EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI

Joan Belo, Maureen Clerc, Daniele Schön

To cite this version:
Joan Belo, Maureen Clerc, Daniele Schön. EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI. Frontiers in Computer Science, Lausanne: Frontiers Media SA, 2021, 3, 10.3389/fcomp.2021.661178 . hal-03215168

HAL Id: hal-03215168
https://hal.archives-ouvertes.fr/hal-03215168
Submitted on 3 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI

Joan Belo *, Maureen Clerc ¹ and Daniele Schön ²

¹ Inria, Université Côte d’Azur, Valbonne, France, ² Dynamics of Cognitive Processes Group (DCP), CNRS, INS, Aix-Marseille Université, Marseille, France

The ability to discriminate and attend one specific sound source in a complex auditory environment is a fundamental skill for efficient communication. Indeed, it allows us to follow a family conversation or discuss with a friend in a bar. This ability is challenged in hearing-impaired individuals and more precisely in those with a cochlear implant (CI). Indeed, due to the limited spectral resolution of the implant, auditory perception remains quite poor in a noisy environment or in presence of simultaneous auditory sources. Recent methodological advances allow now to detect, on the basis of neural signals, which auditory stream within a set of multiple concurrent streams an individual is attending to. This approach, called EEG-based auditory attention detection (AAD), is based on fundamental research findings demonstrating that, in a multi speech scenario, cortical tracking of the envelope of the attended speech is enhanced compared to the unattended speech. Following these findings, other studies showed that it is possible to use EEG/MEG (Electroencephalography/Magnetoencephalography) to explore auditory attention during speech listening in a Cocktail-party-like scenario. Overall, these findings make it possible to conceive next-generation hearing aids combining customary technology and AAD. Importantly, AAD has also a great potential in the context of passive BCI, in the educational context as well as in the context of interactive music performances. In this mini review, we firstly present the different approaches of AAD and the main limitations of the global concept. We then expose its potential applications in the world of non-clinical passive BCI.

Keywords: AAD, EEG, auditory attention, passive BCI, education, art

INTRODUCTION

The ability to discriminate and attend one specific sound source in a complex auditory environment is of utmost importance in the animal world both in terms of avoiding dangers and finding mates. In humans, this ability goes well-beyond survival and reproduction since it is a fundamental skill for efficient communication. Indeed, it allows us to follow a family conversation or discuss with a friend in a bar. In music, this ability is challenged by the simultaneous layering of several instruments playing together, requiring sound source segregation to fully appreciate the ensemble. This ability is also challenged in hearing-impaired individuals and more precisely in those with a cochlear implant (CI). Indeed, due to the limited spectral resolution of the implant, auditory perception remains quite poor in a noisy environment or in presence of simultaneous auditory sources. Thus, being able to enhance the relevant/attended source would facilitate source separation in individuals with CI. However, monitoring the attended auditory source is not easy, as this changes in time.
Recent methodological advances allow now to detect, on the basis of neural signals, which auditory stream within a set of multiple concurrent streams an individual is attending to. This approach, called EEG-based auditory attention detection (AAD), is based on fundamental research findings demonstrating that, in a multi speech scenario, cortical tracking of the envelope of the attended speech is enhanced compared to the unattended speech (Mesgarani et al., 2009; Ding and Simon, 2012; Mesgarani and Chang, 2012; Pasley et al., 2012; Zion Golumbic et al., 2013). Following these findings, other studies showed that it is possible to use EEG/MEG to explore auditory attention during speech listening in a Cocktail-party-like scenario (Ding and Simon, 2012; O’Sullivan et al., 2015; Akram et al., 2016). This field of research has grown rapidly and several new methods and techniques were developed in the last years to improve the first attempts.

Overall, these findings make it possible to conceive next-generation hearing aids combining customary technology and AAD. Importantly, AAD has also a great potential in the context of passive BCI, in the educational context as well as in the context of interactive music performances.

In this mini review, we firstly present the different approaches of AAD and the main limitations of the global concept. We then expose its potential applications in the world of non-clinical passive BCI.

The main rationale behind this mini-review is to bridge the EEG-based AAD and Passive BCI communities and to provide insights about how the emerging synergy will develop. While previous reviews have been published on technical aspects of AAD, this mini-review attempts to briefly present EEG-based AAD in a broader perspective and to guide the reader to the most relevant sources. The methodology used to find and include papers in the current mini-review was as follows. The search was carried on using both Pubmed and Google Scholar. Keywords included machine learning, decoding, encoding, auditory attention, EEG, and speech. Pubmed gave 88 results and Scholar 8,460 results. These results were then filtered with the following exclusion criteria: articles about engineering techniques that are not directly in relation with EEG-based AAD methodology, articles with methods that were not applied to M/EEG data, articles that were not published in a peer-review journal, articles that were cited <1 time. This reduced the number of included articles to 20 (see Table 1).

## EEG-BASED AUDITORY ATTENTION DETECTION METHODS

There are many different AAD methods based on EEG measures. Identifying the attended speaker using cortical activity measurement is possible because the amplitude envelope of the speech stream (a crucial feature for speech comprehension) is represented in the theta and gamma oscillatory activity in the human auditory cortex (Nourski et al., 2009; Giraud and Poeppel, 2012; Kubanek et al., 2013). Attending a source thus results in greater coupling between the envelope of the source and the envelope of neural activity in these bands.

The vast majority of the studies that explored EEG-based AAD performances used two concurrent spatially separated talkers but some of them have explored the impact of speaker number and their location in auditory scene (Schäfer et al., 2018), background noise (Das et al., 2018), reverberation (Fuglsang et al., 2017), number of EEG electrodes (Mirkovic et al., 2015; Bleichner et al., 2016), or even their location (Fiedler et al., 2017) on the performance of AAD algorithms.

One can distinguish two main categories of approaches to detect auditory attention: linear and non-linear models (see Geirnaert et al., 2020 for a comprehensive review of AAD Algorithms).

### Linear Models

In the community of linear models, two main “philosophies” are in competition (see Alickovic et al., 2019 for a complete review on linear models): forward, or encoding (encoding because these models are a description of how the system encodes information), and backward, or decoding, models.

The objective of the forward strategy is to predict the neural response in the neural data (i.e., EEG channels) from the representation of the audio signal via a temporal response function (i.e., an encoder) that describes the linear relationship between a set of neural data and an audio stimulus at certain time points (Crosse et al., 2016). In the simplest case (i.e., one audio signal) a unique representation of the audio signal is created. This representation can be the amplitude envelope (O’Sullivan et al., 2015), the spectrogram of speech signal (O’Sullivan et al., 2017), or the Mel spectrogram for a music signal (Cantisani et al., 2019). Depending on the type of the chosen representation the analysis can be either univariate (an amplitude envelope is a univariate stimulus feature) or multivariate (a spectrogram is a multivariate stimulus feature). Although it is possible to use multivariate TRF with the forward approach, this strategy is, by nature, univariate (Crosse et al., 2016). Afterward, the audio representation is convolved with an unknown channel-specific TRF. To estimate the TRF (i.e., fit the model parameters), an error minimization is performed between the neural response and the one predicted by the convolution (e.g., Mean-Squared Error) using assumptions about noise distribution (Holdgraf et al., 2017). Once the model’s parameters have been estimated, the model is validated on new data. These new data could be from the same dataset used to estimate the parameters (leave-n-out procedure) or from data recorded separately. The validation step is crucial because, to be interpretable, the model should be compatible with new data and make accurate predictions (generalization ability). Finally, the rationale of the forward strategy, in auditory research, is to predict neural data on the basis of the sound’s features.

Backward models work similarly but by predicting the auditory representation based on neural data (Alickovic et al., 2019). A pre-trained neural linear decoder is applied to the neural data to reconstruct the chosen representation (this is the reason why this type of approach is sometimes called “stimulus reconstruction”). The reconstructed representation is compared to the original representations. A high similarity (correlation) indicates a good performance of the model. Two other approaches can also be mentioned: Canonical Correlation...
## TABLE 1 | Table describing main important characteristics of AAD reviewed articles.

<table>
<thead>
<tr>
<th>Article</th>
<th>Data</th>
<th>Method</th>
<th>Subject</th>
<th>Audio features</th>
<th>AV model goodness</th>
<th>AV classification accuracy</th>
<th>Decision window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akram et al., 2016</td>
<td>MEG</td>
<td>Non-linear (SSM)</td>
<td>11</td>
<td>Amp Env</td>
<td>–</td>
<td>74% (Not sure)</td>
<td>60 s (Not sure)</td>
</tr>
<tr>
<td>Bleichner et al., 2016</td>
<td>EEG (+ cEEGrid)</td>
<td>ERP classification</td>
<td>20</td>
<td>–</td>
<td>–</td>
<td>70% (EEG)–66% (cEEGrid)</td>
<td>–</td>
</tr>
<tr>
<td>Cantisani et al., 2019</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>8</td>
<td>Amp Env (AE), Magnitude Spec (MAG), Mel Spec (MEL)</td>
<td>( r = 0.054 ) (AE), ( r = 0.215 ) (MAG), ( r = 0.119 ) (MEL)</td>
<td>F1 score = 51 (AE), 72 (MAG), 73 (MEL)</td>
<td>24 s</td>
</tr>
<tr>
<td>Ciccarelli et al., 2019</td>
<td>EEG</td>
<td>Linear (SR) and Non-linear (DNN)</td>
<td>28</td>
<td>Amp Env</td>
<td>( r = \sim 0.06 ) (Speaker separation = 10(^\circ), SNR = \sim 7.1 db)–( r = \sim 0.14 ) (Speaker separation = 180(^\circ), SNR = \sim 1.1 db) [Attended speaker]</td>
<td>97% (Linear), 81% (Non-linear)</td>
<td>30 s</td>
</tr>
<tr>
<td>Das et al., 2018</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>20</td>
<td>Amp Env</td>
<td>( r = \sim 0.06 ) (Speaker separation = 10(^\circ), SNR = \sim 7.1 db)–( r = \sim 0.14 ) (Speaker separation = 180(^\circ), SNR = \sim 1.1 db) [Attended speaker]</td>
<td>97% (Speaker separation = 180(^\circ), SNR = \sim 1.1 db)–59% (Speaker separation = 10(^\circ), SNR = \sim 7.1 db)</td>
<td>10 s</td>
</tr>
<tr>
<td>de Cheveigné et al., 2018</td>
<td>EEG</td>
<td>Linear (CCA)</td>
<td>8</td>
<td>Amp Env</td>
<td>( r = \sim 0.3 )</td>
<td>~96–~75% (Best CC pairs)</td>
<td>60–10 s</td>
</tr>
<tr>
<td>de Taillez et al., 2017</td>
<td>EEG</td>
<td>Non-linear (NN)</td>
<td>20</td>
<td>Amp Env</td>
<td>–</td>
<td>97.6–67.8%</td>
<td>60–2 s</td>
</tr>
<tr>
<td>Vandecapelle et al., 2020</td>
<td>EEG</td>
<td>Non-linear (CNN:D, CNN:S+D, CNN:S)</td>
<td>16</td>
<td>Amp Env</td>
<td>–</td>
<td>87% (CNN:S+D), 78% (CNN:D), 70.5% (CNN:S) [subject specific]</td>
<td>10 s</td>
</tr>
<tr>
<td>Ding and Simon, 2012</td>
<td>MEG</td>
<td>Linear (SR)</td>
<td>20</td>
<td>Amp Env</td>
<td>( r = \sim 0.2 )</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fiedler et al., 2017</td>
<td>EEG (+ in-Ear-EEG)</td>
<td>Linear (forward)</td>
<td>7</td>
<td>Amp Env</td>
<td>( r = 0.04 )</td>
<td>70%</td>
<td>60 s</td>
</tr>
<tr>
<td>Fuglsang et al., 2017</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>26</td>
<td>Amp Env</td>
<td>( r = \sim 0.07 )</td>
<td>87.1%</td>
<td>40–50 s</td>
</tr>
<tr>
<td>Mesgarani and Chang, 2012</td>
<td>ECoG</td>
<td>Linear (SR)</td>
<td>3</td>
<td>Amp Env</td>
<td>( r = \sim 0.60 )</td>
<td>93.0%</td>
<td>NC</td>
</tr>
<tr>
<td>Miran et al., 2018a,b</td>
<td>EEG and MEG</td>
<td>Linear (SSM)</td>
<td>3</td>
<td>Amp Env</td>
<td>( r = \sim 0.3 ) (EEG)–9 (MEG)</td>
<td>70% (MEG data), 80% (EEG data)</td>
<td>1.5 s</td>
</tr>
<tr>
<td>Mirkovic et al., 2015</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>12</td>
<td>Amp Env</td>
<td>( r = 0.054 ) (Subject-specific decoder)</td>
<td>88.02%</td>
<td>–</td>
</tr>
<tr>
<td>O’Sullivan et al., 2015</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>40</td>
<td>Amp Env</td>
<td>( r = 0.054 ) (Subject-specific decoder)</td>
<td>89% [Subject-specific]</td>
<td>60 s</td>
</tr>
<tr>
<td>O’Sullivan et al., 2017</td>
<td>ECoG</td>
<td>Non-linear (DNN)</td>
<td>6</td>
<td>Spec</td>
<td>( r = \sim 0.4 ) (Attended speaker)</td>
<td>&gt;70% (3 Subjects)</td>
<td>15 s</td>
</tr>
<tr>
<td>Pasley et al., 2012</td>
<td>ECoG</td>
<td>Linear (SR) and Non-linear ()</td>
<td>15</td>
<td>Spec</td>
<td>( r = 0.2–0.3 )</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Schäfer et al., 2018</td>
<td>EEG</td>
<td>Linear (SR)</td>
<td>10</td>
<td>Amp Env</td>
<td>–</td>
<td>61.1%</td>
<td>30 s</td>
</tr>
<tr>
<td>Vandecapelle et al., 2020</td>
<td>EEG</td>
<td>Non-linear (CNN)</td>
<td>16</td>
<td>Pre-processed EEG signal</td>
<td>–</td>
<td>85.1–80.8% [Subject specific]</td>
<td>10–1 s</td>
</tr>
<tr>
<td>Zion Golumbic et al., 2013</td>
<td>ECoG</td>
<td>Linear (SR)</td>
<td>6</td>
<td>Amp Env</td>
<td>( r = \sim 0.15 )</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Articles are sorted by alphabetical order. The method row indicates the type of model used in the article. Amp Env, Amplitude Envelope; Spec, spectrogram; AV, Average; SR, Stimulus Reconstruction; SSM, State-Space Model; DNN, Deep Neural Network; CNN, Convolutional Neural Network; CCA, Canonical Correlation Analysis; s, second.

Analysis (CCA) and Bayesian state-space modeling. Canonical Correlation Analysis is a hybrid model that combines a decoding and an encoding model. This approach, developed by de Cheveigné et al. (2018), aims to minimize the irrelevant variance in both neural data and stimulus by a linear transformation. Concerning Bayesian state-space modeling (Miran et al., 2018b), it is composed of three modules: a dynamic encoder/decoder estimation module, an attention marker extraction module, and a real-time state-space estimator module (see Miran et al., 2018a for a complete description of the model) and this approach was developed in the purpose of real-time decoding of auditory attention.

As mentioned before, in the context of AAD, linear models are generally used with two (or more) concurrent speech streams in order to determine which stream the listener is attending to. In this case, a representation of each auditory source is
created (e.g., speaker 1 and speaker 2). Once the model has been fitted, no matter which approach was chosen, a two-class classifier is used to decide which of the two streams the participant was focused on. To do so, the classifier compares the correlation coefficients between the model output and the original model input representations (e.g., the correlation between the reconstructed envelope and the original audio signals envelopes in backward strategy) over a certain portion of data (decision time windows). The highest correlation indicates which stream the participant was attending to. The length of the decision time window is a crucial parameter because correlation-based measures need a certain amount of information to perform well. However, short decision time windows (<2 s of data) are of interest in BCI for real-time classification.

Generally, AAD performances are assessed with two accuracy metrics: regression accuracy and classification accuracy (Wong et al., 2018). Regression accuracy evaluates the goodness of fit of the model and it is expressed in terms of correlation coefficient (Pearson’s correlation, often ranging 0.1–0.2) between the output of the model and the real value (e.g., speech envelope is correlated with reconstructed envelope for backward models). Classification accuracy, on the other hand, evaluates the ability of the classifier to correctly identify the attended stream for a given decision time window and it is generally expressed in terms of percentage of good classification. Classification accuracy is generally high for long decision time windows (around 85% for 60 s of data) but drops drastically for shorter decision time windows no matter which approach is used.

Recently, Wong et al. (2018) showed that decoding models outperform encoding models in terms of classification accuracies. One of the best classification results obtained so far was 85% with 20-s decision time windows, with the CCA (Geirnaert et al., 2020).

Non-linear Models

Similarly to linear models, several non-linear model architectures are in competition. But non-linear models are still overlooked because they are more complex to implement and interpret. Nevertheless, they were used by a few studies to explore AAD. Vandecapelle et al. (2020) used two convolutional neural networks to determine the attended speaker in a multi-speaker scene by using the direction of the locus of auditory attention. Their method allows them to decode auditory attention with very short decision time windows and with a good classification accuracy (around 80% for 2 s of data). In another study the authors used a fully-connected neural network to reconstruct the speech envelope and estimate the attended speaker (de Tailliez et al., 2017). The classification accuracy obtained with this method appears to be similar to the performance obtained in Vandecapelle et al. (2020) even though the comparison between studies is not straightforward due to differences in experimental and model parameters or accuracy measures (Ciccarelli et al., 2019). However, non-linear models outperform linear models in terms of decision time window/performance ratio. One other potential advantage of this type of model is that it seems more realistic insofar as it may capture the neuronal non-linearity underlying speech perception (O’Sullivan et al., 2015, Mirkovic et al., 2015, de Tailliez et al., 2017).

Limitations of Linear and Non-linear Models

Linear and non-linear models yet suffer from several limitations with respect to AAD. The major problem of linear models lies in the fact that their classification accuracy is strongly influenced by the duration of the decision window. Long windows yield good classification (>80%) while short ones (e.g., 2 s) yield much poorer performance (∼60%). This is due to the fact that (1) short decision windows contain less information (Vandecapelle et al., 2020), (2) EEG signals contain a mixture of several physiological and neural processes. Thus, correlations between predicted and actual data are rather weak (between 0.05 and 0.2) and short decision time windows are particularly sensitive to noise (Geirnaert et al., 2020). Moreover, a huge amount of data is needed to fit the model properly. Therefore, these models are difficult to use in real time situations where the selection of the attended speaker must be performed as fast as possible.

For non-linear models, the principal issue is the risk of overfitting, in particular with small datasets (Vandecapelle et al., 2020). Moreover, comparing performances of several non-linear models on different datasets pointed to a low reproducibility of these algorithms (Geirnaert et al., 2020). Besides fitting issues and physiological noise (and non-relevant neural signal), another source of performance variability resides in inter-individual differences at the cognitive level, such as for instance in working memory (WM) (Ciccarelli et al., 2019), attentional control, cognitive inhibition, but also motivation.

FUTURES PLAUSIBLE APPLICATIONS FOR AUDITORY ATTENTION DETECTION METHODS

Plausible Applications for AAD-Passive Brain Computer Interfaces Systems

Classical active Brain Computer Interfaces (aBCI) exploit the user’s voluntary brain activity to control applications or devices. Several years ago, a new category of BCI, named passive Brain Computer Interfaces (pBCI), emerged. Unlike aBCI, pBCI use involuntary brain activity (e.g., cognitive state) to implicitly modify human-machine interactions (Zander and Kothe, 2011; Clerc et al., 2016). passive Brain Computer Interfaces are generally used to monitor attention, fatigue, or workload in real life situations such as driving situations (Haufe et al., 2014) or air traffic control (Aricò et al., 2016) but they can also be used in less operational contexts. For example, pBCI can be used to provide translation of unknown read words (Hyrskykari, 2006) or to display information on the screen when the user needs it (Jacob, 1990). passive Brain Computer Interfaces also have applications in the field of virtual reality and video gaming (Lécuyer et al., 2008; George and Lécuyer, 2010).

Auditory attention detection algorithms could be coupled with passive BCI to extend the usefulness of such methods to more concrete applications. In the next section, we will describe some possible future applications for AAD-pBCI systems.
AAD-pBCI in Education
Since a few years, studies that explore the relationship between children's attention abilities and screen access have shown that precocious screen access may go along with attentional problems (Christakis et al., 2004; Ponti et al., 2017; Tamana et al., 2019, but see Kostyrka-Allchorne et al., 2017 for a systematic review on the relationship between television exposure and children's cognition). AAD-pBCI systems could be used to improve children's attention ability. Such an attempt was made by Cho et al. (2002) who developed an attention enhancement system for ADHD children using EEG biofeedback and a virtual classroom environment. They showed that it is possible to use pBCI to enhance attention in children with ADHD in a school context. An advantage of real-time AAD applications is that they may allow monitoring children's attention. Moreover, they could be of use in serious game applications aiming at enhancing sustained auditory attention (see for instance Figure 1). Importantly, one can hypothesize that, because sustained attention in a complex auditory scene requires segregation and integration abilities but also inhibition and WM, these functions may also benefit from such applications.

Such a tool could also benefit musicians who must be able to sustain attention for long periods of time (Bergman Nutley et al., 2014). Interestingly, for musicians, this approach could also enhance the ability to share auditory attention across multiple sources, since this is of great importance in ensemble music making. As for the Sustained Attention Enhancement AAD-pBCI System mentioned above, a Divided Attention Enhancement AAD-pBCI System could also take the form of a musical serious game wherein the player has to learn to switch the focus of attention from one source to another and to share attention across multiple sources.

AAD-pBCI in Art
In the field of art, several attempts have been made to bridge EEG and BCI since the 1970s (Vidal, 1973; Rosenboom, 1977; Williams and Miranda, 2018). More recently, works have been done to develop systems to control an instrument (Arslan et al., 2006) or to generate melodies with brain signals (Wu et al., 2010;
FIGURE 2 | Schematic representation of a real-time sound modulation AAD-pBCI system. Based on the real-time EEG data recording, the attended auditory source is continuously tracked by the AAD device. The pBCI device analyses in real-time the user’s intentions (e.g., moving the attended source from the upper left loudspeaker to the bottom left one), translates it into commands and sends it to an external device that will modify the loudspeaker’s parameters accordingly.
Miranda et al., 2011) to name a few. In this sense, there is a place for AAD-pBCI systems to create new kinds of art performances in which brain activity induced by auditory attention could be used to modulate different sound sources (see Figure 2). This could be of particular interest in an immersive listening structure composed of multiple loudspeakers (Pascal, 2020). Such a device would allow the user to select a specific sound source and modify its loudness, spatial location, or motion. In such a setup, the AAD module monitors in real-time the attended source and provides information about the source of interest to the pBCI module. This second module is responsible for analyzing the intentions of the user, translating them into command, and controlling an external device. To do so, the pBCI module classifies among several classes of neural activity induced by different cognitive processes (e.g., imagining a movement of the attended source). Once the user's intention has been detected, the pBCI module translates it into commands that correspond to a particular parameter's modification (e.g., moving the attended source from the upper central loudspeaker to the bottom central one) and sends them to an external device.

**Application in Neuro-Steered Hearing Aids**
The first reason why AAD has been investigated is to enhance hearing aids and more specifically, CI. Cochlear implant are electronic devices that allow deaf people to partly regain hearing aids and more specifically, CI. Cochlear implant are stimulating the auditory nerve. While they perform well when the user is facing a unique speaker (or in quiet environment), in presence of multiple speakers performance drops dramatically because all speakers are amplified indistinctly (e.g., Zeng et al., 2008).

The solution to bypass this limitation is to inform hearing aids of the user's attentional focus. In fact, if the hearing aid was able to “know” which audio source the user is attending to, then it should be able to selectively enhance it. Therefore, combining AAD algorithms and hearing aids technologies, should lead to next-generation hearing aids allowing good performances in complex (or noisy) auditory environments (see for example: Das et al., 2016, 2020; Van Eyndhoven et al., 2017; Cantisani et al., 2020; Geirnaert et al., 2020).

**Other Plausible Applications for AAD-Passive BCI Systems**
One can think about other futuristic applications for AAD, in several distinct domains. For instance, in the entertainment field. It is, for example, possible to develop “auditory games” in which players, equipped with light AAD-pBCI systems, confront each other in musical battles using their auditory attention. In addition to being fun, this kind of game could be interesting to develop cognitive abilities that underlie auditory sustained attention (WM, executive control, etc.) even if it is not its main purpose. Furthermore, such a game could be adapted to a solo or a multiplayer environment.

AAD-pBCI systems could also find applications in the field of domotics. Indeed, a wearable AAD-pBCI system could be useful, in situations where ambient noise is varying constantly (e.g., in a living room), to monitor and adapt in real-time the loudness of the attended sound source (TV, hifi system, home phone, etc.).

**CONCLUSION**
Overall, AAD, by providing real-time cues of the auditory attentional state of an individual, opens new avenues to several applications. After a first stage of fundamental research to understand the links between auditory attention and neural signals, we are now in a second stage of applied research optimizing algorithms in terms of both classification performance and speed. In the next few years, when real-time decoding limitations will be overcome and wearable wireless systems will be developed, AAD could find applications in many domains such as education, art, health, or even domotics and online games.

**AUTHOR CONTRIBUTIONS**
JB and DS conceived the present article. JB wrote the manuscript and designed the figures with the support of DS and MC. All authors contributed to the article and approved the submitted version.

**FUNDING**
JB is funded by Oticon Medical (research agreement n°14294-CRI04) and Region Provence-Alpes-Côte d’Azur. DS is supported by APA foundation (RD-2016-9), ANR-16-CE28-0012-01 (RALP), ANR-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

**REFERENCES**


Bergman Nutley, S., Darki, F., and Klingberg, T. (2014). Music practice and designed the figures with the support of DS and MC. All authors contributed to the article and approved the submitted version.

**AUTHOR CONTRIBUTIONS**
JB and DS conceived the present article. JB wrote the manuscript and designed the figures with the support of DS and MC. All authors contributed to the article and approved the submitted version.

**FUNDING**
JB is funded by Oticon Medical (research agreement n°14294-CRI04) and Region Provence-Alpes-Côte d’Azur. DS is supported by APA foundation (RD-2016-9), ANR-16-CE28-0012-01 (RALP), ANR-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), and the Excellence Initiative of Aix-Marseille University (A*MIDEX).


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Belo, Clerc and Schön. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.