satisfies the required comfort level as closely as possible, is iteratively found by minimising $(c - \hat{c})^2$, with $\hat{c}$ the level above which the attended gain is for $P_0 = 80\%$ of the time given a certain N. This problem is solved for each subject using the built-in Matlab R2021a nonlinear solver *fsolve* with the interior point method and using leave-one-trial-out cross-validation. No attention switches were present when tuning N to ensure that the gain system is always in steady state.

The gain control system is then evaluated by empirically computing the median switch duration (MSD). This is the median time required after a switch for the attended speaker to reach the comfort level c, similar to the expected switch duration proposed in [15]. In contrast [15], the MSD is empirically determined by first computing the switch duration after every switch and then taking the median switch duration per subject. If there are no switches present in a trial, they are artificially produced at the center of the trial by switching the correlations of the attended speaker with the correlations of an unattended speaker.

3.3. Results and discussion

The median switch duration (MSD) of each subject is represented in Figure 1. The median MSD over all subjects is 19 s in the 2-speaker KUL dataset and 13 s in the conversation dataset. The difference in switch duration is not statistically significant based on a Wilcoxon rank sum test ($p = 0.15$). Remarkably, while the conversation data set is a 3-class problem (versus a 2-class problem in the 2-speaker KUL dataset), inevitably leading to lower decoding accuracies, we cannot show a difference in switch duration. This observation can be explained by looking at the average estimated probability that the attended speaker is attended $\bar{p}_a$ and its standard deviation, which is $\bar{p}_a = 52.09 \pm 10.05\%$ for the KUL dataset and $\bar{p}_a = 34.85 \pm 7.24\%$ for the conversations dataset. In both cases, the chance level (50% and 33.33% respectively) is about 0.2 standard deviations away from $\bar{p}_a$. Since the steps in the gain control are directly related to the difference between the probability that a speaker is attended and the chance level, the steps of the attended speaker in the gain control are,

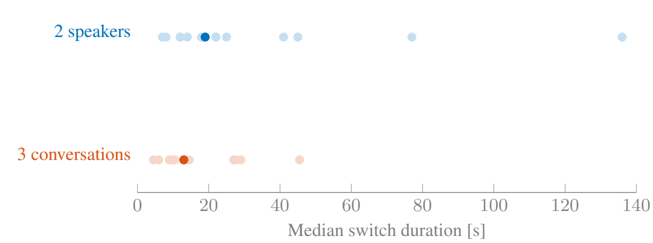

Fig. 1. The median time between the moment a subject switches attention and the moment the gain of the attended speaker is again above comfort level (0.65) is 19s in the Das dataset and 13s in the conversation dataset. The lighter colours represent the median switch duration for each subject, whereas the darker blue and red dot represent the median over all subjects. Smaller is better.

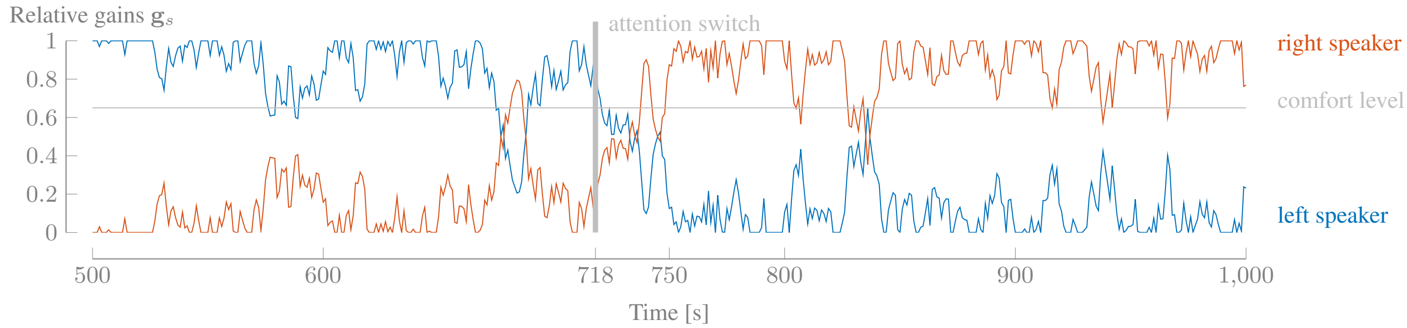

Fig. 2. A representation of the gain sequence of subject 7 from the 2-speaker KUL dataset around the switch introduced between the second and third trial. The two curves represent the gain of the left and right speaker across time. The subject attended the left speaker before and the right speaker after the switch. With a median switch duration of 22 s, subject 7 represents an average subject in the KUL dataset.

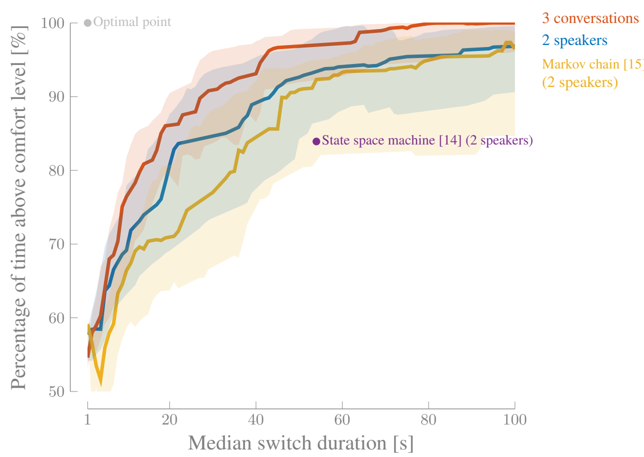

Fig. 3. There is a direct trade-off between the percentage of time that the gain control system stays above the comfort level of $c = 0.65$ in steady state (higher is better) and the median time required until the attended speaker passes above comfort level after an attention switch (lower is better). This trade-off is also represented for the gain control systems proposed in [15] on the 2-speaker KUL dataset. The performance of the state space machine proposed in [14] is only shown using the same hyperparameters as proposed in the paper since the trade-off between stability and speed cannot be easily tuned in this model. The solid line and dots represent the median and the shaded area represents the interquartile range across subjects for each dataset.

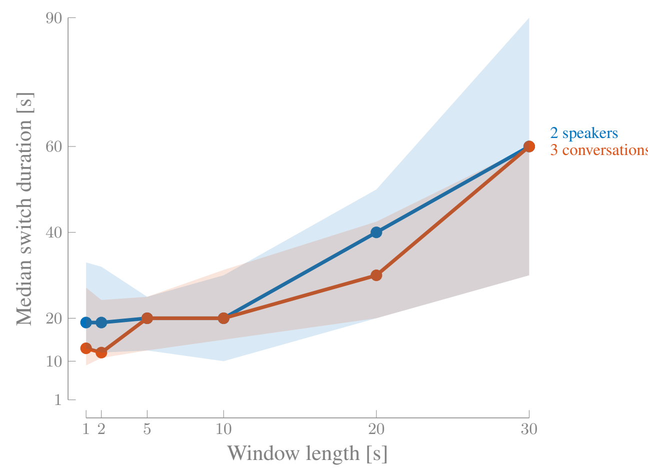

Fig. 4. The median time to switch from one speaker to another is lower when the AAD algorithm uses many short windows, rather than only a few long windows (lower is better). The shaded area represents the interquartile range for each dataset.

therefore, comparable. Note that the low values for $\bar{p}_a$ are caused by the short window lengths of 1 s, where the LS AAD algorithm achieves very low accuracies [1–3].

As an example, the gain control system of Subject 7 (whose median switch duration is 22s) is shown in Figure 2. It is clearly visible that the gain mostly stays above the comfort level $c = 0.65$ in steady state. The gain of the attended speaker rises above the comfort level 20s after the attention switch.

The ideal trade-off between stability and switching speed is naturally context-dependent. We therefore investigate the relation between the MSD and the percentage of time that the gain control system is above comfort level P_0 in a steady state. This is achieved by varying the hyperparameter N in (8) between 0.1 and 5 and then computing both the MSD and the percentage of time above comfort level for each value of N. These resulting curves are then linearly interpolated to enable comparisons between subjects. The median and IQR of the interpolated results are shown in Figure 3. For both datasets, the gain of the attended speaker rarely dips below the comfort level when an MSD over 40 s is allowed.

To compare with state-of-the-art gain control systems, the median percentage of time above comfort level and median MSD of the gain control system achieved by the Markov chain proposed by Geirnaert et al. in [15] and the state space machine proposed by Aroudi et al. in [14] are also added in Figure 3. We used the hyperparameters that were suggested by the corresponding papers as we did not find specific hyperparameter settings that substantially improved results. The Markov chain [15] was tested using the correlations obtained from non-overlapping 1 s windows and by varying the hyperparameter N between 0.1 and 5. Similar to [14], the state space machine was tested using overlapping windows of 15 s shifted with

0.25 s since the machine was too unstable on shorter windows for practical use. These two models were only tested on the 2-speaker KUL dataset, since they were nor designed, nor easy to transform to datasets with more than two competing speakers.

Until now, we only investigated the gain control system using 1 s windows. In Figure 4, we demonstrate how the window length influences the MSD of the gain control system. Similar to the results in [15], we observe that it is generally better to use short, low-accuracy windows than long windows where AAD algorithms can classify the attended speaker with much more certainty [1]. However, there is no significant difference in the MSD of windows shorter than 10 s. Note that the MSD can only be a multiple of the window length since only non-overlapping windows were used. This explains the large variance in MSD at larger window lengths.

4. CONCLUSION

We proposed and evaluated a new gain control system for auditory attention decoding (AAD). The system is designed to work along with any AAD model and in scenarios with an arbitrary number of speakers. Furthermore, the gain control system is designed to be simple in use with only a single hyperparameter that directly controls the trade-off between the speed and the stability of the gain control system.

The gain control system was evaluated on both a dataset with two speakers and a pilot dataset with three conversations. Using the least-squares AAD algorithm, it takes on average 19 s for the two speaker dataset and 13 s for the dataset with three conversations to switch attention from one person to another, given a reasonable level of stability. Surprisingly, the gain control system achieved similar results on both datasets, despite the fact that the decoding accuracy is significantly different for both datasets.

5. REFERENCES

[1] Simon Geirnaert, Servaas Vandecappelle, Emina Alickovic, Alain de Cheveigné, Edmund Lalor, Bernd T. Meyer, Sina Miran, Tom Francart, and Alexander Bertrand, "Neuro-Steered Hearing Devices: Decoding Auditory Attention From the Brain," *IEEE Signal Processing Magazine*, vol. 38, no. 4, pp. 89–102, 2021.

[2] James A. O'Sullivan, Alan J. Power, Nima Mesgarani, Siddharth Rajaram, John J. Foxe, Barbara G. Shinn-Cunningham, Malcolm Slaney, Shihab A. Shamma, and Edmund 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, 2015.

[3] Wouter Biesmans, Neetha Das, Tom Francart, and Alexander 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.

[4] Alain de Cheveigné, Daniel E. Wong, Giovanni M. Di Liberto, Jens Hjortkjær, Malcolm Slaney, and Edmund Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[5] Sina Miran, Sahar Akram, Alireza Sheikhattar, Jonathan Z Simon, Tao Zhang, and Behtash Babadi, "Real-time tracking of selective auditory attention from M/EEG: A Bayesian filtering approach," *Frontiers in Neuroscience*, vol. 12, pp. 262, 2018.

[6] Jaswanth Reddy Katthi, Sriram Ganapathy, Sandeep Kothinti, and Malcolm Slaney, "Deep Canonical Correlation Analysis For Decoding The Auditory Brain," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 2020, IEEE.

[7] Nicolas Heintz, Simon Geirnaert, Tom Francart, and Alexander Bertrand, "Unbiased Unsupervised Stimulus Reconstruction for EEG-Based Auditory Attention Decoding," in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2023, Institute of Electrical and Electronics Engineers (IEEE).

[8] Tobias Taillez, Birger Kollmeier, and Bernd T. Meyer, "Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech," *European Journal of Neuroscience*, vol. 51, no. 5, pp. 1234–1241, Mar. 2020, Publisher: Blackwell Publishing Ltd.

[9] Simon Geirnaert, Tom Francart, and Alexander Bertrand, "Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns," *IEEE Transactions on Biomedical Engineering*, 2020.

[10] Servaas Vandecappelle, Lucas Deckers, Neetha Das, Amir Hossein Ansari, Alexander Bertrand, and Tom Francart, "EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks," in *Auditory EEG Signal Processing symposium*, Feb. 2019.

[11] Gregory Ciccarelli, Michael Nolan, Joseph Perricone, Paul T. Calamia, Stephanie Haro, James O'Sullivan, Nima Mesgarani, Thomas F. Quatieri, and Christopher J. Smalt, "Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods," *Scientific Reports*, vol. 9, 2019.

[12] Sahar Akram, Alessandro Presacco, Jonathan Z. Simon, Shihab A. Shamma, and Behtash Babadi, "Robust decoding of selective auditory attention from MEG in a competing-speaker environment via state-space modeling," *NeuroImage*, vol. 124, pp. 906–917, Jan. 2016, Publisher: Academic Press Inc.

[13] Alessandro Presacco, Sina Miran, Behtash Babadi, and Jonathan Z. Simon, "Real-Time Tracking of Magnetoencephalographic Neuromarkers during a Dynamic Attention-Switching Task," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. July 2019, pp. 4148–4151, Institute of Electrical and Electronics Engineers Inc., ISSN: 1557170X.

[14] Ali Aroudi, Eghart Fischer, Maja Serman, Henning Puder, and Simon Doclo, "Closed-Loop Cognitive-Driven Gain Control of Competing Sounds Using Auditory Attention Decoding," *Algorithms*, vol. 14, no. 10, pp. 287, Oct. 2021.

[15] Simon Geirnaert, Tom Francart, and Alexander Bertrand, "An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307–317, 2020.

[16] Miguel Angel Lopez-Gordo, Simon Geirnaert, and Alexander Bertrand, "Unsupervised Accuracy Estimation for Brain-Computer Interfaces based on Selective Auditory Attention Decoding," 2024.

[17] Neetha Das, Wouter Biesmans, Alexander Bertrand, and Tom Francart, "The effect of head-related filtering and ear-specific decoding bias on auditory attention detection," *Journal of Neural Engineering*, vol. 13, no. 5, pp. 056014, Oct. 2016.