# Fast, accurate, unsupervised, and time-adaptive EEG-based auditory attention decoding for neuro-steered hearing devices

**Simon Geirnaert[1][2], Rob Zink[1], Tom Francart[2], and Alexander Bertrand[1]**

**Abstract** More than 5% of the world's population suffers from disabling hearing loss. Hearing aids and cochlear implants are crucial for improving their quality of life. However, current hearing technology does not work well in cocktail party scenarios, where several people talk simultaneously. This is mainly because the hearing device does not know which speaker the user is attending to, and so which speaker should be amplified relative to the background noise. In this project, we have developed novel signal processing algorithms for electroencephalography (EEG)-based auditory attention decoding to steer the hearing device towards the attended speaker based on the user's attention. We propose algorithms that are fast, accurate, and able to adapt automatically to (changes in) the EEG data of individual users. These are crucial ingredients towards the realization of practically viable neuro-steered hearing devices.

**Keywords** BCI – EEG – Auditory attention decoding – Hearing device

## 1. Introduction

### *1.1 The auditory attention decoding problem*

The World Health Organization estimates that more than 5% of the world's population, or 430 million people, suffers from disabling hearing loss and requires rehabilitation. This group is expected to grow to more than 7% by 2050 [1]. As hearing loss tremendously impacts society both on an economic and individual level (e.g., social isolation, loneliness), effective assistive hearing devices such as hearing aids and cochlear implants are required to restore communication by improving speech intelligibility.

While newly developed hearing devices more and more contain advanced speech enhancement and noise suppression algorithms, they still underperform in so-called 'cocktail party' scenarios, where multiple persons are talking simultaneously. In these situations, hearing devices lack a fundamental piece of knowledge: They do not know which speaker the hearing device user *wants* to attend to. Correspondingly, they have no information about which speech signal to enhance (i.e., the attended speaker) and which others to consider as noise and thus suppress (i.e., the unattended speaker(s)). We refer to this problem as the *auditory attention decoding* (AAD) problem.

While the AAD problem can be tackled using heuristics such as eye gaze direction or by simply selecting the loudest speaker, they fail in several practical scenarios, for example, when listening to a public address

[1] STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Heverlee, Belgium
[2] ExpORL, Department of Neurosciences, KU Leuven, Leuven, Belgium
S. Geirnaert, T. Francart, and A. Bertrand are also affiliated with Leuven.AI – KU Leuven Institute for AI, Leuven, Belgium
Email: {simon.geirnaert,tom.francart,alexander.bertrand}@kuleuven.be

system, eavesdropping, or when driving a car. Selecting and thereby amplifying the wrong (unattended) speaker can be avoided by pursuing a more ideal strategy: decoding the auditory attention from where it originates, i.e., the brain. As shown by Mesgarani and Chang [2], certain characteristics (such as the speech envelope) of the attended speech signal are better encoded in the brain than those of the unattended speech signal(s), which, in turn, opens the possibility to develop algorithms to decode the auditory attention from brain signals. Incorporating such AAD algorithms in a hearing device could then lead to a new type of brain-computer interface (BCI) technology to assist hearing-impaired people: a *neuro-steered hearing device* (Figure 1). There exist other use cases of AAD in BCIs as well, such as in consumer earphones and other hearables [3].

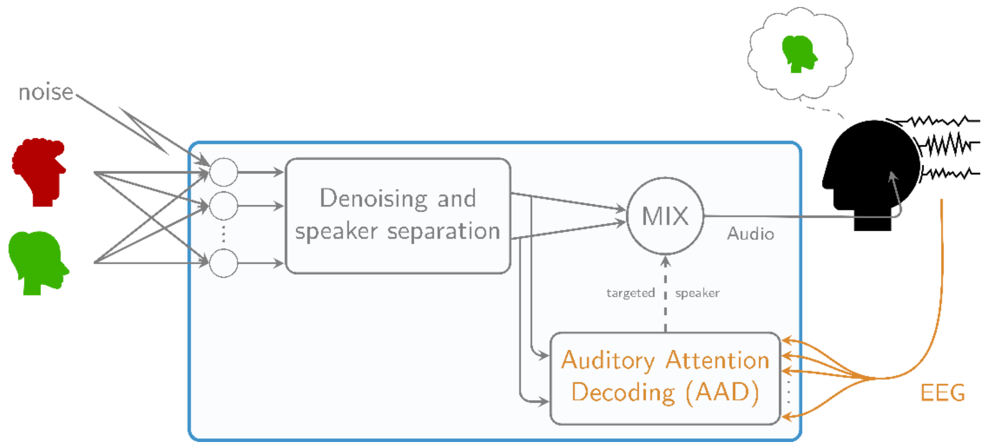


*Figure 1: In a neuro-steered hearing device, the AAD algorithm informs the hearing device about which speaker needs to be amplified based on the EEG signals. Based on Figure 1 in [4].*

*1.2 Neuro-steered hearing devices*

Figure 1 shows a conceptual overview of a neuro-steered hearing device with its main ingredients: a speaker separation and enhancement block, an AAD block to determine the attended speaker based on the brain signals of the listener, and a mixing block that mixes the separated speech signals based on the retrieved information about the attended speaker.

We use electroencephalography (EEG) as a neurorecording modality because it is non-invasive, wearable, and relatively cheap. These are all crucial features for the widespread usage of neuro-steered hearing devices during daily-life activities. Furthermore, EEG has an excellent temporal resolution, which is critical for tracking fast modulations in speech and for low-latency processing in BCIs in general.

One of the most common paradigms for AAD – *stimulus reconstruction* (SR) – is based on the principle of neural tracking of speech signals, i.e., the auditory cortex tracks time-varying characteristics of the attended speech stimulus [5]. Correspondingly, it has been shown that in a cocktail party scenario with two competing speakers, it is possible to reconstruct from the neural signals a spectrogram [2] or envelope [6] that better reflects the spectrogram or speech envelope of the attended speech signal than the unattended one. The SR algorithm then exploits this stronger neural tracking of the attended speech signal for AAD by reconstructing an envelope from the EEG signals of the listener using a neural decoder and correlating this reconstructed envelope with the individual speech envelopes of the competing speakers [7]. The speech envelope of the speaker that exhibits the highest correlation is then identified as the attended speaker (Figure 2a). This neural decoder can be modeled and trained in various ways. In [4], we provide an extensive review of the different ways to train the neural decoder, as well as other AAD algorithms beyond SR.

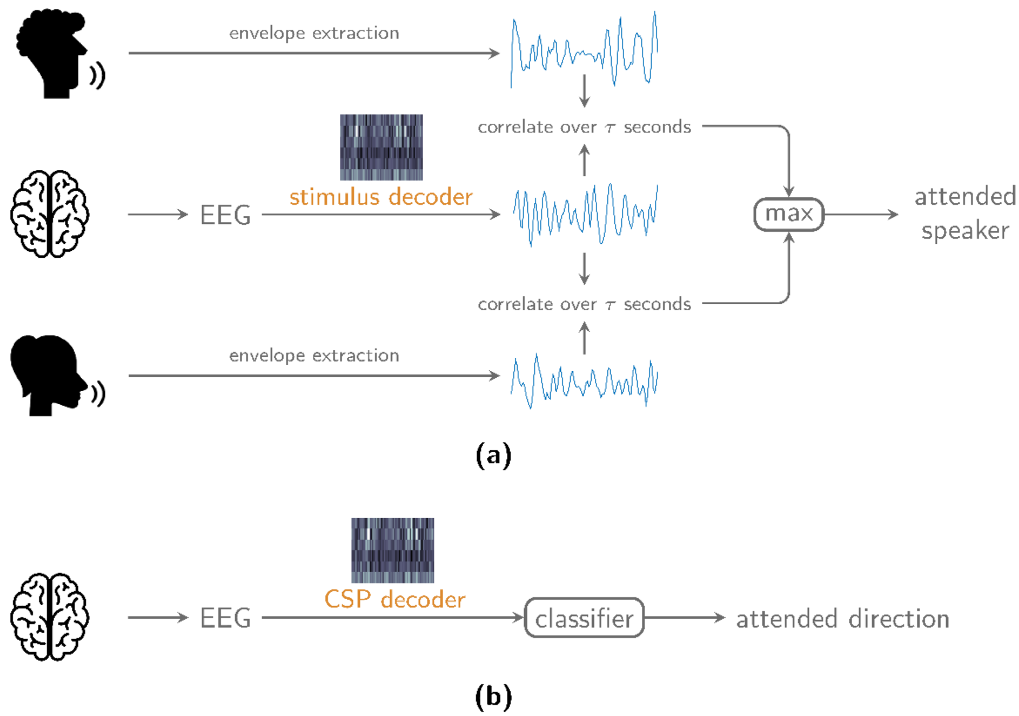


*Figure 2:* ***(a)*** *In the SR algorithm, a neural stimulus decoder is applied on the EEG to reconstruct the attended speech envelope. Based on Figure 3a in [4].* ***(b)*** *In the CSP algorithm, the attended spatial direction is directly decoded from the EEG lateralization patterns.*

*1.3 Two fundamental problems with stimulus reconstruction*

The SR algorithm suffers from two critical limitations that are dealt with in this chapter:

1. The neural stimulus decoder is traditionally trained in a *supervised* manner and remains *fixed over time* [7]. To be able to train this neural decoder to reconstruct the attended speech envelope from the EEG in a data-driven way, the attended speaker needs to be known at training time. The necessity for attention labels during training requires a cumbersome calibration session for each new user, in which the user is instructed to attend to a particular speaker. Furthermore, this also prevents adapting the neural decoder to changes in the EEG signal characteristics, making it suboptimal.
2. The SR algorithm suffers from a quickly decreasing accuracy when reducing the number of EEG/speech envelope samples to make a decision (the decision window length) (see Figure 4a). This is an effect of the notoriously low signal-to-noise ratio of the stimulus-following neural response in the EEG, making the estimation of the correlation coefficients highly susceptible to interference from other neural processes when reducing the decision window length [4]. As a result, the SR algorithm is too slow in detecting switches in auditory attention, resulting in too large delays in changing the relative gain between the competing speakers in the hearing device [8].

Therefore, we have developed novel AAD signal processing algorithms to overcome both challenges, i.e., by designing an unsupervised and time-adaptive SR algorithm (Section 2) and exploiting an alternative AAD paradigm: decoding the spatial focus of auditory attention (Section 3).

## 2. Unsupervised and time-adaptive stimulus reconstruction

### *2.1 Supervised training of the neural stimulus decoder*

Traditionally, the neural stimulus decoder in the SR algorithm (Figure 2a) is a linear spatio-temporal filter, integrating EEG channels and post-stimulus time lags to reconstruct a sample of the attended speech envelope [7, 8]. Given a training set of EEG data and speech envelopes of the competing speakers, the filter weights of this neural decoder can be found by minimizing the mean squared error (MMSE) between the reconstructed envelope from the EEG and the attended speech envelope[3]. This training procedure is conceptually summarized in Figure 3a. A crucial feature of this training procedure is that it is *supervised*, i.e., it requires the ground-truth attention labels to select the correct speech envelope as the target during training.

Depending on the training data used in this supervised training procedure, two versions of the neural stimulus decoder exist: a user-specific or user-independent decoder [7, 9]. The former is trained with data from the end user, while the latter is trained with data from other users. As expected, the user-specific decoder leads to higher AAD performances (see Figure 4a), as it is tailored towards the brain of the specific end user [7, 9]. However, it lacks the plug-and-play characteristic of the user-independent decoder that can be pre-implemented on a hearing device, whereas the user-specific decoder requires a cumbersome training session for each new end user [9]. In this training session, the user then needs to be instructed to listen to one of multiple competing speech signals to be able to generate the required ground-truth labels.

In [9], we proposed an *unsupervised but user-specific* stimulus decoder to combine the best of both worlds. As it is unsupervised, i.e., it does not require ground-truth attention labels, it retains part of the plug-and-play characteristic of the user-independent decoder, while it can achieve higher performances because it still uses user-specific data. In Section 2.2, we explain how this unsupervised training procedure works using a batch of EEG/speech envelope training data but *without* ground-truth attention labels, while its extension to a time-adaptive updating procedure is explained in Section 2.3.

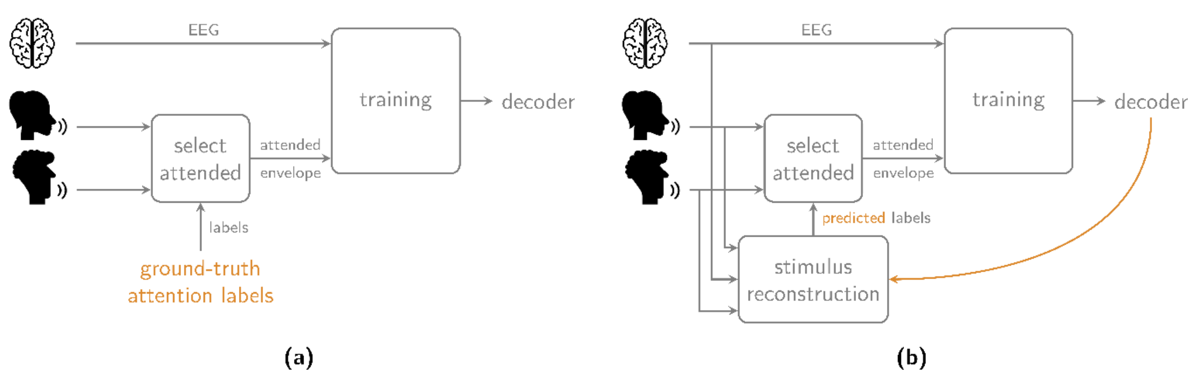


*Figure 3: **(a)** The neural stimulus decoder is traditionally trained in a supervised manner, i.e., the ground-truth attention labels are required to select the attended speech envelope. **(b)** In [9], we propose an*

[3] An extensive mathematical explanation can be found in, e.g., [9].

*unsupervised training procedure where the ground-truth attention labels are replaced by the predicted labels when applying the learned decoder.*

*2.2 Unsupervised training of the neural stimulus decoder*

Again assume the availability of a batch of EEG/speech envelope training data, however, now *without* knowing which of the two competing speakers corresponds to the attended speaker at each point in time. Correspondingly, the training scheme as proposed in Figure 3a has become unfeasible due to the unavailability of the ground-truth attention labels. Therefore, in [9], we proposed an alternative unsupervised training scheme that replaces the ground-truth attention labels with the predicted labels on the training data using a previously trained decoder, as conceptually shown in Figure 3b. The key idea is that this creates a closed-loop, iterative predict-and-update procedure, where

1. the trained decoder is used to (re)predict attention labels on the training data using SR,
2. the (re)predicted attention labels are used to (re)select the attended speech envelope such that the neural stimulus decoder can be retrained. Step 1 can then be repeated.

When using an MMSE-based linear spatio-temporal filter as stimulus decoder, we have shown that this procedure converges after a few iterations (under some mild conditions) and that it can be interpreted as a fixed-point iteration algorithm. This interpretation explains the self-leveraging effect in the iterative updating procedure, i.e., it leverages its own predicted labels to converge to a better decoder, even when starting from a random decoder. Furthermore, in the spirit of transfer learning, we have developed a version that allows incorporating labeled data from other users than the end user (as in the user-independent decoder) in the updating procedure.

Figure 4a shows the AAD accuracy (number of correct decisions) versus the decision window length (number of time samples used to make a decision) for the supervised user-specific decoder, the supervised user-independent decoder, the unsupervised user-specific decoder when starting from a random decoder, and the unsupervised user-specific decoder with user-independent side-information[4]. The unsupervised but user-specific decoder substantially outperforms the user-independent decoder while retaining the ‘plug-and-play’ characteristic, not requiring ground-truth attention labels. Adding labeled user-independent side information in the unsupervised updating procedure even enables accuracies close to the optimal supervised user-specific decoder.

This unsupervised training procedure opens the possibility for a fully automatically updating decoder over time, as explained in the following section.

[4] All details about the data and experiments can be found in [9].

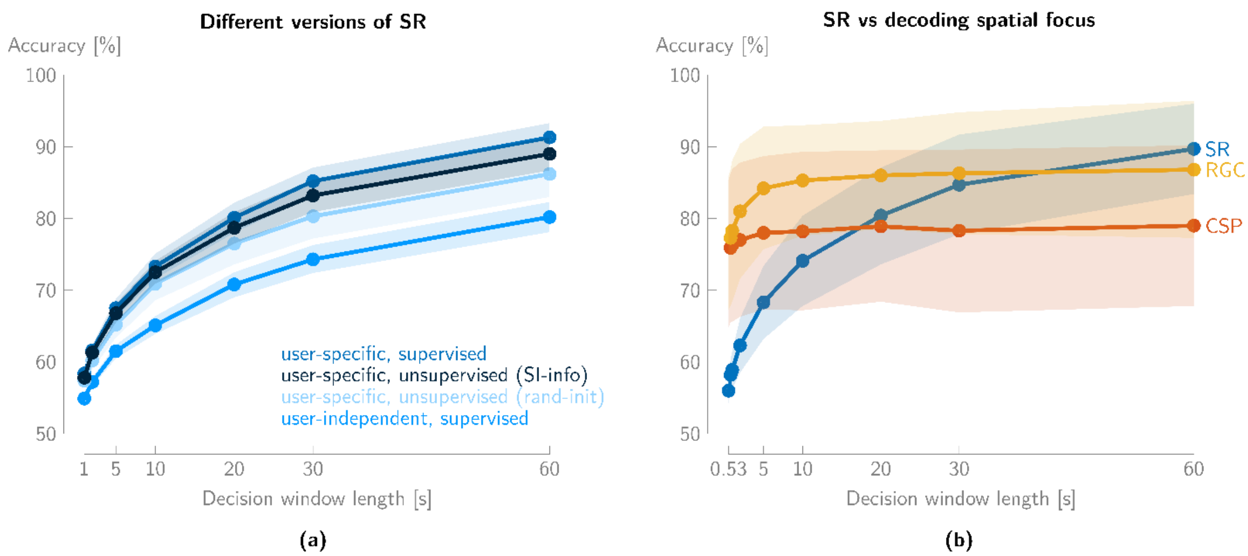


*Figure 4:* ***(a)*** *Different versions of the linear SR algorithm, all showing a degrading accuracy for short decision window lengths. Based on Figure 5a from [9].* ***(b)*** *Decoding the spatial focus of auditory attention with CSPs or RGCs gives a substantial improvement for the shorter decision window lengths. Based on Figure 1 from [18].*

### *2.3 Time-adaptive unsupervised updating of the neural stimulus decoder*

While the unsupervised training procedure for the neural stimulus decoder in Section 2.2 removes the necessity of having access to ground-truth attention labels during training, it still assumes the availability of a batch of EEG/speech envelope data. Furthermore, it results in a decoder that remains fixed during operation, not adapting to the long-term signal changes that are characteristic of EEG. These non-stationarities in the EEG originate, for example, from the inherent non-stationarity of neural signals, changing electrode-skin contact impedances, and shifting or loosening electrodes. Modifying the unsupervised training algorithm from Section 2.2 to be time-adaptive is therefore crucial to adapt to these non-stationarities, in order to obtain higher performances. Furthermore, it also alleviates the need for a dedicated training session for each new end user, fully realizing the plug-and-play potential of the unsupervised decoder.

Consider now the practical use case where EEG data and speech envelopes of the competing speakers are continuously being recorded. In [10], we proposed a single-shot predict-and-update scheme based on a recursive implementation of the linear neural stimulus decoder. The key idea is to predict the attention label using the previous decoder on the new incoming data window and to use this predicted label to update the parameters of the decoder using exponential weighting. This exponential weighting hyperparameter then determines the tradeoff between the speed of adaptation to non-stationarities and the accuracy of the resulting decoder. From our experiments, a good choice for this hyperparameter led to an adaptation time of 20 minutes to adapt the decoder to a new end user, starting from a random decoder.

Figure 5a compares the pretrained fixed supervised decoder with the proposed time-adaptive unsupervised decoder when simulating disconnecting EEG electrodes during an AAD experiment (on a 64-

channel BioSemi EEG system)[5]. After disconnecting EEG electrodes, the time-adaptive unsupervised decoder finds alternative ways of reconstructing the attended speech envelope that are clearly as effective as before. The pretrained fixed supervised decoder cannot take these disconnections into account and, therefore, shows a decreasing performance. From the moment that more than three electrodes are disconnected, it is outperformed by the time-adaptive decoder, which requires no pretraining and updates fully automatically. Furthermore, as shown in Figure 5b, the time-adaptive unsupervised decoder performs at least as well as the fixed supervised decoder in an AAD experiment that reflects the practical use-case of a neuro-steered hearing device in which data is collected across multiple recording days. The time-adaptive decoder then adapts to the changes in, for example, EEG setup, electrode impedances, conditions, and state of mind of the user, *without* any external intervention or supervision.

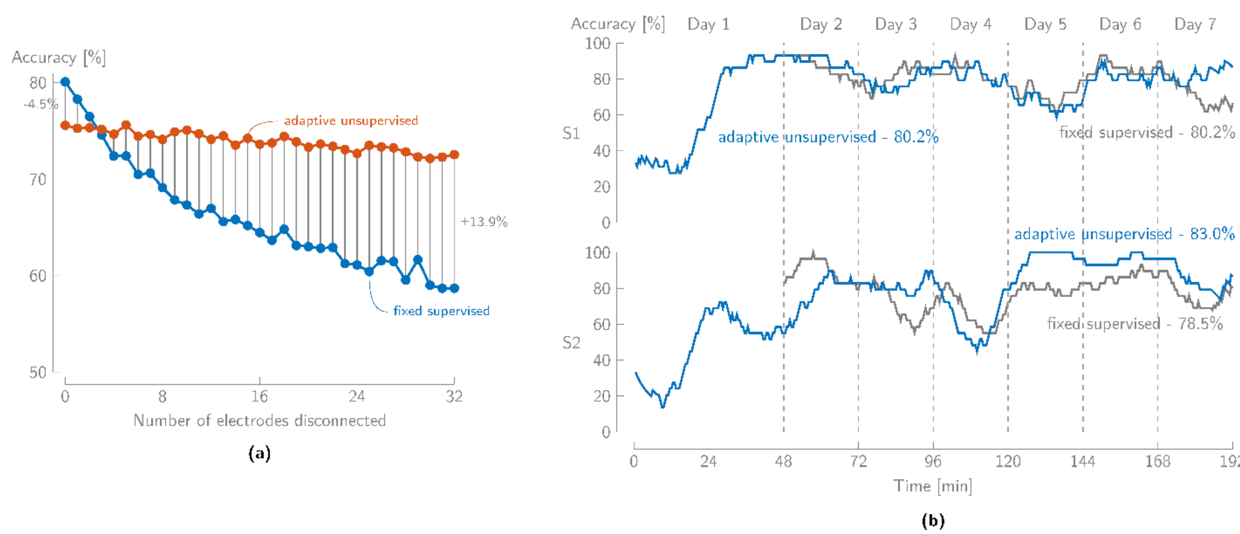


*Figure 5: **(a)** When disconnecting electrodes from a 64-channel BioSemi EEG system during an AAD experiment, the pre-trained fixed supervised decoder shows degrading performances, while the time-adaptive unsupervised decoder can find alternative ways of reconstructing the attended speech envelope that are as effective. **(b)** The time-adaptive unsupervised decoder performs at least as good as the fixed supervised decoder in an AAD experiment across multiple recording days (two participants). Based on Figure 7a and 8b in [10].*

These various results show the added value of the developed time-adaptive unsupervised SR algorithm that allows adaptation to the EEG signal changes and to a specific end-user. As such, it is a crucial enabler towards the online application of AAD in practical neuro-steered hearing devices.

**3. Decoding the spatial focus of auditory attention**

While the availability of a time-adaptive unsupervised SR algorithm represents an essential step towards practical AAD, it is still based on the SR algorithm that is too slow to adequately detect switches in auditory attention (see Section 1.3 and [8]). Therefore, there has been an increasing interest in decoding the auditory attention by tapping into other characteristics of the neural activity, i.e., other than neural tracking of the speech envelope. One of these characteristics is based on the hypothesis that the neural

---

[5] All details about the data and experiments can be found in [10].

signals change depending on the direction of spatial attention to the attended speaker. For example, it has been shown that there exist alpha-power lateralization patterns depending on the attended spatial location in a competing-stimuli experiment [11, 12]. Bednar and Lalor [13] tried reconstructing the attended sound source trajectory from the EEG in a moving competing speaker scenario. Specifically for AAD, Vandecappelle et al. [14] proposed to use a convolutional neural network to discriminate between left and right attended solely based on the EEG. Using such an AAD algorithm also impacts the conceptual overview of a neuro-steered hearing device (Figure 1), as it does not require the individual speech signals anymore for AAD itself. Furthermore, the attended spatial direction is a different piece of information that now comes out of the AAD block, which can be used, for example, to steer a beamformer towards the correct location.

In [15], we proposed an alternative algorithm to decode the spatial focus of auditory attention based on the popular common spatial patterns (CSP) filtering algorithm in BCI literature [16]. The idea is to apply a neural CSP decoder on the EEG signals to perform a smart dimensionality reduction that amplifies the discriminative patterns between, for example, left/right attended. The attended direction can be classified by computing features based on the energy of the output signals in a particular decision window (Figure 2b). Figure 4b shows that exploiting this alternative paradigm with the CSP algorithm substantially improves accuracy on the very short decision windows, which is paramount to detect switches in auditory attention. Furthermore, we have shown that this also works in a three-class scenario with multiple directions of attention and background babble noise, with a reduced number of EEG channels around the ear, and in a user-independent context [15]. Lastly, in [17], we employed a nonlinear extension based on Riemannian geometry-based classification (RGC) [18] to obtain even higher performances (Figure 4b), especially in the multi-class scenario [19].

The high AAD accuracy at short decision windows of these algorithms that decode the spatial focus of auditory attention make them excellent candidates as fast and accurate decision-making AAD algorithms. However, these algorithms do not seem to work well in every scenario, and it is currently unclear what boundary conditions are required to make them work. Furthermore, there seems to be a considerable time dependency in the CSP decoding, in the sense that decoding performance drops or sometimes completely fails for test segments that are recorded (much) later in time than the segments used to train the decoder [19]. Further research towards these boundary conditions, the underlying neural processes that drive the CSP decoding, and the time dependency is required to realize the potential that these algorithms hold.

## 4. Conclusion and future challenges

In this project, we have contributed several key components towards the practical application of AAD in neuro-steered hearing devices. The time-adaptive unsupervised AAD algorithm is a crucial enabler towards the online application of AAD, being able to automatically adapt to the end-user and the non-stationarities in the EEG data. Furthermore, a faster and more accurate AAD algorithm could be obtained by exploiting the varying EEG lateralization patterns based on the attended spatial direction. These innovations lead to several follow-up challenges, such as evaluating the CSP and RGC algorithms in various listening scenarios to identify their boundary conditions, resolving the time-dependency that exists in the CSP algorithm between training and testing data, and building a time-adaptive unsupervised CSP/RGC algorithm, for example, by combining it with the time-adaptive unsupervised SR algorithm.

More generally, the main signal processing-related challenges to realize practical neuro-steered hearing devices are mainly situated in integrating all the different building blocks (see Figure 1). Firstly, while several papers investigate the combination of speech enhancement algorithms with SR [20-23], a similar integration between beamforming algorithms and algorithms that decode the spatial focus of attention would be interesting to research. An exciting alternative approach is combining speech enhancement and AAD in an all-in-one algorithm [24, 25]. Secondly, an adaptive gain control system should be installed based on the AAD decisions that takes the tradeoff between speed and accuracy [8] and subjective listening parameters into account. Thirdly, a practical neuro-steered hearing device requires a wearable and concealable EEG recording system using miniaturized EEG sensors, such as in-ear [26] or around-the-ear [27] EEG. While AAD has been tested with channel selection algorithms [28] and wearable EEG systems [29], there is not yet a practically viable solution. Finally, the human in the loop needs to be considered as well. It is therefore paramount to investigate neurofeedback effects on AAD in a closed-loop system in real-life scenarios [30]. Using the time-adaptive SR algorithm, such a closed-loop system could tap into the concept of co-adaptive learning, in which both the human and algorithm learn over time, hopefully resulting in improved performance.

**Acknowledgements** This work was supported by an Aspirant Grant from the Research Foundation – Flanders (FWO) (for S. Geirnaert – 1136219N), FWO project nos. G0A4918N and G081722N, the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant agreement No 802895), and the Flemish Government (AI Research Program).

This project summary is written for the BCI Award 2022 and is based on the collection of papers in the PhD thesis of Simon Geirnaert [19].

**References**

[1] World Health Organization, "World Report on Hearing", Technical Report, 2021.

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

[3] J. Belo, M. Clerc, and D. Schön, "EEG-Based Auditory Attention Detection and Its Possible Future Application for Passive BCI," *Frontiers in Computer Science*, vol. 3, no. 661178, 2021.

[4] S. Geirnaert, S. Vandecappelle, E. Alickovic, A. de Cheveigné, 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.

[5] S. J. Aiken and T.W. Picton, "Human Cortical Responses to the Speech Envelope," *Ear and Hearing*, vol. 29, no. 2, pp. 135-157, 2008.

[6] 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. 11854-11859, 2012.

[7] 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.

[8] S. Geirnaert, T. Francart, and A. Bertrand, “An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307-317, 2020.

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

[10] S. Geirnaert, T. Francart, and A. Bertrand, “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.

[11] J. R. Kerlin, A. J. Shahin, and L. M. Miller, “Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party”,” *Journal of Neuroscience*, vol. 20, no. 2, pp. 620-628, 2010.

[12] M. Wöstmann, B. Herrmann, B. Maess, and J. Obleser, “Spatiotemporal dynamics of auditory attention synchronize with speech,” *Proceedings of the National Academy of Science of the United States of America*, vol. 113, no. 14, pp. 3873-3878, 2016.

[13] A. Bednar and E. C. Lalor, “Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG,” *NeuroImage*, vol. 205, no. 116283, 2020.

[14] 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, e56481, 2021.

[15] 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.

[16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis,” *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41-56, 2007.

[17] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, “Multiclass Brain-Computer Interface Classification By Riemannian Geometry,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920-928, 2012.

[18] S. Geirnaert, T. Francart, and A. Bertrand, “Riemannian Geometry-Based Decoding of the Directional Focus of Auditory Attention Using EEG,” in Proceedings of the *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1115-1119, 2021.

[19] S. Geirnaert, “Signal Processing Algorithms for EEG-based Auditory Attention Decoding,” PhD thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 2022. Available: https://theses.eurasip.org/theses/921/signal-processing-algorithms-for-eeg-based/.

[20] 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.

[21] C. Han, J. A. O'Sullivan, Y. Luo, J. Herrero, A. D. Mehta, and N. Mesgarani, "Speaker-independent auditory attention decoding without access to clean speech sources," *Science Advances*, vol. 5, no. eaav6134, 2019.

[22] A. Aroudi and S. Doclo, "Cognitive-Driven Binaural Beamforming Using EEG-Based Auditory Attention Decoding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 862–875, 2020.

[23] N. Das, J. Zegers, H. Van hamme, T. Francart, and A. Bertrand, "Linear versus deep learning methods for noisy speech separation for EEG-informed attention decoding," *Journal of Neural Engineering*, vol. 17, no. 4, 046039, 2020.

[24] M. Hosseini, L. Celotti, and É. Plourde, "End-to-End Brain-Driven Speech Enhancement in Multi-Talker Conditions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1718-1733, 2022.

[25] 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.

[26] S. L. Kappel, M. L. Rank, H. O. Toft, M. Andersen, and P. Kidmose, "Dry-Contact Electrode Ear-EEG," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 150–158, 2019.

[27] S. Debener, R. Emkes, M. De Vos, and M. G. Bleichner, "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear," *Scientific Reports*, vol. 5, no. 16743, 2015.

[28] 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.

[29] 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.

[30] R. Zink, S. Proesmans, A. Bertrand, S. Van Huffel, and M. De Vos, "Online detection of auditory attention with mobile EEG: closing the loop with neurofeedback," *bioRxiv*, 218727, 2017.