KU LEUVEN

FACULTEIT INGENIEURSWETENSCHAPPEN

EEG-based Auditory Attention Detection with Neurofeedback Training at Home

Stijn Proesmans

Thesis submitted for the degree of
Master of Science in
Biomedical Engineering

Thesis supervisor:
Prof. dr. ir. Sabine Van Huffel

Assessors:
Prof. dr. ir. M. Van Hulle
Dr. ir. B. Hunyadi

Mentor:
Rob Zink

Academic year 2016 – 2017

Preface

This thesis is the final work for my Master's degree in Biomedical Engineering at the KU Leuven, and is entitled "EEG-based Auditory Attention Detection with Neurofeedback Training at Home". It reports on the research that I carried out from September 2016 to May 2017.

I am very grateful that I was granted the opportunity to perform research in the field of Neurotechnology, which I am passionate about. I could not have been more pleased with the topic of my research, which not only involved a data analysis part, but also an experimental part during which I collected data from human subjects. Moreover, my research objectives were in direct relationship with the improvement of the quality of human life in the future, which made every new-read article, every tested hypothesis, even every time cleaning the experimental equipment, worthwhile.

The outcome of this thesis would not have been the same without the support of others. First of all, I would like to thank my mentor Rob Zink for his excellent guidance. Thanks to his vast knowledge in the field and his pleasant personality, meetings with him were both productive and enjoyable. Secondly, I would like to express my gratitude to my supervisor Sabine Van Huffel, and to both assessors Marc Van Hulle and Borbála Hunyadi, for evaluating my work and for their valuable feedback at the mid-term evaluation. Finally, I would like to thank my family and friends, and in particular my close friend Fien and my sister Charlotte, for their mental support throughout this journey.

Stijn Proesmans

Contents

Abstract

EEG-based auditory attention detection (AAD) is possible in a controlled two-speaker scenario, which opens doors for the development of a brain-computer interface (BCI) controlled by auditory attention in real-life. The success of a BCI depends on the "computer" and on the "brain" of the subject using the BCI. In the current study, we address the influence of the subject on AAD, which has so far been left unaddressed.

In the first part of this study, we assess whether subjects can be trained to improve their AAD performance by being provided with online neurofeedback (NFB) about ongoing performance. NFB is thereby intended to facilitate subjects in self-regulating own brain activity. We evaluated the effects on AAD of one week of so-called Neurofeedback Training at the subject's home. Training effects were compared between an experimental subject and a control subject, to investigate the specific effect of NFB. Our results suggest that NFB training can be effective in improving AAD performance. Training would mostly facilitate the subject in eliciting more stable neural responses to the auditory attention task, which results in a more consistent detection accuracy. Learning effects were found to increase throughout training sessions and partly remain after training, but decrease over time. These results are promising for further study on a larger population. Furthermore, we show that AAD and NFB training are feasible at the subject's home, which is an important step towards the usage of AAD in real-life.

In the second part of this study, we look for subject-specific characteristics that underlie the inter-subject performance variability that has been observed for AAD. We evaluated the relationship between a number of neurophysiological markers and AAD performance, by analyzing the EEG data collected in a previous AAD study. We identified theta power in centro-parietal areas, and frontal and central midline theta, as potential neurophysiological predictors of AAD performance. Future studies should investigate which personality traits underlie these theta power differences in AAD and how they affect performance. Assessing a subject's personality would eventually allow to customize AAD and training to the individual subject.

List of Figures and Tables

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

AAD	Auditory Attention Detection
BCI	Brain-Computer Interface
CS	Calibration Session
EEG	Electroencephalography
FUS	Follow Up Session
HRTF	Head-Related Transfer Function
ITR	Information Transfer Rate
MEG	Magnetoencephalography
MI	Mental (or Motor) Imagery
MCQ	Multiple-Choice Questions
MSE	Mean-Squared Error
NFB	Neurofeedback
SC	Control Subject
SN	Neurofeedback Subject
TS	Training Session

Symbols

CA	Correlation between reconstructed envelope and input attended envelope (Pearson)
CUA	Correlation between reconstructed envelope and input unattended envelope (Pearson)
C_{RR}	Auto-correlation matrix of the EEG data
C_{RS}	Cross-correlation vector between the attended speech envelope and the EEG data
$D(\tau, n)$	linear spatiotemporal AAD decoder
$R(t, n)$	EEG signal recorded at electrode n
$S(t)$	Input speech stream envelope
$\hat{S}(t)$	Reconstructed speech stream envelope
TL	Trial length
τ	Time lag between stimulus onset and EEG signal
p	Two-sample t-test p-value
p_w	Wilkoxon signed rank test p-value
r, P	Pearson's correlation coefficient and p-value
ρ, P	Spearman's correlation coefficient and p-value

Chapter 1

Introduction

In auditory attention detection (AAD), the goal is to detect which speaker a person is listening to, in a multi-speaker environment, based on that person's neural activity. Recent studies have shown that it is possible to detect auditory attention, in a two-speaker scenario, based on electroencephalography (EEG) recording [1]-[9]. This has opened doors for future studies on developing a brain-computer interface (BCI) that can be controlled by auditory attention in online, real-life applications. One of the most relevant applications of AAD would be in neuro-steered hearing-aids, where it could steer the hearing aid with the goal of amplifying the attended speaker's voice.

There are still many challenges that need to be overcome before and AAD-BCI could be used in real-life – such as the reduction of time and number of electrodes needed for detection, robustness to complex sound environments, interference of background mental activity and motor artefacts, only to name a few. Previous AAD studies have addressed a number of these challenges by improving signal processing and machine learning techniques [2]-[4] and assessing their performance in more realistic listening situations [2]-[9] – i.e., by focusing on the "computer" side of the BCI. However, AAD performance has been reported to vary substantially between subjects [7, 9]. "Indeed, as the name suggests, a BCI requires the interaction of two components: the subject's brain and the computer" [36]. The subject has to produce EEG patterns that can be detected by the computer, which employs signal processing and machine learning techniques to do so. "For reliable BCI performance during real-life applications, these EEG patterns need to be clear, stable and distinct. Self-regulation of brain activity is a skill that can and must be learned by the user in order to get control of the BCI" [36]. So far, the influence of the subject on AAD has not been addressed. In the current study, the main focus is on the subject – i.e., the "brain" side of the BCI. We address the influence of the subject on AAD performance in two ways.

In the first part of this study, we assess whether subjects can be trained to improve their AAD performance by being provided with online neurofeedback (NFB) about ongoing performance. NFB is thereby intended to facilitate subjects in self-regulating own brain activity. So-called Neurofeedback Training has been proven to be a successful method in facilitating self-regulation of neural activity in previous studies. It has been widely applied as a therapeutic or cognitive enhancement tool, where self-regulation is intended to improve some behavioral function [27]-[29]. Moreover, it has been used as a method to improve control of many BCIs such as those based on mental-imagery (MI) tasks [30]-[33]. To our best knowledge, our study is the first to evaluate the effects of extensive Neurofeedback Training on AAD performance. We thereby assess whether AAD and training can be successful at the subject's home, which would be an important step towards the usage of AAD in real-life.

In the second part of this study, we look for subject-specific characteristics that cause subjects to vary in AAD performance. Inter-subject performance variability has been observed with BCIs based on mental-imagery (MI) tasks, which has led the BCI community to look for predictors of MI-BCI performance – "i.e., individual characteristics that correlate with the command classification accuracy" [36]. For AAD, it is not entirely known which are the individual characteristics that impact performance and could thus explain the inter-subject variability. Identifying such predictors would allow to optimize AAD performance by customizing the design of both BCI and training to the individual subject [36].

In this introductory chapter, we will first elaborate on the topic of auditory attention detection (AAD), in Section 1.1. In order to understand how auditory attention can be detected based on neural activity, we will provide the reader with relevant knowledge about the neural representation of speech and selective auditory attention. Furthermore, this section will give an overview of recent AAD developments and current state-of-the-art. Next, in Section 1.2 we will introduce the reader to the concept of Neurofeedback Training and illustrate how it has been proven successful as a BCI training method. In Section 1.3, we will discuss literature on predictors of BCI performance. We will end this chapter by summarizing the goals of our study and how they are to be achieved, and by presenting the structure of the thesis text, in Section 1.4.

1.1 Auditory Attention Detection (AAD)

1.1.1 The Cocktail Party Problem and its Neural Underpinnings

Humans show the remarkable ability to attend to the voice of one speaker in an environment where there are multiple speakers at the same time. It is still largely unknown how the brain solves this problem, which was first described as the "cocktail party problem" by Cherry in 1953 [10].

In contrast to the visual system, where the neural mechanisms of scene analysis and visual selective attention have been widely explored, little is known about these topics in the auditory system. However, significant progress has been made since research showed that cortical oscillations phase-lock to the envelope of speech [12]. Furthermore, it was found that attended speech is more represented in the brain than unattended speech, which is suppressed, especially at a higher cognitive level [13, 14]. In a review study of different magnetoencephalography (MEG) experiments, Simon et al. [15] provided proof for the assumption of neurally encoded auditory objects. This study also showed that lower order auditory areas represent both attended speech and background noise, while higher areas preferentially represent the attended speech stream.

1.1.2 Detecting Auditory Attention

Based on these findings, several studies have proven that it is possible to detect attention in a controlled two-speaker scenario, where the task is to attend to one speaker and ignore the other ("cocktail-party listening task"). The problem has been approached in a number of different ways. One approach is to extract important features from the cross-correlation of neural response and speech envelope to train a linear classifier [16]. Another approach is to build a forward regression model from sound to the neural responses [17, 18].

So far the most promising and profoundly studied approach is the Stimulus Reconstruction Approach, which is based on backward modelling. The input stimulus is reconstructed by mapping in the reverse direction – i.e., from the neural data back to the stimulus. Stimulus reconstruction has been shown to be very sensitive to selective attention. Speech spectrograms reconstructed based on multi-electrode surface recordings were dominated by the key temporal and spectral features of the target speaker [13]. Speech envelopes reconstructed from MEG data typically correlated more strongly with the attended speech than the unattended [14]. These results provide an important insight into the neural representation of speech and neural mechanisms of selective auditory attention. However, for studies on large populations and for use in real-life BCI applications, methods based on surface recordings or MEG are not suitable. The former is invasive, while the latter carries a high cost and lack of portability.

1.1.3 EEG-based Stimulus Reconstruction

O'Sullivan et al. [1] were the first to decode attentional selection with the stimulus-reconstruction approach, based on electroencephalography (EEG) recordings. During training phase, a linear spatiotemporal decoder was constructed that maps from the multi-electrode EEG recordings, and time-shifted versions of these recordings, to the envelope of the attended speech stream. The decoder was optimized by minimizing the mean-squared error (MSE) between the reconstructed stream and the actual attended input stream. During testing, the decoder was employed to reconstruct the envelope of the attended speech stream. This reconstruction was then compared with both input stream envelopes by assessing correlation. Attention was correctly detected if the reconstructed envelope was more correlated with the attended than with the unattended stream envelope. Trials of 60 s were used to make the detection decision and data was measured from 128 electrodes. Similarly to the decoder that estimated the attended speech envelope – i.e., the attended decoder –, a second decoder was constructed which estimated the unattended speech envelope – i.e., the unattended decoder. Figure 1.1 illustrates the decoding strategy.

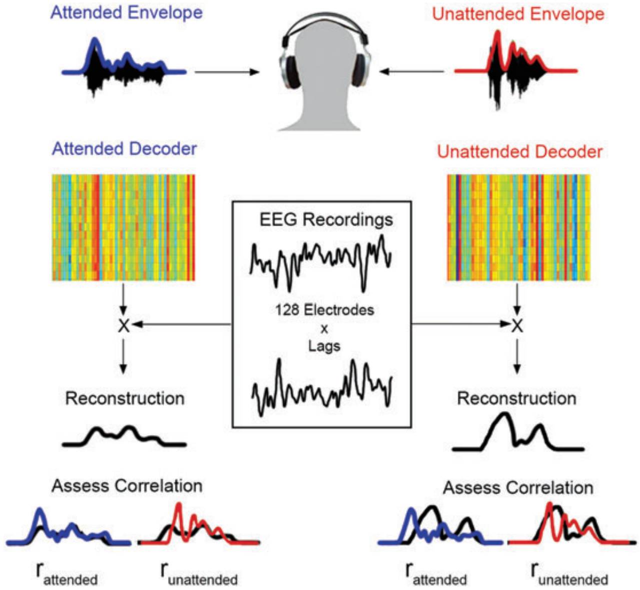


Figure 1.1: Illustration of the decoding strategy in O'Sullivan et al. [1]. The (un)attended speech envelope is reconstructed from EEG recordings at different electrodes and time-shifted versions of these recordings. Correlations between the reconstructed envelope and the actual attended and unattended envelope are assessed.

This study proved that EEG-based AAD is possible, with (attended) decoder performance ranging from 82% to 89% across subjects. Furthermore, by constructing decoders on individual time-shifted versions of the EEG-recordings, they identified that neural processing at approx. 200 ms must be critical for solving the cocktail party problem. These results not only provided important insights into human cognition, they opened doors for developing an EEG-based BCI that could detect auditory attention in real-life applications.

Several studies repeated the experimental paradigm presented by O'Sullivan et al., while slightly adapting it, working towards real-life applications. Mirkovich et al. [2] implemented an offline iterative procedure to eliminate the least contributing channels, thereby reducing the amount of channels needed for successful detection. Secondly, they studied the effect of training data duration on decoding performance. In [3], Das et al. created a more realistic listening scenario, by head-related filtering of the presented audio stimulus, and assessed its effect on decoder performance. Besides, they compared performance of "ear-specific decoders" – i.e., decoders trained and tested only on one ear (left or right) – with decoders trained and tested on both ears. Biesmans et al. [4] included knowledge of the auditory periphery into the speech envelope extraction and compared different envelope extraction methods. Furthermore, they introduced a more natural way of combining recordings and trained the decoder in an "all-at-once" way, which reduced sensitivity to regulation parameters. Aroudi et al. [5] assessed the influence of using noisy – and thus more realistic – speech signals as reference signals when making the detection decision.

AAD detection appeared to be robust to white and speech-shaped noise. In [6], performance of a concealed multichannel around-the-ear EEG system was evaluated. The system allowed identification of the attended speaker above chance-level. Van Eyndhoven et al. [7] combined the AAD detection with speech segregation and noise suppression algorithms – two technologies that will make up the future neuro-steered hearing aids. They extracted the attended speaker from microphone array recordings of noisy speech mixtures based on the outcome of EEG-based AAD, and showed that this extraction was feasible and robust. In [8], Fuglsang et al. used advanced acoustic simulations to recreate real-world acoustic scenes in the laboratory, with varying amounts of reverberation and number of interfering talkers. Across the different listening environments, they found that the attended speaker could be accurately decoded irrespective of the different distortions in the acoustic input.

The study that is most relevant for the current study was performed by Zink et al. [9]. They implemented a fully-automated closed-loop system that could perform the AAD in real time, with trials of 10 s, and which allowed to give visual feedback about the subject's ongoing performance. Furthermore, they brought the experiment outside of the lab by using mobile EEG hardware. By proving the feasibility and performance of their experimental set-up, they paved the way for studies investigating the effect of neurofeedback on AAD performance.

1.2 Neurofeedback Training

1.2.1 Concept

Neurofeedback (NFB) training has been defined as a method to self-regulate one's own brain activity [20]-[22]. In a closed-loop BCI application, neural activity is measured, and a sensory representation – whether visual, auditory or tactile – is fed back to the user, in real-time, to facilitate self-regulation of neural activity.

"It is well established that people can learn to take control of various features of their neural activity through a training process that involves the online display of ongoing changes in the EEG" [19]. NFB training thereby engages several learning mechanisms in the brain of which operant conditioning is considered as the main one [19, 20]. For an in-depth review of the theories and models that have been proposed to explain the neural mechanisms of neurofeedback learning, we kindly refer the reader to [22].

Neurofeedback training has been applied as a therapeutic tool to normalize deviating brain activity; or as a cognitive enhancement tool for healthy subjects in so-called "peak-performance training" [20]. Examples are the use of NFB training in neurorehabilitation for motor learning in post-stroke recovery [24]-[26], in treatment of attention deficit hyperactivity disorder (ADHD) [27] or epilepsy [28]; or NFB training for improving attentional abilities [29]. These studies are based on (the assumption of) a causal relationship between specific brain oscillations and cognition, behavior or motor function – that is, self-regulation of specific brain activity is intended to produce beneficial changes in cognitive, behavioral or motor function.

Effectiveness of NFB training can then be measured using two independent variables: (1) changes in EEG activity and (2) cognitive or behavioral changes of a targeted function [21].

1.2.2 Implications for BCI Training

In the current study, NFB training is also used as a method to mediate self-regulation of brain activity. However, self-regulation is not intended to induce any beneficial changes in behavior – that is, increased auditory performance in adverse listening situations. The intention here is for the subject to gain control over the AAD-BCI by learning how to produce stable, clear and distinct EEG patterns that are correctly interpret by the AAD decoder. A sensory representation of decoder performance assists the subject during this so-called “subject learning” process. BCI-control training not only involves the subject; the decoder is updated throughout the training process to include more and more subject data and adapt to the statistical nature of evoked brain states – i.e., “machine learning”.

Related research has shown that EEG-based BCIs can be controlled by motor-imagery tasks, where the user imagines the movement of specific body parts to command the BCI. Within only a few days of training, discrimination of two brain states (e.g., left- versus right-hand movement imagination) could be reached and classification accuracy was nearly 100% [31]. The standard “Graz-BCI” protocol [31]-[33] has been adopted by many studies, where NFB was either developed specifically for training (e.g. a cursor on a computer screen) or provided by the responses of the device that was controlled (e.g. a robotic arm). In a study by Zich et al. [24], 16 participants were trained in the motor imagery (MI) task over four consecutive days. They found that online MI-BCI accuracy increased significantly, caused by the underlying brain activity that had changed (most importantly, a reduction in ipsilateral event-related desynchronization of sensory motor rhythms). Another study explored functional and structural changes after four weeks of MI NFB training [25]. Besides reporting increased classification accuracies and effectively changed EEG activity (stronger contralateral EEG), they provided a proof-of-concept for a low-density mobile EEG system for efficient NFB training at the participant’s home.

For the AAD-BCI, there is no standard training protocol, nor a reference study. It is thus not known how and to what extent neural responses to the auditory attention task can be self-regulated. However, it is interesting to mention that NFB training has been able to facilitate self-regulation of activity from human auditory areas. In [34], they showed that real-time functional magnetic resonance imaging (fMRI)-based feedback of regional cortical activity from the auditory area enabled a group of individuals to increase the level of activation during a selective auditory attention task. Another study assessed the effects of slow cortical potential (SCP) NFB training on one patient with chronic tinnitus [35]. They revealed close to normal changes of resting state activity in cortical areas that are considered to be involved in tinnitus generation, with the patient reporting decreased tinnitus loudness and pitch after training.

1.3 Predictors of BCI performance

AAD performance has been reported to vary substantially between subject [7, 9], a finding that has not been addressed by the AAD community yet. Inter-subject performance variability has been observed with BCIs based on mental-imagery (MI) tasks, which has led the BCI community to look for predictors of MI-BCI performance – i.e., “individual characteristics that correlate with the command classification accuracy” [36]. In [37], subjects were asked to perform several MI tasks. Before the experiment, subjects’ personality and cognitive profile was assessed using psychometric questionnaires. During the experiment, different neurophysiological patterns were computed, such as EEG power in certain frequency bands. The authors explored the relationship between MI-BCI performance and a number of these psychological and neurophysiological markers. While they found no relevant relationships with neurophysiological markers, strong correlations between MI-BCI performances and mental-rotation scores (reflecting spatial abilities) were revealed.

To our best knowledge, there is no literature on predictors of AAD performance. However, previous AAD studies found a significant correlation between AAD performance and behavioral performance during the cocktail-party listening task [1, 9], which thus can be considered as a behavioral predictor of AAD performance. Furthermore, a number of neurophysiological markers have been proposed by literature to play an important role in auditory attention. In [38], they showed that modulation of alpha power during a dichotic listening task could predict accurate recall of the to-be-attended stimuli. Besides, they noted that “high overall alpha power indicates suppression of neural activity in visual areas in the occipital lobe, which might support enhanced neural function in the auditory modality” [38]. In [39], the authors suggested that beta-band oscillations in EEG are associated with selective-attention induced unmasking of target speech in a cocktail-party listening situation.

1.4 Research Goals and Structure of the Text

The first goal of our study is to assess whether Neurofeedback Training can be used as a method to enhance AAD performance, and if training (and thus AAD) is feasible in the home environment. To achieve this goal, we will evaluate the effects on AAD of one week of intensive training at the subject’s home. Training effects will be compared between an experimental subject and a control subject, to investigate the specific effect of NFB. Realization of the first goal will be reported in Chapter 2.

The second goal of our study is to identify neurophysiological predictors of AAD performance – i.e., neurophysiological markers that correlate with AAD performance, across subjects, and could thus predict a subject’s AAD performance. To achieve this goal, we will analyze the EEG data collected in a previous AAD study [9], and explore the relationship between a number of neurophysiological markers and AAD performance. Chapter 3 reports on the realization of the second goal. The thesis will be brought to a close in Chapter 4, which unifies the findings of Chapter 2 and Chapter 3.

Chapter 2

The AAD Neurofeedback Training Experiment

As we explained in the previous chapter, auditory attention can be detected, in a two-speaker listening situation, based on EEG activity. This has opened doors for developing a BCI, controlled by auditory attention, that could be used in real-life (e.g. neuro-steered hearing aids). Several studies have contributed to this end by mainly focusing on the "computer" side of the BCI. However, a large inter-subject performance variability has been reported for AAD [7, 9] which has not been addressed by the AAD community yet. In the current study, we assess whether Neurofeedback Training can be used as a method to enhance AAD performance, and if training (and thus AAD) is feasible in the home environment. In this chapter, we will evaluate the effects on AAD of one week of intensive training at the subject's home, and outline the experimental work carried out to do so.

In Section 2.1, we will discuss our materials and methods. This includes the AAD paradigm, the Neurofeedback Training Paradigm and an overview of the Experiment Timeline. In Section 2.2, we will explain how we analyzed the data to assess the effects of NFB training on AAD. Results will be presented in Section 2.3, and discussed and concluded in Section 2.4. In that section, we will also discuss implications for applications and revise some of the used methods.

Besides analyzing the effect of NFB training, the large amount of subject-specific data that was collected during our experiment allowed us to make several additional post-hoc analyses. These are not related to the effects of NFB training and are therefore treated in a separate section, Section 2.5.

2.1 Materials and Methods

2.1.1 AAD Paradigm

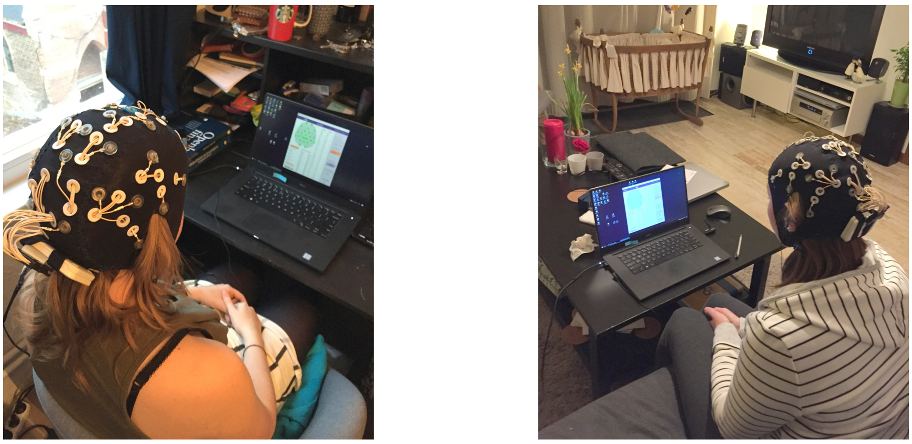
The AAD experimental paradigm was based on the original paradigm by O'Sullivan et al. [1]. Main adaptations were the application of head-related transfer functions (HRTFs) to the stimuli [3], the audio envelope extraction method and training of decoders in an "all-at-once" way [4], and the presentation of NFB during training sessions [9]. Implementation of the AAD setup was adopted from [9].

Participants

Two subjects took part in this study, aged 21 and 26, both female. They were Dutch (Flemish) native speakers and reported normal hearing and normal vision (with or without glasses). Before and after the experiment, subjects filled out the SSQ12 test [40], which has been proposed by literature for self-assessing auditory performance in various listening situations [40].

The experiment consisted of eight AAD sessions, divided into a decoder calibration phase, an AAD training phase and a follow up phase (as we will further explain in Topic 2.1.3). During training phase, one participant, further referred to as Subject N (SN), was presented with NFB about ongoing AAD performance. The other participant, Subject C (SC), did not receive any NFB throughout the entire study, but participated in as many sessions as SN. Sessions took place at the subject's home, in a comfortable seated position at a desk or in the living room. Figure 2.1 shows an illustration of the recording environment for both subjects. Note that an additional subject participated in a short (~30 min) pilot experiment in which AAD setup and some NFB features were tested.

Before engaging in the experiments, all participants signed an informed consent which informed them about the experiment's procedure, goals and eventual discomforts (see Appendix A). Our study was approved by the ethics committee of the KU Leuven.



(a) SC (b) SN

Figure 2.1: Illustration of the recording environment for a) SC and b) SN.)

Behavioral Task

During a session, subjects were presented, during twenty-four minutes, with two different fiction stories at the same time. Stories were played in both ears by means of earphones. One story was more prominent in the left ear, while the other more in the right ear; as if one speaker was positioned at 90° to the left of the subject, while the other speaker at 90° to the right. Subjects were instructed to attend to the story on either left or right side, while ignoring the story on the other side. They were asked to maintain visual fixation on a sphere situated at the center of the computer screen, and to minimize motor activity.

Figure 2.2 shows an overview of the design of an AAD session. The twenty-four minute presentation was divided into four blocks of six minutes. After each block, there was a short (~3 min) break during which the subjects were given a questionnaire (see Appendix A) containing four multiple-choice questions (MCQ) about the story fragment they just attended to. Intention was to motivate subjects to execute the task well, and to keep track of behavioral performance. Subjects also indicated task difficulty (i.e., difficulty of concentrating on one story while ignoring the other) during the last block and their interest in the attended story fragment on a ten-point scale.

At the start of the next block, both stories continued where they left off at the end of the previous block, and subjects continued attending to the same story. Right and left ear attendance were altered within a session – i.e., during the first two blocks, subjects attended to the story on their right side; the following two blocks, the to-be-attended story continued on their left side. Intention was to avoid "overoptimistic results" as described by Das et al. [3]. They found that decoders trained and tested on only one side performed better than decoders trained and tested on both sides. However, one-side attendance is not guaranteed in real-life listening situations.

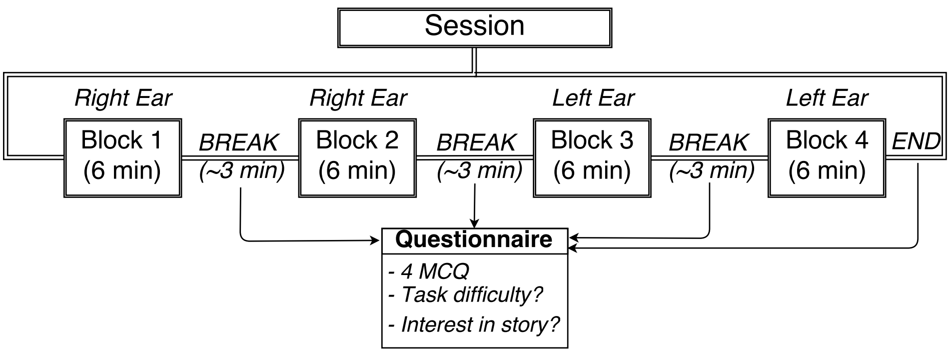


Figure 2.2: Overview of an AAD session. Each session consisted of four 6 min blocks. Ear attendance was switched from right to left at mid-session. After each block, a questionnaire was filled out (4 multiple-choice questions (MCQ), task difficulty rating, interest in story rating) during a short (~3 min) break.

Stimuli

Seventeen Dutch fiction stories, narrated by male speakers, were used as stimuli. The audio files were preprocessed before being presented as stimuli. We therefore slightly adapted an implementation provided by Das et al. [3]. Silences longer than 500 ms were truncated to 500 ms in order to prevent the subject from losing attention and switching attention to the to-be-ignored stream. Stories were then concatenated with other stories in streams, so as to get eight speech streams, each of length greater than 24 min. Thereby, each story was only used once. Streams were paired to form 4 stream pairs. Throughout the study, these pairs would each be presented twice, whereby subjects altered attention so that each stream was only attended once and ignored once. Streams were divided into 4 parts of 6 min each, each part serving as a block fragment.

Instead of presenting a stream pair in a dichotic way, by separately playing one stream in each ear, streams were filtered using head-related transfer functions (HRTFs). Result was a stereo audio stimulus in which both speech streams were played in both ears, but with different intensities and delays. Thereby, positions of the speakers were simulated at 90° to the left and right of the subject. Compared to the dichotic listening condition, this simulated a more realistic listening scenario and would lead to better AAD performance, as was showed by Das et al. [3]. Stream amplitudes were normalized to have the same root mean square (RMS) intensity and the stimulus was presented to the subject through low-cost consumer earphones (Sennheiser MX 475).

Data Acquisition

The data-acquisition setup was adopted from [9]. While subjects performed the behavioral task, their EEG was measured using 24 Ag/AgCl passive scalp electrodes (Easycap, Herrsching, Germany) placed according to the 10-20 standard system (see Appendix A). Electrode impedances were kept below 10 kOhm by using an abrasive electrolyte gel that was applied to each electrode. The electrodes were connected with a SMARTING mobile 24-channel EEG amplifier (mBrainTrain, Belgrade, Serbia), which wirelessly (via Bluetooth) transmitted the EEG data to a notebook computer. The computer collected the data through Openvibe, where it was stored for offline analysis, and simultaneously streamed to Matlab via the labstreaminglayer interface (LSL). Every 10 seconds, a chunk of data comprising the last 10 seconds of 24-channel EEG data was streamed to Matlab where it was analyzed online. The audio stimuli were played via Openvibe and were synced with the EEG recordings via the Openvibe audio triggers. The stimuli were also preloaded into Matlab for the online analysis part. All further data analysis (online and offline) was done in Matlab.

Data Preprocessing

Preprocessing of EEG and audio data was adopted from [9]. EEG data was bandpass filtered between 1 and 8 Hz (Butterworth) and down-sampled to 20 Hz (during the online analysis, this was done every 10 s for the new incoming 10 s chunk of data). As a reference for the attention detection, the noise-free unfiltered speech streams were used. Speech envelopes were obtained by taking the absolute value of the audio waveforms with power-law compression with exponential 0.6. These were then low-pass filtered (Butterworth) with a cut-off frequency of 8 Hz. In [4], various audio envelope extraction methods were compared, and it was found that adding a non-linear power law amplitude compression to the envelope extraction significantly improved AAD performance. We did not investigate optimality of data preprocessing techniques, as it was not the focus of our study. All methods used have been proven successful in performing AAD in previous work.

Stimulus Reconstruction

Detection of auditory attention was based on the EEG-based method of stimulus reconstruction which was first presented by O'Sullivan et al. [1]. Goal is to reconstruct the envelope of the attended speech stream S based on neural recordings R via a linear spatiotemporal decoder D. In case of N electrodes, $R(t,n)$ represents the EEG signal recorded at electrode n. A decoder $D(\tau,n)$ maps from $R(t,n)$ to the attended speech stream envelope $S(t)$ as follows:

$$\hat{S}(t) = \sum_{n} \sum_{\tau} D(\tau,n) R(t+\tau,n) \tag{2.1}$$

where $\hat{S}(t)$ denotes the estimated envelope of the attended speech stream and τ represents a time lag. $\hat{S}(t)$ is thus estimated from the EEG recordings at all N electrodes and time-lagged versions of those recordings. This way, information in the EEG at later time points is included in estimating the speech stream envelope at earlier time points. The decoder D is computed that minimizes the mean-squared error between the actual and reconstructed envelope. Analytically, D is found by:

$$D = C_{RR}^{-1} C_{RS} \tag{2.2}$$

where C_{RR} is the auto-correlation matrix of the EEG data (across all electrodes and time lags), and C_{RS} is the cross-correlation vector between the attended speech envelope and the EEG data (across all electrodes and time lags).

Trial Length

In determining ideal trial length in an online NFB application, one must consider a trade-off between temporal resolution and the accuracy of the given NFB. The shorter the trial length, the more real-time the feedback could be provided, however, the less accurate the NFB would represent the actual performance of the subject (because of lesser decoder accuracy). In [9], trial length of 10 s resulted in reasonable decoder accuracies (81.9% on average) while still allowing for near-real-time feedback. Trial length of 10 s was therefore used throughout our experiment as well, for training as well as testing the decoders.

Decoder Training and Attention Detection

Decoders were trained based on the stimulus reconstruction method described hereabove, and in the same way as in [9]: EEG recordings as well as corresponding speech envelopes were divided into trials of 10 s. EEG trials were then time-shifted in a 0 - 300 ms time lag interval with shifts of 50 ms. The resulting 7 versions of an EEG data trial composed the R matrix as in Formula 2.1. The corresponding attended speech envelope trial composed the S vector. For each training trial, C_{RS} and C_{RR} were computed. These matrices were averaged over all training trials, after which the decoder was computed with Formula 2.2. This is different from the approach in [1], where decoders were computed for each trial separately and were then averaged over training trials. According to previous studies, the current approach this yields a correlation matrix that is better conditioned than the per-trial correlation matrices [3]. Furthermore, it reduces sensitivity to a regularization parameter [4]. Averaging the correlation matrices is equivalent to training the decoder in an "all-at-once" way, in which a single decoder is optimized to minimize the mean-squared error over the entire training data set [4].

During testing, attention could be detected by applying the decoder to the EEG data of the test trial which resulted in a reconstructed speech envelope (Formula 2.1). Pearson's correlation coefficients were computed for the correlations between the reconstructed envelope and the actual attended and unattended envelopes, respectively, further referred to as CA ("correlation attended") and CUA ("correlation unattended"), respectively. Attention during a trial was correctly detected if CA was greater than CUA for that trial. CA can be seen as the "reconstruction accuracy" of the attended decoder.

Similarly to decoders that estimated the attended speech envelope – i.e., attended decoders –, decoders were constructed that estimated the unattended speech envelope – i.e., unattended decoders. The S vector in Formula 2.1 was then composed of the unattended speech envelope and the reconstructed envelope $\hat{S}$ was an estimate thereof. Decoding was correct when CUA was greater than CA, and CUA can be seen as the "reconstruction accuracy" of the unattended decoder. Unattended decoders were only computed for post-hoc analysis and were not used for providing NFB during the online application.

2.1.2 Neurofeedback Training Paradigm

Empirical data concerning an "optimal" paradigm for BCI-control training do not exist [19]. However, experiences with therapeutic NFB training can help in the design of BCI training procedures, since both types of training have the common goal of facilitating self-regulation of brain activity. Therefore, in developing our NFB training paradigm, we were guided by several studies reviewing the key aspects relevant to the development of EEG-NFB methodologies [19]-[23]. Furthermore, we were guided by two studies that each developed a MI-BCI training paradigm and successfully applied it to enhance self-regulation of neural activity and BCI control [24, 25]. Developing a NFB training paradigm included designing a NFB cue and a training protocol.

Neurofeedback Cue Design

Designing the NFB cue included deciding which measure was fed back to the subject, in what modality, and in which format.

Performance Measure In clinical practice, NFB is usually a representation of EEG power in a certain frequency band, or a ratio of power in different frequency bands. Which activity is targeted is based on evidence for an association between activity and some behavior [20]. Training has been shown to be more successful when a smaller number of frequency bands, and a bigger number of electrodes was involved in the NFB representation [21].

In BCI-control training, the goal is not the self-regulation of EEG-activity itself. Here, self-regulation implies producing brain activity that is easier to detect by the "computer" in order to increase BCI performance. To achieve this goal, NFB should thus represent this performance.

In the case of AAD, the "computer" is a pre-trained linear spatiotemporal decoder that maps from neural activity to the attended speech envelope. The decoder performs well when correlation of the reconstructed attended speech envelope is high with the actual attended speech envelope, and low with the unattended envelope – that is, when CA-CUA is high. CA-CUA is thus a measure of ongoing AAD performance and was chosen to be represented by the NFB. The subject's goal is to maximize performance by eliciting strong neural fluctuations with the attended speech stream, and weak fluctuations with the unattended – thus, by self-regulating brain activity. CA-CUA could be represented proportionally or in a binary way (e.g. CA-CUA $<$ or $>$ 0). Proportional feedback was opted as it has been suggested to be more effective [20].

Feedback Modality According to [20], there are still too few systematic studies comparing the effects of different feedback modalities for specific protocols. Therefore, feedback modality selection is often based on practical considerations.

In our study, two modalities were considered: auditory and visual. When considering future application in neuro-steered hearing aids, where auditory feedback will be given by the device (i.e., amplification of attended stream while suppression of background noise), auditory feedback could be preferred for training as well. Interesting to mention is a study that implemented automatic gain control to adjust the amplitudes of attended and unattended source with the goal of increasing signal-to-noise ratio [16]. However, because auditory feedback would interfere with the behavioral task, it would make training more complex and it would make it more difficult to assess for NFB-specific effects. In [20], they mention the importance for training efficacy of avoiding the crossover of sensory input in which the feedback signal stimulates the same brain regions as those targeted during training. Furthermore, neuro-steered hearing aids are not the only application of AAD. Robust AAD could be used to control other devices (e.g. a sound recording device) which might as well present the user with visual feedback.

Although some studies have shown effective BCI training using only auditory stimuli, visual feedback has turned out to be superior in most BCIs [19, 21]. It has therefore been used in most BCI training studies. Two examples are [24] and [25] where visual feedback successfully induced self-regulation of brain activity and increased MI-BCI control. For this and for the aforementioned reasons, we chose to present the NFB visually. However, it would be interesting for future work to implement an auditory feedback condition as well to compare both feedback modalities.

Feedback Format Changes in brain activity that reflect successful NFB training should be rewarded or positively reinforced [19]. If the represented performance measure reaches a certain threshold, a rewarding stimulus should be presented. Typical visual feedback stimuli are a moving cursor, a bar or sphere varying in size or objects changing in color [19]. In our study, we implemented a moving bar and shrinking/growing sphere which were both tested in a pilot experiment. We opted for a sphere centered at the screen to minimize eye movement artefacts.

Feedback is often presented in a game-like format, in which the task is to reach a specified goal. Such formats help to maintain the subject's motivation and attention [19]. However, the feedback signal should not distract the subject from the behavioral task to be performed. Two different formats were implemented and tested in a pilot experiment: 1) a sphere with a size proportional to the online CA-CUA difference in combination with a circle indicating the biggest sphere size achieved throughout the session, serving as a target to surpass; 2) a sphere with radius growing proportionally to the online CA-CUA difference, with the goal of maximizing sphere size throughout the session. The former mentioned format was found to highly frustrate, discourage and therefore distract the subject. Reason was the frequent and unfair presentation of negative feedback: a shrinking sphere, even though not always corresponding with incorrect attention detection, was always experienced as negative performance. The shrinking/growing sphere, on the other hand, was not reported to be discouraging or distracting and correct detection always corresponded with a rewarding signal – i.e., an increase in sphere size. We therefore chose this format.

Generally, thresholds are adjusted to keep the task challenging enough as the subject improves. In [41], they address the problem of finding the optimal task difficulty for brain self-regulation learning. Training is effective only if the task is challenging enough, however, too challenging tasks could cause the subject to feel frustrated. In therapeutic practice, thresholds are set so that the reward is received about 60 - 70% of the time [19]. In our case, since initial decoder performance was expected between 55 and 85% – i.e., CA-CUA greater than zero in 55 to 85% of trials – the initial threshold for reward was set to zero. This threshold was eventually adopted throughout the entire experiment.

A final consideration was made regarding the presentation of negative feedback. In [9], online CA-CUA was represented by a color-changing sphere. A green sphere indicated above threshold performance; a red sphere indicated CA-CUA was below threshold. It was found that, when subjects were suddenly presented with negative feedback, they were more prone to perform bad during the next trial as well. The authors ascribed it to a "negative surprise effect" which would cause lower attentional response to the behavioral task. To avoid this effect and the aforementioned frustration related with negative feedback, we decided to only present positive feedback.

To summarize: In our experiment, NFB was presented as a blue sphere positioned at the center of the computer screen. In case of above-threshold performance (CA-CUA > 0), sphere radius grew proportionally to the performance measure (CA-CUA). In case of below-threshold performance (CA-CUA < 0), the sphere did not grow nor shrink (with the subject not being explicitly informed what this meant). The experimental subject (SN) was instructed to maximize sphere size by focusing on the to-be-attended speech stream. Figure 2.3 shows an illustration of a sample NFB cue during the first four trials of a training session.

TRIAL	trial 1	trial 2	trial 3	trial 4
NFB CUE				
R	$R_1 = R_{start}$	$R_2 = R_1 + k*0.21$	$R_3 = R_2$	$R_4 = R_3 + k*0.33$
CA-CUA	0.21	-0.06	0.33	...

Figure 2.3: Sample NFB cue during the first four trials of a training session. R is sphere radius. CA-CUA is computed at the end of each trial. During trial 1, the sphere size was determined by a predefined R_{start}. From then on, R was equal to R in the previous trial augmented with the proportionally (k) weighted CA-CUA value that was computed at the end of the previous trial, if that value was > 0. Otherwise, R was equal to R in the previous trial.

Training Protocol

When designing a training protocol, important considerations are the session length, training intensity and the use of a control condition.

Session Length When determining session length, one must consider a subject's ability to focus on the training. The longer the session, the more difficult for the subject to stay focused towards the end. On the other hand, shorter sessions allow for less practice trials for control-training, and too short sessions might end just when the subject gets immersed in the task. Common session duration in therapeutic studies is about 20 - 40 min [20]. In [24], MI-training sessions of approx. 24 min facilitated the learning effect by avoiding exhaustion and boredom. We chose a session length of 24 min, which is comparable to AAD session length in previous studies [1, 2, 3, 9]. Besides, 24 min (= 144 trials) of test data per session was deemed to be sufficient for proper comparison of session performances during post-hoc data analysis.

Furthermore, it should be decided whether or not a session is interrupted by breaks, and if these breaks are self-paced or defined before the experiment. It was decided to divide the session in 4 blocks of 6 min, with a short (~3 min) break between blocks, which allowed for assessing of behavioral scores by means of a questionnaire (see Topic 2.1.1). Previous AAD studies [3, 9] used the same practice.

Training Intensity Training intensity involves the number of training sessions and the amount of rest (in days) between them. According to [20], for learning to self-regulate brain activity, little is known about the optimal amount of training sessions in a certain time interval, nor the length of an effective gap between them. Therefore, in the decision of training intensity, we were guided by successful practice in MI-BCI studies. In [24], a total of four training sessions (i.e. one per day for four consecutive days) resulted in enhanced self-regulation and MI-BCI accuracy in all 16 subjects trained. In the current study, we decided to conduct four training sessions in one week, with one day of rest between consecutive sessions. The day of rest was believed to enhance training efficiency, since sleep and rest have been found to facilitate auditory learning [42].

Control Condition To allow the assessment of the specific effects of NFB, it is important to use a control condition. It enables the control of repetition-related training effects and non-specific effects that may be caused by the overall setting (e.g. trainer-subject interactions) [20].

One option is to use a group of subjects that participate in as many training sessions as the experimental group but do not receive NFB. Another option is to present the control group with sham- or pseudo-feedback. In that case, the control subjects receive an artificial computer-generated feedback signal or a replay of the feedback signal derived from another subject/another session. A sham group would allow to control the effect of attention effort that accompanies any EEG-NFB training [21]. However, feedback credibility is an important issue to consider [20] : if a control subject becomes aware that his/her actions do not affect the feedback signal, this might result in passive behavior ("learned helplessness" [20]) and decreased performance. Moreover, inclusion of a sham-feedback control group has not been common practice in BCI-control training and was not done in the MI-BCI training studies that were revised [24, 25]. Therefore, we decided not to provide feedback to the control subject.

2.1.3 Experiment Timeline

Each subject participated in eight AAD sessions divided into a decoder calibration phase, an AAD training phase and a follow up phase. Recording time was kept as consistent as possible, between sessions and between the two subjects, to avoid performance fluctuations depending on the time of the day (due to the influence of non-specific effects such as sleepiness, hunger etc.). Figure 2.4 presents an overview of the experiment timeline, which we further explain here.

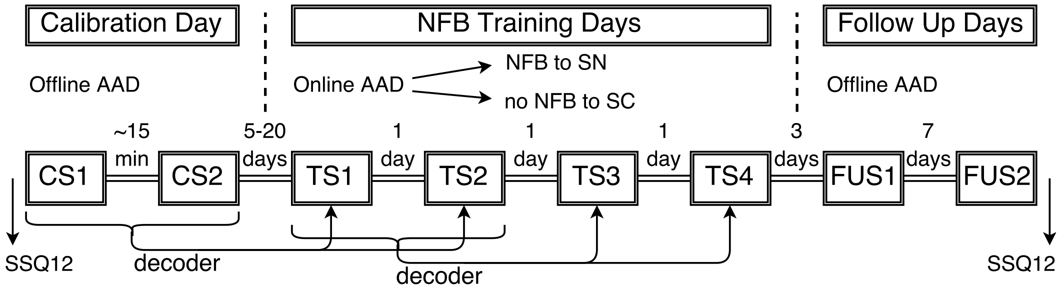


Figure 2.4: Overview of the experiment timeline. The experiment consisted of 8 sessions: 2 calibration sessions (CS), 4 training sessions (TS) and 2 follow up sessions (FUS). Amount of time between each session is indicated. During TS1 & TS2 (resp. TS3 & TS4), AAD was detected online based on a decoder trained on CS1 & CS2 (resp. TS1 & TS2). During training phase, SN was presented with online NFB on AAD performance; SC was not. Before (at CS1) and after (at FUS2) the experiment, subjects self-assessed auditory performance with the SSQ12 test [40].

Calibration Day

To be able to provide NFB in an online application, a decoder must be available. A grand average decoder could be constructed from data from previous experiments, by averaging decoders across subjects. In our study however, where the focus lies on the individual, we opted to construct a subject-specific decoder for every subject. It is assumed that NFB will then be more effective, since exact characteristics may vary across subjects as a function of age, task performance capabilities, brain volume etc. [20]. Besides, subject-specific decoders were reported to generally perform better than grand average decoders [2], which is beneficial for the accuracy of the NFB provided during the online sessions. Each subject participated in two calibration sessions (CS) which took place on the same day. At the end of the Calibration Day, a subject-specific decoder could thus be constructed, based on 48 min of EEG training data. According to previous literature, such a decoder would be reasonably accurate [9] and additional training data would not result in substantially better performance [2].

Neurofeedback Training Days

Five to twenty days after the Calibration Day, subjects started a series of four training sessions (TS). Only one training session was given per day, with one day of rest between consecutive training sessions. During these sessions, attention was detected online for both subjects, but NFB was only presented to SN. This subject was asked to maximize NFB output (i.e., sphere size) by executing the auditory attention task. NFB thereby intended to facilitate the subject in self-regulating brain activity towards higher AAD performance. During the first two training sessions, the subject-specific pre-trained decoder that was constructed after the calibration sessions was used for online detection. After TS2, a new decoder was constructed, for both subjects, based on the data acquired during TS1 and TS2 (48 min of training data). This decoder performed the online detection during following training sessions 3 and 4.

Follow Up Days

Neurofeedback training is beneficial for BCI control, only if after training, when subjects stop receiving NFB about their performance, they still perform better than before training. Permanent learning effects are usually assessed after a period of time without training.

To assess for permanent learning effects in our study, each subject was measured two times more during two follow up sessions (FUS). The first one took place four days after the last training session; the second one another eight days later. During follow up sessions, neither one of the subjects received NFB.

2.2 Analysis

Our main interest was the effect of NFB training on AAD. In this section, we will present how we analyzed our data in order to evaluate this effect.

2.2.1 Offline and Online AAD Performances

We used the leave-one-trial-out cross-validation approach to evaluate offline AAD performance. In this approach, each trial of a certain data set is decoded by a decoder trained on all other trials of that data set. Performance is then the percentage of trials that is correctly decoded. This way, we computed offline AAD performance on the data of each block (= 36 trials) and of each session (= 144 trials), for both subjects.

We computed online AAD performance, per training block and per training session, as the percentage of trials that was correctly decoded during that block or session. We computed this for both subjects. Furthermore, we computed CA-CUA mean and variance per block and per session, for each subject, to get a better insight into the effects of training on AAD.

2.2.2 Evaluation of Training Effects

Our main interest was the effect of training across sessions. However, it is interesting to assess for learning effects within a training session as well. For instance, subjects could perform better and better towards the end of a session through task adaptation or subject learning. On the other hand, effects such as fatigue and attention loss could cause a decreased performance towards the ends of sessions. For the same reasons, performance could change within a training block (i.e., a block within a training session). We evaluated all three effects. Furthermore, we assessed whether (NFB) training had an effect on the spatiotemporal characteristics of the decoder across sessions. Here, we explain how each of these effects was analyzed.

Across Sessions

We tracked online AAD performance across training sessions, for both subjects. Performance during training sessions was compared with offline performances on calibration sessions. The latter reflect the subjects' initial AAD performance. Furthermore, we assessed for permanent learning effects by evaluating offline performance on follow up sessions. Mean and variance of CA-CUA were compared across sessions as well, as they provide a better insight into the effects of training on AAD.

Within a Training Session

To evaluate changes within a training session, we computed mean performance per training block, by averaging across the four training sessions. Similarly, we computed CA-CUA mean and variance per training block. All three parameters were compared across training blocks, for both subjects. Furthermore, we compared CA-CUA of trials during 1^{st} halves of training sessions with CA-CUA of trials during 2^{nd} halves, using the Two-sample T-test.

Within a Training Block

To assess for a change in performance within a training block, we compared performances during the 1^{st} halves of training blocks with performances during the 2^{nd} halves. We thereby used the Wilcoxon Signed Rank test. Furthermore, we compared CA-CUA of trials during 1^{st} halves with CA-CUA of trials during 2^{nd} halves, using the Two-sample T-test.

On Spatiotemporal Decoder Characteristics

For both subjects, we analyzed the spatiotemporal characteristics of the session decoders to assess for changes across sessions. When reconstructing the speech envelope, a decoder includes information from 7 time-lagged versions of the EEG signals, measured at 24 electrodes. A decoder is thus a 24 x 7 matrix: for each of the 24 electrodes, it contains one weight for each of the 7 time-lagged versions of the EEG signal measured at that electrode. These weights reflect the importance of a certain time lag, at a certain electrode. A decoder thus contains spatial information about the relative importance of each electrode, and temporal information about the relative importance of each time lag.

Per session, we assessed the importance of each of the 24 electrodes. This corresponds to the overall weight each electrode was given by the decoder. To obtain an overall weight for each electrode, we averaged the absolute decoder weights across all time lags. Result is one value for each of the 24 electrodes, indicating its importance. The distribution of these overall weights across the scalp can be displayed as a topographic map.

Furthermore, we assessed the importance of each of the 7 time lags per session. This corresponds to the overall weight each time lag was given by the decoder. To obtain an overall weight for each time lag, we averaged the absolute decoder weights across all electrodes. Result is one value for each of the 7 time lags, indicating its importance. This allowed us to identify the most important time lag for each session, corresponding to the time-shifted version of the EEG that contributed the most to the envelope reconstruction.

2.2.3 The Specific Effect of Neurofeedback

In order to evaluate the specific effect of NFB, training effects were compared between both subjects. We tracked AAD performance, mean CA-CUA and CA-CUA variance of SN relatively to that SC. These relative parameters were computed for each session session as:

$$Relative\ Parameter = \frac{Parameter\ SN - Parameter\ SC}{Parameter\ SC}.100\%$$

2.3 Results

We will first present the behavioral results and overall online and offline AAD performances. These will allow us to validate our experimental setup. We will then present the results regarding the effects of NFB training.

2.3.1 Behavioral Results

Individual scores per session on the multiple-choice questions are shown in Table 2.1. On average, SC correctly answered 86.71 ± 6.20% of the questions that were asked in a session. For SN, this was 89.84 ± 5.73%. These results show that both subjects successfully performed the behavioral task. Similar results were found in previous AAD experiments [3, 9]. Average ratings of task difficulty and interest per session are shown in Table 2.2 and Table 2.3, respectively.

The SSQ12 test [40] served as a self-assessment test of auditory performance in various listening situations. A higher score indicated better performance. At the start of the first session (CS1), SC scored 94/120, while SN scored 60/120. After the last session (FUS2), SC scored 89/120, while SN scored 75/120.

Table 2.1: Session scores on multiple-choice questionnaires (in %).

	CS1	**CS2**	**TS1**	**TS2**	**TS3**	**TS4**	**FUS1**	**FUS2**	**AVG±SD**
SC	81.25	100	81.25	87.5	81.25	87.5	87.5	87.5	86.71 ± 6.20
SN	100	87.5	81.25	87.5	87.5	87.5	93.75	93.75	89.84 ± 5.73

Table 2.2: Average difficulty rating per session (on 10).

	CS1	**CS2**	**TS1**	**TS2**	**TS3**	**TS4**	**FUS1**	**FUS2**	**AVG±SD**
SC	5.5	6	4.75	6.5	5	6	5.25	7	5.75 ± 0.77
SN	6	4.75	4.5	4.75	5.5	5	4.25	5	4.97 ± 0.56

Table 2.3: Average interest rating per session (on 10).

	CS1	**CS2**	**TS1**	**TS2**	**TS3**	**TS4**	**FUS1**	**FUS2**	**AVG±SD**
SC	6.25	6.25	6.5	7.5	6.25	7	6.5	5.25	6.44 ± 0.65
SN	6.25	3.25	4.25	4.5	5	5.5	4.75	4.25	4.72 ± 0.90

2.3.2 Offline and Online AAD Performances

Average (±SD) and range of offline AAD performances (in %) are shown in Table 2.4, for blocks and sessions, and for both subjects. Average (±SD) and range of online performances during training blocks and training sessions (in %) are presented in Table 2.5, for both subjects. These performances are comparable with previous AAD experiments [2, 9], which confirms the working of our AAD setup, and its robustness to measuring in the home environment.

Table 2.4: Offline AAD performances on blocks and sessions: range and AVG±SD for both subjects.

OFFLINE		MIN	MAX	AVG±SD
Block	SC	61.11	91.67	76.74±7.53
	SN	52.78	97.22	76.39±10.49
Session	SC	73.61	84.72	79.69±3.91
	SN	69.44	88.19	78.99±5.88

Table 2.5: Online AAD performances on training blocks and training sessions: range and AVG±SD for both subjects.

ONLINE		MIN	MAX	AVG±SD
Block	SC	72.22	88.89	78.3±5.67
	SN	66.67	88.89	79.17±6.95
Session	SC	75.69	80.56	78.30±2.62
	SN	77.08	81.94	79.17±2.20

2.3.3 Training Effects and the Effect of Neurofeedback

Across Sessions

Figure 2.5 shows online performances during training sessions, for both subjects. Offline performances during calibration and follow up sessions are shown as well. Dash-dotted lines indicate the on- and offset of the presentation of NFB to SN (from left to right, respectively). It is clear that for both subjects, performance increased from the initial performance at CS1, but there was no consistent increase across sessions. When comparing SN with SC, there seems to be a strong correlation between the session performances of both subjects. This correlation was indeed found strong and significant (r=0.81, P=0.015). Furthermore, while SN had a lower initial AAD performance than SC, SN outperformed SC on TS1. With exception of TS3, SN continued to perform better than SC throughout the rest of the experiment.

Figure 2.6 shows relative performance of SN to SC. While SN performed almost 6% worse than SC initially, SN became increasingly better than SC throughout training sessions. Although, there was an inconsistency at TS3. Furthermore, during FUS1, SN still outperformed SC. However, her relative performance had decreased since TS4 and decreased more towards FUS2.

Figure 2.7 shows mean CA-CUA per session, for both subjects. Mean CA-CUA was higher during training sessions compared to the initial value, and the effect was greater for SN than for SC. However, relatively to SC, SN did not show a consistent increase in mean CA-CUA throughout training sessions (Figure 2.8).

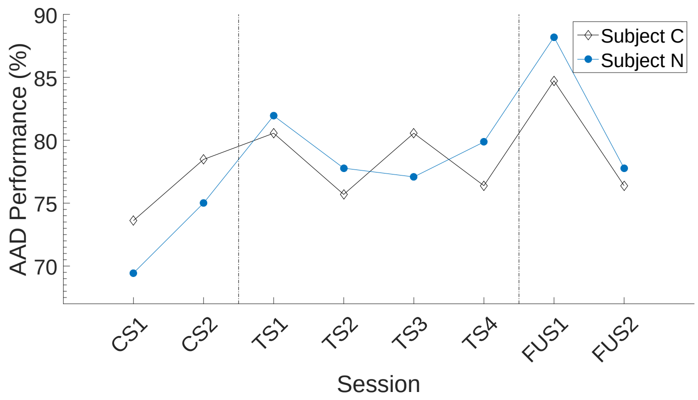


Figure 2.5: Evolution of session performance for both subjects. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left and right, respectively).

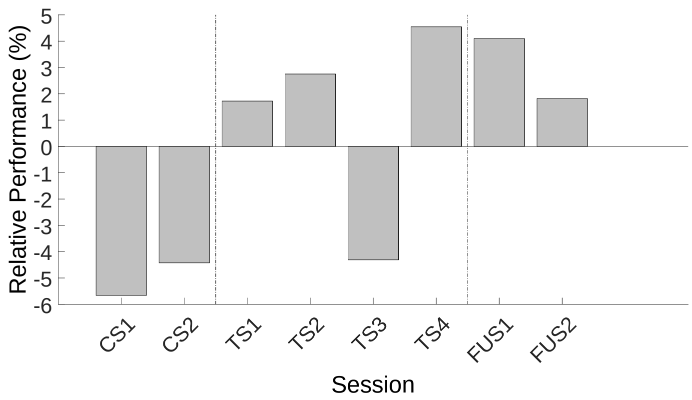


Figure 2.6: Evolution of relative session performance of SN to SC. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left to right, respectively).

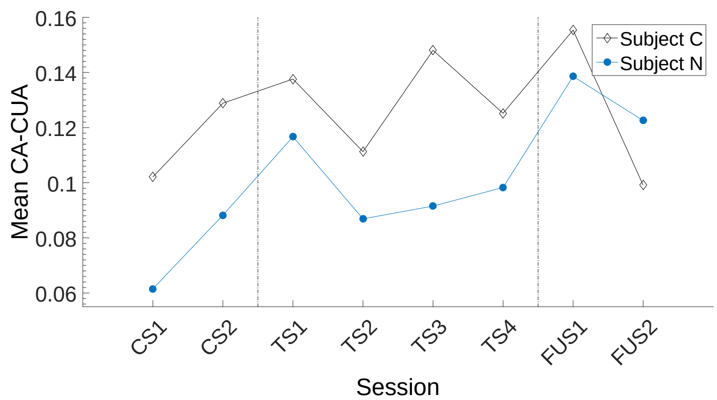


Figure 2.7: Evolution of mean CA-CUA for both subjects. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left to right, respectively).

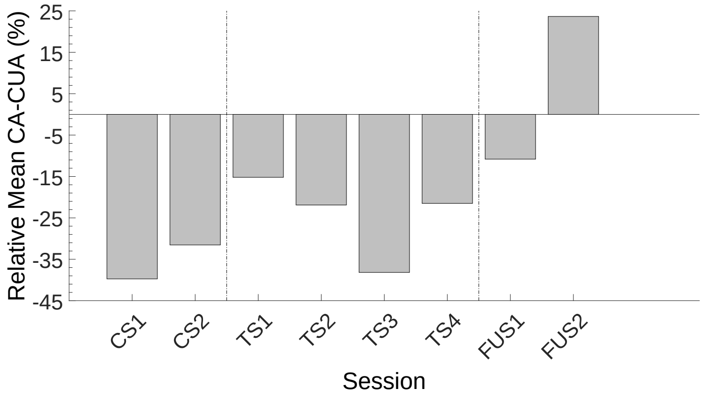


Figure 2.8: Evolution of relative mean CA-CUA of SN to SC. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left to right, respectively).

Figure 2.9 shows CA-CUA variance per session, for both subjects. Results show that both subjects started with more or less the same CA-CUA variance. While variance of SC was almost consistently higher than initial variance, variance of SN was consistently lower. Furthermore, relatively to SC, SN showed a consistently decreasing variance throughout training sessions (from 38% less variance during TS1 to 50% less variance during TS4). This effect remained during FUS1, although it had diminished since TS4, and diminished more towards FUS2 (Figure 2.10).

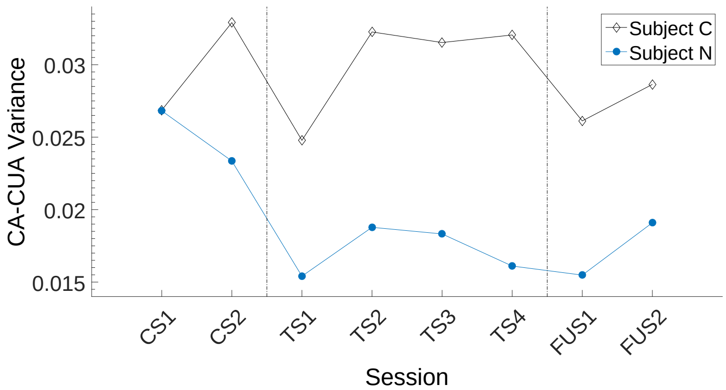


Figure 2.9: Evolution of CA-CUA variance for both subjects. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left to right, respectively).

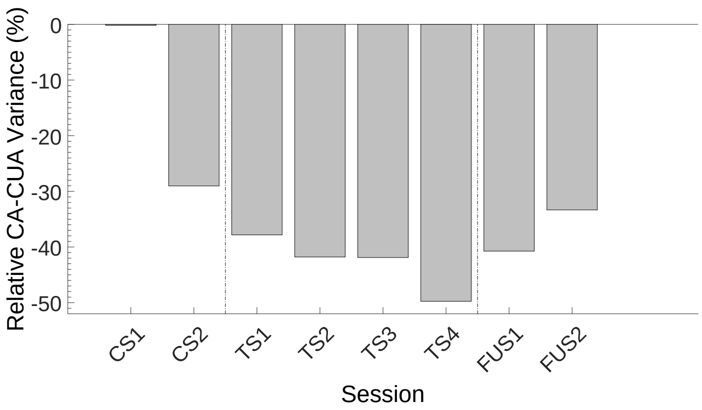


Figure 2.10: Evolution of CA-CUA variance of SN relatively to SC. Evolution of CA-CUA variance for both subjects. Dash-dotted lines indicate the on- and off-set of NFB to SN (from left to right, respectively).

Within a Training Session

Figure 2.11 shows mean performance, and mean and variance of CA-CUA per training block. Results show that online decoding performance of SN increased during the 1st half of a training session, but dropped strongly during 2nd half. Similar effects are seen for CA-CUA mean and variance which decrease, respectively increase, towards the end of a training session. SC showed less change, although performance also decreased towards the session end. Furthermore, we found that for SN, mean CA-CUA was significantly lower during the 2nd half than during the 1st half of a training session ($p<0.001$). For SC however, CA-CUA did not significantly change between session halves ($p=0.260$).

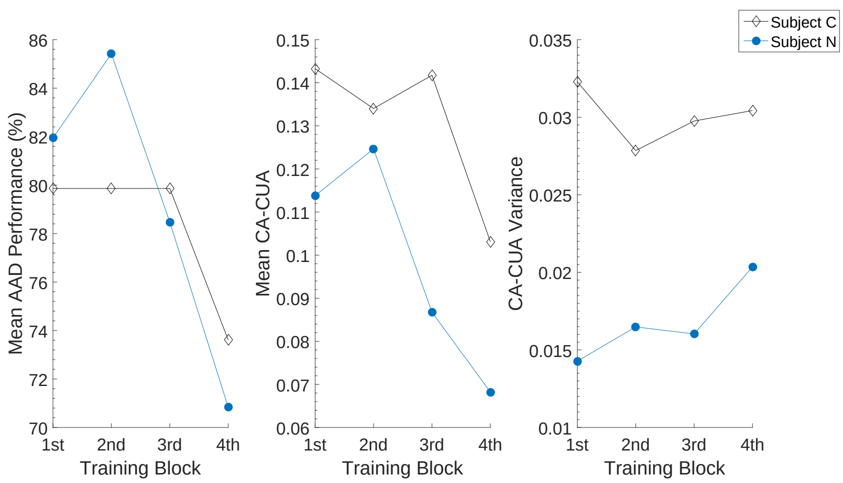


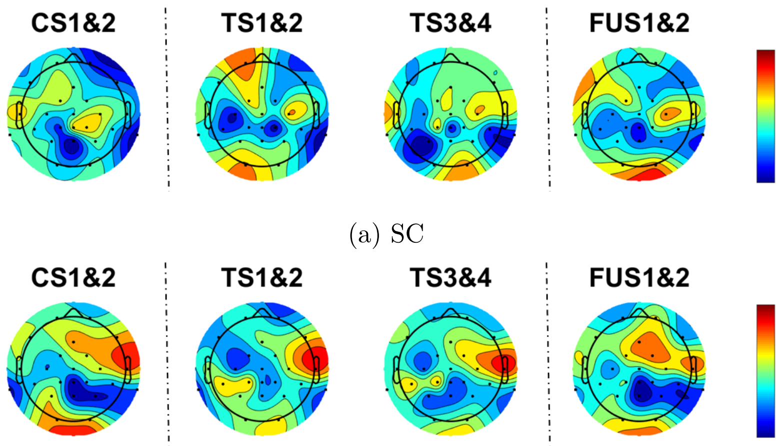
Figure 2.11: Dynamical change in performance within a training session. Mean Performance (left), mean CA-CUA (middle) and CA-CUA variance (right) across an average training session.

Within a Training Block

We found no significant change in median performance between the 1st half and 2nd half of a training block, for neither subject (SC: p_w=0.540; SN: p_w=0.910). However, for SC, CA-CUA had a significantly lower mean value during the 2nd half of a training block (p=0.049) than during the 1st. For SN, CA-CUA did not significantly change in mean (p=0.443).

On Spatiotemporal Decoder Characteristics

Figure 2.12 shows topographic maps of the spatial distribution of decoder weights (averaged across time lags) throughout the experiment, for both subjects. For visual comparison, we averaged topographies per 2 sessions. Red indicates a relatively high overall weight for the electrode; Blue indicates a relatively low overall weight. Note that colors are only relative within a topographic map, and dash-dotted lines indicate the on- and offset of the presentation of NFB to SN (from left to right, respectively).



(a) SC

(b) SN

Figure 2.12: Topographic maps of the spatial distribution of decoder weights throughout the experiment, for a) SC and b) SN. (Red: high weight. Blue: low weight). Dash-dotted lines indicate the on- and offset of NFB to SN (from left to right, respectively)

The spatial distribution for SN appeared quite consistent throughout the experiment. Distribution for SC seemed much less consistent than for SN. We could quantify the consistency in spatial distribution, for both subjects, by computing the correlations between consecutive distributions (averaged per 2 sessions) (Table 2.6). These confirmed that spatial distribution was indeed more consistent for SN than for SC.

Table 2.6: Correlations between consecutive spatial distributions, for both subjects.

	CS to TS	**Within TS**	**TS to FUS**
SC	r=0.30, P=0.150	r=0.40, P=0.051	r=0.57, P=0.004
SN	r=0.66, P=0.000	r= 0.76, P=0.000	r=0.72, P=0.000

Regarding the temporal characteristics, there was no consistent shift in most important time lag, for neither subject. For SC, the most important time lag was 200 ms for all sessions except TS3, where it was 100 ms. For SN, the most important time lag varied between 150 ms and 200 ms.

2.4 Discussion and Conclusion

We aimed to assess whether Neurofeedback Training can be used as method to enhance AAD performance, and whether training (and thus AAD) is feasible at the subject's home. The AAD paradigm was based on the paradigm in [1] and implementation of the AAD setup was adopted from a previous study [9]. We developed a Neurofeedback Training paradigm based on an extensive study of literature on EEG-Neurofeedback Training [19]-[23]. Besides, we were guided by successful practice in previous studies on BCI training [24, 25]. Two subjects participated in our experiment. We measured each subject during eight 24 min sessions, on different days. These consisted of two calibration sessions (CS), four training sessions (TS) and two follow up sessions (FUS). During calibration sessions, we assessed the subjects' initial AAD performance and we trained a decoder that could be used for online detection during the training sessions. During these training sessions, attention was detected online for both subjects, although only one subject, Subject N (SN), was presented with online NFB on AAD performance. This subject was thereby stimulated to maximize NFB output. The other subject, Subject C (SC), did not receive NFB, which allowed the control of non-NFB-specific training effects. Follow up sessions served to assess for permanent learning effects. During these sessions, neither subject was presented with NFB.

We evaluated training effects across sessions, within a training session, and within a training block. Therefore, we tracked AAD performance, and mean and variance of CA-CUA. To reveal the specific effect of NFB, these three parameters were compared between both subjects, by computing the relative parameter value of SN to SC. Furthermore, we evaluated the effect of (NFB) training on the spatiotemporal characteristics of the decoder across sessions.

In this section, we will discuss our results regarding the effect of NFB training on AAD. Furthermore, we will discuss implications for future applications and revise some of the used methods.

2.4.1 Training Effects and the Effect of Neurofeedback

AAD performance of neither subject consistently increased across sessions, which would suggest that there was no training effect. However, we hypothesize that performance during a session depended on which audio stimulus was presented during that session, which could have masked the learning curve. Our hypothesis is based on the fact that a strong correlation was noticed between the session performances of both subjects, who were presented with the same order of audio stimuli. Post hoc, we further analyzed the influence of the audio stimulus on AAD performance (Section 2.5).

In the assessment of the specific effect of NFB, the relative performance of SN to SC accounted for the potential influence of the audio stimulus and for other non-NFB-specific effects. We found that while SN performed almost 6% worse than SC initially, SN outperformed SC after one training session and became increasingly better throughout training. This might justify the assumption of a NFB-specific learning effect. Inconsistency was noted at TS3. One hypothesis is that SN had an "off day" during TS3, which caused her to perform less well in the behavioral task. Although, this was not noticeable in any of the behavioral measures. Another hypothesis is that AAD performance of SN was influenced by the change in online decoder: As we explained in Topic 2.1.3, the online decoder was updated after TS2. On TS3, SN's neural activity was thus decoded differently than before, which changed its relationship with the online NFB that was presented. SN could have needed some time to get used to this new relationship, before being able to again improve AAD performance during TS4.

Furthermore, during FUS1, when SN was deprived of NFB, her relative performance remained increased. However, it had diminished since the last training session, and diminished more towards FUS2. This suggests that part of the learning effect remained after training, but that it diminished over time.

When tracking mean CA-CUA across sessions, it was interesting to notice that it was lower for SN than for SC on all but the last session. This shows that between subjects, lower mean CA-CUA does not mean lower AAD performance. Furthermore, although relative mean CA-CUA increased at the onset of NFB for SN, it decreased throughout the rest of the training. It could thus not explain the relative increase in performance. This is an unexpected finding, as NFB was intended to facilitate SN in maximizing CA-CUA, which would thus result in a relative increase in mean CA-CUA. However, we hypothesize that a growth in sphere radius did not clearly reflect the proportionality of CA-CUA, and that the NFB signal was thus rather experienced as binary ("growth or no growth").

Relative CA-CUA variance showed the exact inverse pattern of relative performance: it decreased throughout training sessions (with exception of TS3), and gradually increased again during follow up sessions. This suggests that NFB had effect on AAD performance, mostly be causing a more consistent CA-CUA instead of a more distinct CA-CUA. The former can be related to the "stability" of the subject's neural responses to the behavioral task, the latter to the "distinctiveness". Again, the effect increased throughout training, and some effect remained after training, although it diminished over time.

We found no learning effects within a training session. AAD performance of both subjects dropped towards the end of a training session, which might imply that the duration of 24 min exceeded the subjects' attentional limits. However, we hypothesize that most of the change within a session was caused by left-right differences: From the 1^{st} to the 2^{nd} half, attended ear switched from right to left. This might have caused a decreased performance during 2^{nd} half, as AAD performance has been shown to be generally better on right ear than on left ear trials [3]. This can be related to a phenomenon termed the right ear advantage [43]. Therefore, to evaluate the effects of NFB within a session, better practice would have been to randomize ear attendance alternation between sessions.

Within a training block, performance did not significantly change. Although, for SC, CA-CUA was lower during the 2^{nd} half than during the ^{1st}, which could imply that attention of SC decreased towards the end of a training block.

When assessing the effect of NFB on the spatiotemporal decoder characteristics, we found that the most important time lag did not consistently change across sessions, for neither subject, which suggests that it was not affected by training. It mostly varied between 150 ms and 200 ms (across sessions and subjects), which corresponds to the finding in AAD literature that neural processing at ~150 - 200 ms is critical for decoding auditory attention to speech [1, 2].

We found strong correlations between the spatial distributions of decoder weights across sessions for SN. These were stronger for SN than for SC. Since during TS1&2 and TS3&4, the pre-trained decoder was based on CS1&2 and TS1&2, respectively, these correlations (Table 2.6) thus reflect the similarity between the spatial distribution during training sessions and the spatial distribution of the decoder used for online detection. This is an interesting result, as it suggests that NFB was effective. Indeed, if NFB is effective, the spatial distribution during training sessions is expected to be similar to that of the pre-trained decoder on which the online NFB is based. The subject then "provides" the information needed for correct decoding at those electrodes where the pre-trained decoder "searches" this information.

All in all, these results suggest that NFB training has been effective in improving AAD performance. Training would thereby mostly facilitate the subject in eliciting more stable neural responses to the behavioral task, which results in a more consistent CA-CUA. This effect would be accompanied by an effective modulation of the spatial distribution of neural activity. Learning effects would increase throughout training sessions, and would partly remain after training, but would decrease over time. It is important, however, to interpret these results with caution. First of all, the relative AAD performance increase could have been caused by a relative increase in behavioral performance. This hypothesis is backed by the findings of O'Sullivan et al. [1]. Across subjects, they found a significant correlation between behavioral performance and CA-CUA variance. They therefore postulated "that good behavioral performance follows as a result of consistency in sustaining attention, and that this consistency should be measurable in terms of a consistent CA-CUA" [1]. Our results did indeed show an increasing relative consistency in CA-CUA of SN compared to SC, which thus might imply that an increasing relative behavioral performance was the cause of the increasing relative AAD performance. Although, this hypothesis could not be backed by our behavioral results, as no correlation was found between relative AAD performances and relative scores on the multiple-choice questionnaires. A second remark is that our control condition did not account for the effect of attention effort that accompanies any EEG-NFB training ([21]). Again, this effect could have explained the relative AAD performance increase through a relative increase in behavioral performance of SN to SC. Thirdly, and maybe most importantly, a valid conclusion about the effectiveness of NFB on AAD performance cannot be made based on the comparison of only two subjects. An extensive set of individual factors could have influenced our results such as learning susceptibility, age, degree of brain plasticity, attentional capacity, auditory performance etc. Nevertheless, our study does provide promising results for further study on a larger population, and proves the feasibility of AAD and NFB training at the subject's home.

2.4.2 Implications for Applications

In our study, the goal of NFB training was to improve AAD performance by facilitating self-regulation of neural activity. It has not yet been empirically confirmed whether self-regulation would cause a "transfer effect" on behavior – i.e., changes in auditory performance in a cocktail-party listening situation. Results on the SSQ12 test [40] suggest that auditory performance of SN was higher after training than before. Although not a conclusive result, it might be promising for future research on using AAD NFB training as a therapeutic tool to treat the symptoms of hearing loss, or to enhance auditory performance of healthy subjects. Indeed, there is evidence for a relationship between hearing loss and the ability to neurally track speech [44]. Since AAD is based on the neural tracking of speech, improving AAD performance might alleviate the symptoms of hearing loss.

2.4.3 Revision of the Methods Used

Sham-Feedback Control Group

Our results suggest that attention effort has been an important factor in explaining the observed relative performance increase of SN to SC. Therefore, future research should consider implementing a sham-feedback condition to account for this effect.

Study Blinding

We adopted a single-blind study design – i.e., the subject did not know whether she belonged to the experimental or the control group. However, for future studies, better practice would be double- (both subject and researcher do not know the group assignment) or triple- (a third party assessing the effects) blind study design. This would allow to rule out non-specific factors such as expectancy effects [20].

Online Trial Length

In our decision regarding online trial length (TL), we considered a trade-off between temporal resolution and the accuracy of the given NFB. Based on a previous study [9], we decided to use a trial length of 10 s. However, better practice might have been to determine the trial length that maximizes the information transfer rate (ITR) [45] of the NFB. Every NFB presentation contains a certain amount of information, depending on the probability that it is based on correct detection, that is:

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 \frac{1 - P}{N - 1} \text{ [45]} \tag{2.3}$$

where B is in bits/presentation, N is the number of sound sources that can be selected, and P the probability that the correct one is selected. In our case, $N = 2$ and P is equal to the online decoder performance.

As trial length decreases, NFB is presented at a higher rate. Although, each presentation contains less information as decoder performance decreases. Choosing the trial length that maximizes the ITR (2.4) would thus maximize the amount of information presented to the subject during a training session, which could be beneficial for training efficiency.

$$ITR = B . \frac{1}{TL} \text{ [45]} \tag{2.4}$$

2.5 Post-hoc Additional Analyses

Besides analyzing the effect of NFB training on AAD, the large amount of data collected for each subject allowed us to make several additional post-hoc analyses. These regarded the influence of the audio stimulus on AAD performance, the use of unattended decoders, and the effect of training duration on decoder performance. Each will be discussed in a separate topic.

2.5.1 The Influence of the Audio Stimulus on AAD Performance

In Section 2.4, we stated our hypothesis that AAD performance depended on which audio stimulus was presented. It was based on the fact that a strong across-sessions correlation was noticed between the performances of both subjects, who were presented with the same order of audio stimuli. The influence of the audio stimulus has not been addressed by the AAD community yet. The large amount of audio stimuli that we used in our experiment provided us the opportunity to further investigate this effect.

Analysis

First, we further investigated the performance correlation by comparing offline block performance, and CA-CUA mean and variance, between both subjects. We then addressed the question of which stimulus factors could have influenced performance and how. For this part, we averaged AAD performance, difficulty rating and interest rating across both subjects. We hypothesized that performance could have been influenced at different levels of the AAD process: Certain stimulus factors could have acted at the behavioral level, by determining task difficulty; Other factors could have impacted AAD in a more direct way, by acting on the (neural) mechanisms and methods on which AAD is based.

Factors acting at the behavioral level The audio stimulus is a determining factor for the difficulty of the behavioral task. Certain audio or story features could make the task more or less difficult. Think of factors such as clearness of the attended speaker's voice, difficulty of vocabulary, distracting sounds in the to-be-ignored stream, how interesting both stories are etc. To test whether AAD performance was influenced by the difficulty of the task (as perceived by the subjects), we evaluated the across-blocks correlation between performance and difficulty rating (both averaged across both subjects). Furthermore, we assessed whether interest in to-be-attended story was a determining factor of perceived task difficulty, by evaluating the across-blocks correlation between interest rating and difficulty rating (both averaged across both subjects).

Factors acting on the (neural) mechanisms of AAD We hypothesized that there were other stimulus factors that could have impacted AAD performance without affecting difficulty of the behavioral task. Their impact would have been directly on the (neural) mechanisms and methods on which AAD is based. For example, these factors could have influenced fluctuation of EEG activity with (attended) speech, thereby influencing stimulus-reconstruction and the detection decision. To test our hypothesis, we evaluated, per block, eighteen time domain features of the attended audio stream envelope. They are explained in further detail in Appendix B. Matlab scripts for feature calculations were developed in a study by Koolen et al. [46]. To assess whether these features influenced task difficulty and/or AAD performance, we computed for each feature its across-blocks correlation with difficulty rating and with AAD performance, respectively (both averaged across subjects). For those features that did correlate significantly with AAD performance, we further evaluated their across-blocks correlations with CA-CUA mean and with CA-CUA variance, respectively (both averaged across subjects).

Results

Figure 2.13a shows offline block performances for both subjects throughout the experiment. Figure 2.13b shows block performance of SN versus SC. We found a moderate but significant correlation between subjects for block performances (r=0.46, P=0.009) and for mean CA-CUA (r=0.39, P=0.029), but not for CA-CUA variance (r=0.24, P=0.179). Furthermore, we found a slightly stronger correlation for mean CA (r=0.52, P=0.002), while mean CUA did not correlate significantly (r=0.25, P=0.164).

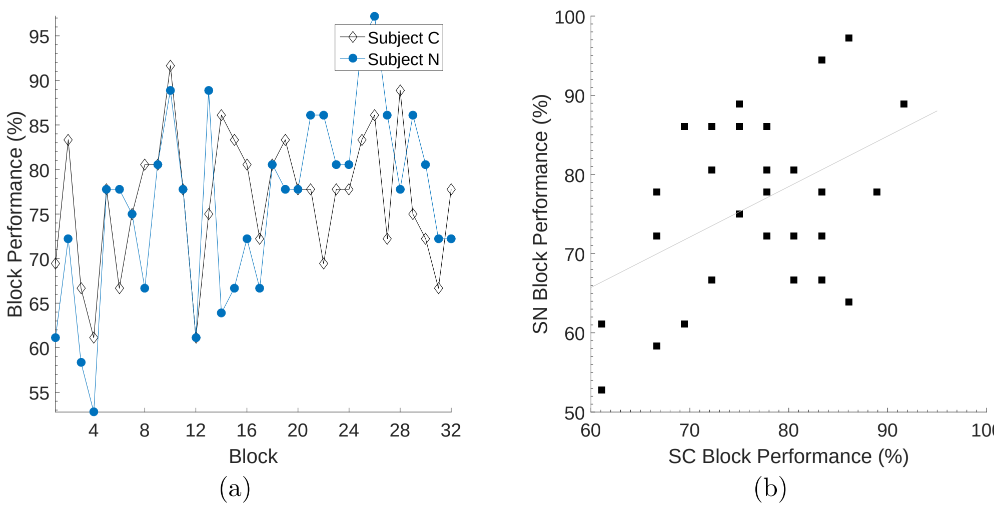
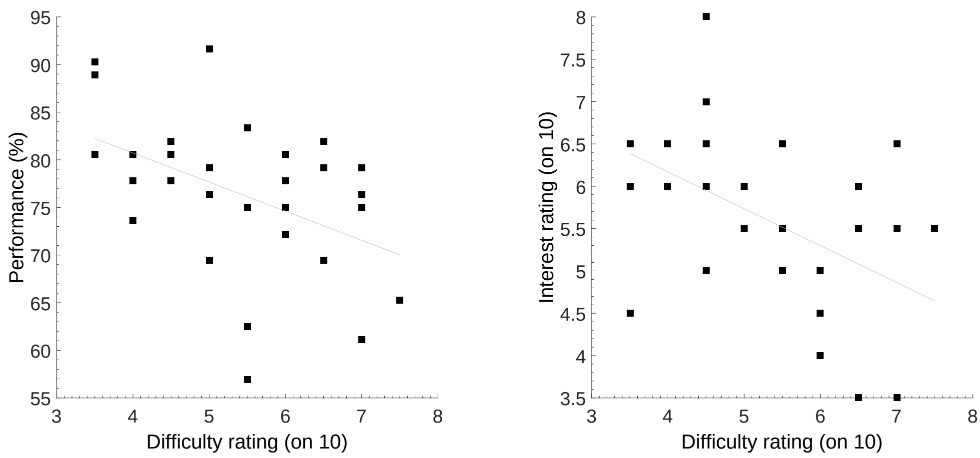


Figure 2.13: Performance correlation between both subjects. a) Block performances for both subjects; b) Block performance of SN versus SC (r=0.46, P=0.009).

Factors acting at the behavioral level We found a significant negative correlation (r=−0.46, P=0.009) between block difficulty rating and block AAD performance (Figure 2.14a), and between block difficulty rating and block interest rating (r=−0.47, P=0.006)(Figure 2.14b).



(a) (b)

Figure 2.14: Difficulty rating versus a) AAD performance (r=−0.46, P=0.009); b) Interest rating (r=−0.47, P=0.006) on blocks (all averaged across subjects).

Factors acting on the (neural) mechanisms of AAD Table 2.7 presents the correlations, across blocks, of eighteen time domain features of the attended envelope with difficulty rating and with AAD performance, respectively. Significant correlations are marked in bold. None of these features showed significant correlation with difficulty rating. However, ten features showed a significant correlation with performance. We added visual illustrations of these correlations in Appendix B. Note that, by visual inspection, the correlation of feature 2 with performance seemed to be mainly caused by outlier values. This correlation was therefore no longer considered as significant. Furthermore, note that features 1-5-6-9-14-15 and features 2-7-8-18 strongly correlated among each other (abs(r)>0.80).

Results further showed that Feature 11 and Feature 13 correlated strongly with both mean CA-CUA and variance, but not significantly. However, we found significant correlations between mean CA-CUA and Feature 15 (r=0.38, P=0.030), Feature 16 (r=0.57, P=0.001) and Feature 18 (r=−0.38, P=0.032), respectively (see Appendix B).

Discussion and Conclusion

AAD performances of both subjects significantly correlated throughout the experiment, which confirms our post-hoc hypothesis that performance was influenced by the audio stimuli. The stimulus mostly determined how well the decoder could

Table 2.7: Time domain features of the to-be-attended audio stream envelope: Correlations with Task Difficulty and with Performance. Significant correlations are marked in bold.

		Difficulty		Performance	
N	Feature	r	P	r	P
1	LineLength	-0.14	0.442	**0.41**	**0.019**
2	RMSamp	-0.09	0.621	**-0.51**	**0.003**
3	Slope	0.11	0.553	0.11	0.560
4	Activity	-0.26	0.146	0.11	0.544
5	Mobility	-0.05	0.800	**0.42**	**0.016**
6	Complexity	-0.02	0.898	-0.33	0.064
7	Kurtosis	0.15	0.405	0.26	0.149
8	Skewness	0.16	0.389	0.34	0.056
9	NonLinEnergy	-0.10	0.577	**0.42**	**0.018**
10	ZeroCrossings	-0.31	0.083	0.12	0.505
11	Minima	0.13	0.465	**-0.41**	**0.020**
12	Maxima	-0.08	0.658	0.21	0.249
13	ARModErr	-0.02	0.909	**0.39**	**0.027**
14	VarFirstDer	-0.17	0.344	**0.50**	**0.004**
15	VarSecDer	-0.24	0.185	**0.56**	**0.001**
16	ZCFirstDer	-0.28	0.117	**0.55**	**0.001**
17	ZCSecondDer	-0.30	0.096	0.23	0.197
18	Mean	0.05	0.768	**-0.57**	**0.001**

reconstruct the attended speech envelope – i.e., CA, the reconstruction accuracy of the attended decoder. How well unattended speech was reconstructed (i.e. CUA) was less affected. This is an unexpected finding, as our stimuli were all selected and processed in a way that matched recent publications [3, 4, 9]. It suggests that seemingly similar sets of stimuli can differ in their influence on AAD. Results in previous and future AAD studies may therefore need to be (re)considered regarding which audio stimulus is (was) used. Ultimately, this highlights the need for a "standardized" audio bank that could be used to fairly compare results between future AAD studies and subjects. Other auditory paradigms (e.g. based on auditory oddball [48] or auditory steady-state responses (ASSRs) [49]) have the advantage of standardized audio stimuli that are constant.

Our results suggest that performance was influenced by the difficulty of the behavioral task. Hence the audio stimulus, as a determining factor of task difficulty, could have influenced AAD at the behavioral level. The subject's level of interest in the attended story was one of the stimulus factors acting at this level, as we found a significant correlation with task difficulty. Furthermore, we could identify nine features of the attended envelope that correlated significantly with AAD performance, but not with task difficulty. These stimulus factors are thus suggested to have influenced AAD performance by acting directly on the (neural) mechanisms on which

AAD is based (e.g., the neural fluctuation with speech). Their effect was most evident on the mean CA-CUA which deviated in relation to the features. Regarding AAD, knowing which stimulus factors affect performance can be very useful. First of all, it provides an important insight into the auditory neuroscience aspects, by showing how well certain speech features are neurally represented and how much they contribute to speech stream segregation. Secondly, the attention detection method could be implemented in such a way that it automatically adapts to these factors, in order to increase decoding accuracy. Future studies should further investigate which stimulus factors have an important impact on AAD, and use this knowledge to improve their methods.

2.5.2 Unattended Decoders

Unattended decoders have been shown to perform worse than attended decoders [1]. This can be expected, since unattended speech is much less represented in the brain than attended speech, and thus much more difficult to reconstruct. Therefore, the use of the unattended decoder in making the detection decision has been discarded so far. However, we hypothesized that the information that the unattended decoder provides – i.e., a reconstruction of the unattended speech stream – could still be useful when used in combination with the information of the attended decoder. Because the unattended decoder perform worse overall, does not mean that on a trial basis, the information provided by the unattended decoder is useless. Therefore, we aimed to include this information into the detection.

Analysis

Post-hoc, we evaluated offline unattended decoder performance, per session and for both subjects. We then combined the attended and unattended decoder in two ways: 1) by making a weighted combination of the information provided by both decoders; 2) by employing the decoder that is "most certain" to make the detection decision.

Weighted Combination During testing, both attended (att) and unattended (unatt) decoder computed CA and CUA for each test trial. We implemented a detection algorithm in which the decision was made based on a parameter d, combining information of both decoders as:

$$d = [CA - CUA](att) + w.[CUA - CA](unatt)$$

Detection was correct if $d > 0$. If, for a certain trial, [CA-CUA] (att) was negative, the unattended decoder could turn the detection decision around by indicating the unattended stream (with a high [CUA-CA] (unatt)). However, a weight w was given to the information of the unattended decoder, since its information was still less accurate than that of the attended decoder. In the same way, the unattended decoder could turn around a correct decision made by the attended decoder, if [CUA-CA] (unatt) was very negative.

We tested performance of this combined decoder strategy, per session, with increasing weight for the unattended decoder (w from 0 to 2 with steps of 0.1). This allowed us to identify the optimal weight w and corresponding optimal performance, per session, which we compared with attended decoder performance.

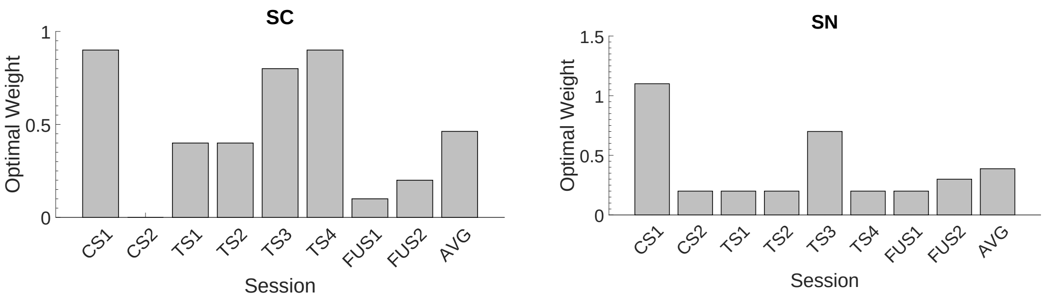
Most Certain Decoder We hypothesized that, whenever a decoder's reconstructed envelope had a small difference in correlation with both speech streams, the decoder was "not very certain" which was the attended, and which was the unattended stream. For a given trial, CA-CUA could thus be seen as a measure of how certain the decoder was in its detection decision for that trial.

Based on this hypothesis, we implemented a decision algorithm in which for each trial, the "most certain" decoder – i.e., the one for which absolute CA-CUA was the highest – decided the detection. We tested performance of this strategy for both subjects and compared its performance with that of the attended decoder.

Results

Average offline session performance with unattended decoders was 61.98±6.29% for SC and 61.46±5.89% for SN. Results per session are shown in Appendix B.

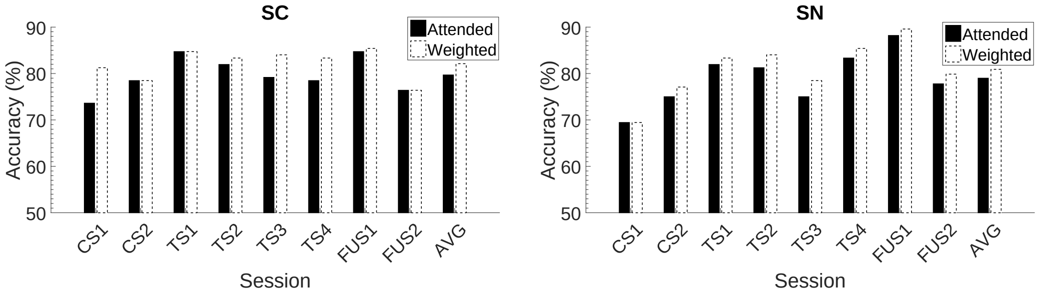
Weighted Combination Figure 2.15 shows the optimal unattended decoder weight w per session and for both subjects. Average (±SD) optimal weight was 0.46±0.36 for SC, and 0.39±0.34 for SN. Figure 2.16 shows AAD performance with the attended and with optimally weighted decoder, per session and for both subjects. For SC, session performance with the weighted combination was on average (±SD) 82.12±3.19%. For SN, this was 80.90±6.14%.



(a) (b)

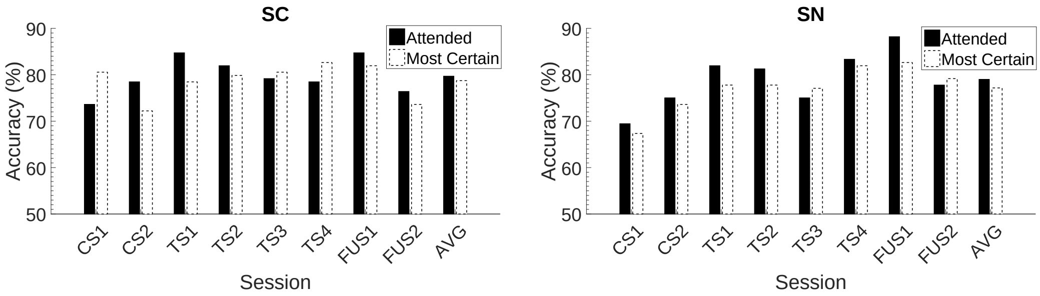
Figure 2.15: Optimal unattended decoder weights per session for a) SC and b) SN.

Most Certain Decoder Figure 2.17 shows AAD performance with the attended and with the most certain decoder strategy, per session and for both subjects. Average session performance with the most certain decoder was 78.73±3.82% for SC, and 77.17±4.87% for SN.



(a) (b)

Figure 2.16: Performance of attended vs. optimally weighted decoder for a) SC and b) SN.



(a) (b)

Figure 2.17: Performance of attended vs. most certain decoder for a) SC and b) SN.

Discussion and Conclusion

For both subjects, and for almost all sessions, giving some weight to the unattended decoder could increase detection accuracy. The most certain decoder, which corresponds to an equally weighted combination, performed on average only 1-2% worse than the attended decoder alone. For some sessions, performance was even better. These results show that the information provided by the unattended decoder can indeed be useful for improving decoding performance. However, since optimal w varied across subjects and sessions, which weight should be given to the unattended decoder is uncertain. When using a combined decoder for online detection, this weight should be decided upfront. Therefore, further research is needed on identifying a generally applicable unattended decoder weight, and on developing more robust combined decoders that could be used in online applications.

2.5.3 Effect of Training Duration on Decoder Performance

The effect of (decoder) training duration on decoder performance has been previously addressed in [2]. However, in that study, subject-specific training data was limited to 30 min. In our study, we collected up to 3 h 12 min of subject-specific data, which provided us the opportunity study this effect in more depth.

Analysis

We evaluated the effect for both subjects separately. Post-hoc, we trained decoders on incremental amounts of training data. The training set started from the data collected during the first block of CS1 (6 min), and incremented including one more block, towards the last block of TS4 (144 min). We tested online performance of these decoders on the (subject-specific) data collected during both follow up sessions (48 min of test data).

Results

Figure 2.18 shows online performance of decoders trained on incremental amounts of data, for both subjects. It is clear that performance increased the most, for both subjects, when training duration increased from 6 min to 12 min. For SC, performance seemed to stagnate from 18 min of training data. For SN, performance continued increasing until training duration was about 102 min.

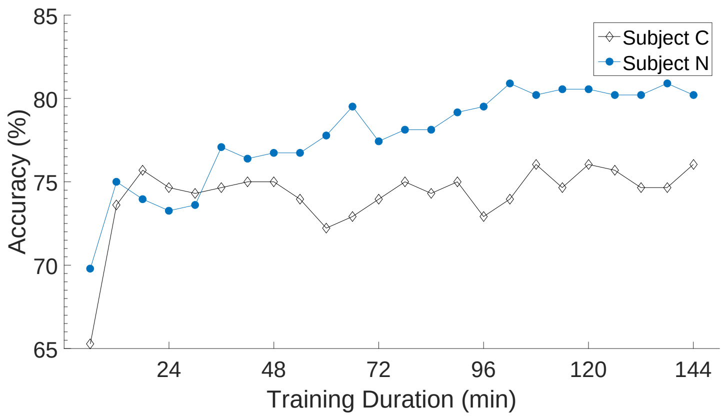


Figure 2.18: Effect of training duration on online decoder performance during FS1&FS2, for both subjects.

Discussion and Conclusion

The effect of training duration on decoder performance seems to be subject-dependent, which is in correspondence with the findings in [2]. Although our results might have been influenced by the NFB condition for SN, they show that, even beyond 30 min, including more training data can still result in more robust online decoder performance.

Chapter 3

Neurophysiological Predictors of AAD Performance

A large inter-subject performance variability has been reported for AAD [7, 9]. In this study, we address the influence of the subject on AAD performance in two ways. In the previous chapter, we assessed whether subjects can improve their AAD performance, by means of Neurofeedback Training. In this chapter, we look for subject-specific characteristics that underlie the inter-subject performance variability. We will explore the relationship between a number of neurophysiological markers and AAD performance, with the aim of identifying predictors of performance – i.e., neurophysiological markers that correlate with AAD performance.

In Section 3.1, we will explain which data we used and how we analyzed it. In Section 3.2, we will present our results which will be discussed and concluded in Section 3.3.

3.1 Methods

Our analysis was based on the data collected in a previous AAD study [9]. In that study, twelve subjects participated in eight AAD sessions of approx. 6.5 min each. Their EEG was measured using 24 Ag/AgCl passive scalp electrodes placed according to the 10-20 standard system (see Appendix A) and AAD setup was similar to the setup explained in Topic 2.1.1. For each subject, we evaluated overall AAD performance and a set of overall neurophysiological markers.

3.1.1 Evaluation of Subject AAD Performance

We evaluated overall offline AAD performance by means of leave-one-trial-out cross-validation on all subject-specific data (Topic 2.2.1). We thereby used a trial length of 10 s.

3.1.2 Evaluation of Subject Neurophysiological Markers

We evaluated overall EEG power in five different frequency bands (Table 3.1), at each of the 24 electrodes. We then defined ten scalp areas (Figure 3.1) in which the power in each band was averaged across all electrodes in the area. This resulted in five overall power values (one for each of the five frequency bands) in each of the ten areas. These were normalized to the total power in the area (i.e., to the power in all five bands).

Table 3.1: Five defined frequency bands

Name	Symbol	Frequency Band
delta	δ	0.5 - 3 Hz
theta	θ	3 - 8 Hz
alpha	α	8 - 13 Hz
beta	β	13 - 25 Hz
gamma	γ	25 - 100 Hz

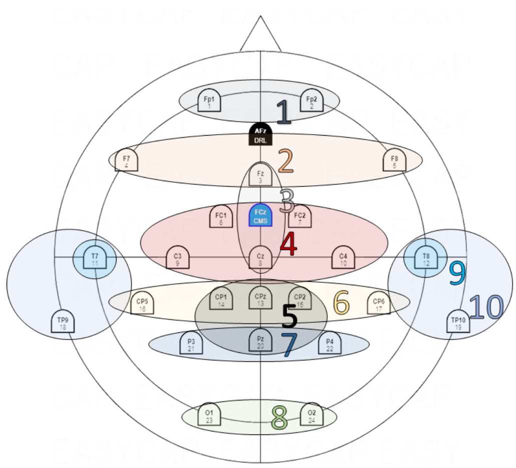


Figure 3.1: Ten defined scalp areas in which power was averaged across the electrodes.

3.1.3 Correlation Analysis

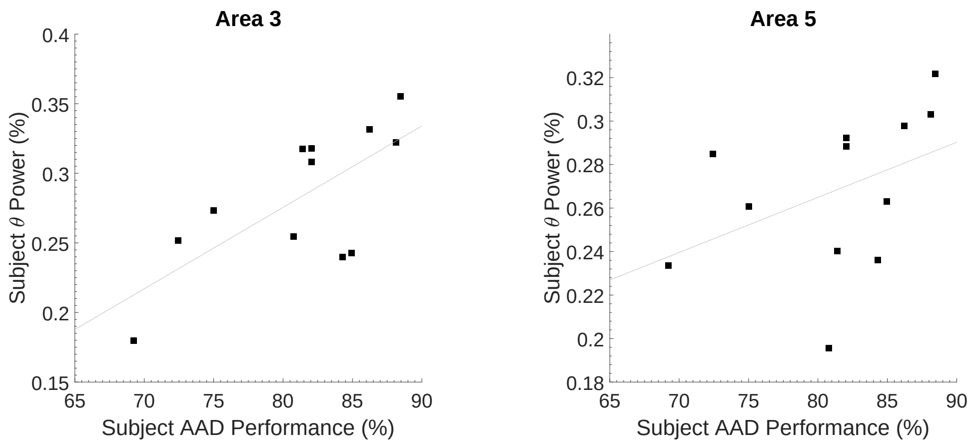
We looked for neurophysiological markers that could predict a subject's AAD performance. We therefore assessed whether the (normalized) overall power in a certain frequency band, in a certain area, correlated with overall AAD performance (across subjects).

3.2 Results

For delta, alpha, beta and gamma power, we did not find any significant across-subjects correlation between AAD performance and the power in a certain area. However, it is interesting to mention that alpha power showed a negative correlation with subject performance in all ten areas, while gamma power showed a positive correlation in all areas. Furthermore, we found that theta power positively correlated with subject performance in all areas. For four areas, this correlation reached significance at the level of 0.10, and for two at the level of 0.05 (Table 3.2). Figure 3.2 shows subject overall performance versus subject (normalized) overall theta power in the two most significant areas, respectively, for all 12 subjects.

Table 3.2: Areas where (relative) overall θ power (%) correlated with subject AAD performance (Spearman's ρ).

Area	**Electrodes**	ρ	**P**
3	Fz, Cz	0.64	0.026
5	CPz, CP1, CP2, Pz	0.71	0.010
6	CPz, CP1, CP2, CP5, CP6	0.50	0.097
7	Pz, P3, P4	0.52	0.082



(a) (b)

Figure 3.2: Subject (nomalized) overall theta power versus subject performance for a) Area 3 (Fz, Cz) (ρ=0.64, P=0.026) and b) Area 5 (CPz, CP1, CP2, Pz) (ρ=0.71, P=0.010), for all 12 subjects.

3.3 Discussion and Conclusion

Based on these results, we hypothesize that low overall alpha power and high overall gamma power are associated with high AAD performance. Furthermore, our results indicate that theta power in centro-parietal areas (area 5, 6 and 7), and frontal and central midline theta (area 3) are able to predict how well a subject performs in AAD. This can be related to a number of studies that associated theta power in frontal and parietal areas with sustained attentional state and cognitive processing during mental tasks [50]-[52]. Moreover, in [51] they showed that frontal midline theta (Fm theta) shows individual differences and is related to mental activity and certain personality traits such as anxiety.

These results are promising for future studies with more subjects. Future research should investigate which personal traits underlie these theta power differences in AAD and how they affect performance. Assessing a subject's personality would eventually allow to customize AAD and training to the individual subject.

Chapter 4

Conclusion

AAD holds a promising potential for controlling a device in real-life (e.g. neuro-steered hearing aids). Such a BCI, as any other, needs the cooperation of both the "computer" and the subject's "brain". In the current study, we addressed the influence of the subject on AAD performance in two ways.

In the first part of this study, we assessed whether subjects can be trained to improve their AAD performance by means of Neurofeedback (NFB) training at home. We developed a NFB training paradigm based on an extensive study of related literature on EEG-NFB training [19]-[23], and successful practice in previous studies on BCI training [24, 25]. This included the design of a NFB cue, and the development of an effective training protocol.

Two subjects took part in our experiment, which consisted of eight AAD sessions divided into a decoder calibration phase, a training phase and a follow up phase. During training phase, which consisted of four sessions over the course of one week, auditory attention of both subjects was detected online with a closed-loop BCI, although only one subject was presented with NFB about ongoing performance. NFB thereby intended to facilitate this subject in self-regulating brain activity to improve AAD performance. The other subject did not receive NFB, which allowed the control of non-NFB specific learning effects. Implementation of the BCI was based on a previous study [9].

We tracked performance of both subjects throughout the experiment to evaluate training effects, and compared both subjects to assess the specific effect of NFB. We found no consistent increase in performance across sessions, for neither subjects, but we hypothesized that the evaluation of a training effect was masked by differences in the audio stimuli across sessions. This was based on a strong correlation that was noticed between the performances of both subjects, who were presented with the same order of audio stimuli. Nevertheless, by comparing both subjects, we could assess the specific effect of NFB. Our results suggest that NFB was effective in improving AAD performance, mostly by facilitating the trained subject to elicit more stable neural responses to the behavioral task. Besides, NFB seemed to cause a modulation of the spatial distribution of neural activity that fitted the spatial distribution of decoder weights of the online decoder. The NFB effect increased throughout training sessions, and partly remained after training, but diminished over time. Although promising, these results should be interpret with caution since we did not account for the effect of attention effort that accompanies EEG-NFB training [21]. Therefore, future studies should use a sham-feedback control group that allows the control of this effect. Moreover, an extensive set of individual factors could have influenced our results such as learning susceptibility, age, degree of brain plasticity, attentional capacity, auditory performance etc. Nevertheless, with this proof-of-concept investigation, we show the feasibility of AAD and NFB training at home, and provide results that are promising for further study on a larger population.

The large amount of data that was collected during the NFB training experiment allowed us to make several post-hoc analyses. First of all, we proved our hypothesis that the seemingly similar sets of audio stimuli that were used in our experiment differed in their influence on AAD. This highlights the need for a "standardized" audio bank that could be used to fairly compare results between future AAD studies and subjects. By identifying some of the stimulus factors that had an important impact on performance, we provide important insights into the auditory neuroscience aspects of AAD. Secondly, we showed that the information provided by the unattended decoder can be used to improve detection accuracy. Thirdly, we investigated the effect of training duration on decoder performance and found that the effect is subject-dependent. Moreover, we found that online decoder performance can increase until training duration reaches up to 102 min.

In the second part of this study, we looked for subject-specific characteristics that underlie the inter-subject performance variability. Therefore, we evaluated the relationship between a number of neurophysiological markers and AAD performance, by analyzing the EEG data collected in a previous AAD study [9]. The goal was to identify predictors of AAD performance – i.e., markers that correlate with performance. We identified theta power in centro-parietal areas, and frontal and central midline theta, as potential neurophysiological predictors of AAD performance. These markers have been related to cognitive performance during mental tasks and to personality traits [50]-[52]. Future research should investigate which personal traits underlie these theta power differences in AAD and how they affect performance. Assessing a subject's personality would eventually allow to customize AAD and training to the individual subject.

Appendices

Appendix A

Illustrations to Section 2.1

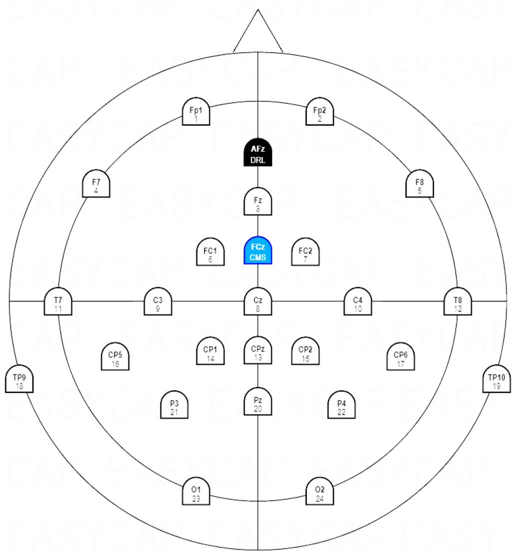


Figure A.1: Electrode positions. 24 Ag/AgCl passive scalp electrodes (Easycap, Herrsching, Germany), placed according to the 10-20 standard system with positions: FP1, FP2, Fz, F7, F8, FC1, FC2, Cz, C3, C4, T7, T8, CPz, CP1, CP2, CP5, CP6, TP9, TP10, Pz, P3, P4, O1 and O2.

'Geïnformeerde toestemming'

Titel van het onderzoek:
EEG Auditory attention detection with neurofeedback training in a home environment

Naam + contactgegevens onderzoeker:
Rob Zink
Researcher, STADIUS, KU Leuven
E: rob.zink@esat.kuleuven.be
T: +32 16 32 17 99 KU Leuven

Methodologie van het onderzoek:
Zie 'Uitnodiging om deel te nemen aan onderzoek voor het verzamelen van EEG data'

Duur van het experiment:
+/- 2 uur tijdens een kalibratiesessie
+/- 2 uur elke dag gedurende 4 dagen binnen een periode van 8 dagen (trainingsfase)
+/- 2 uur elke dag gedurende 2 opvolgdagen ongeveer één, respectievelijk twee weken na de laatste dag van de trainingsfase

Gelieve aan te kruisen indien akkoord:

☐ Ik begrijp wat van mij verwacht wordt tijdens dit onderzoek.

☐ Ik stem in met het dragen van de electrodenkap. Ik begrijp dat ik wellicht enkele gevoelige plekjes krijg op mijn hoofdhuid die binnen enkele uren vanzelf verdwijnen.

☐ Ik geef toestemming om het onderzoek te laten doorgaan bij mij thuis en geef toestemming aan de onderzoeker om in mijn woning aanwezig te zijn gedurende het onderzoek.

☐ Ik weet dat er risico's of ongemakken kunnen verbonden zijn aan mijn deelname:
- ☐ Mijn haar kan vochtig en verward zijn na het onderzoek.

☐ Ik bevestig dat ik het formulier 'Uitnodiging om deel te nemen aan onderzoek voor het verzamelen van EEG data' (pagina-1) heb gelezen en eventuele onduidelijkheden heb gevraagd aan de onderzoeker.

☐ Ik heb normaal gezichtsvermogen (eventueel met bril) en geen beschadiging aan mijn gehoor.

☐ Ik neem uit vrije wil deel aan dit onderzoek.

☐ De resultaten van dit onderzoek kunnen gebruikt worden voor wetenschappelijke doeleinden en mogen gepubliceerd worden.

☐ Mijn naam wordt daarbij niet gepubliceerd, anonimiteit en de vertrouwelijkheid van de gegevens is in elk stadium van het onderzoek gewaarborgd.

☐ Ik behoud het recht om op elk moment mijn deelname aan het onderzoek stop te zetten en ik weet dat daaruit geen nadeel voor mij mag ontstaan.

☐ Voor eventuele vragen, klachten, verdere opvolging, weet ik dat ik na mijn deelname terecht kan bij:
- ☐ Rob Zink - rob.zink@esat.kuleuven.be
- ☐ smec@kuleuven.be

Ik heb bovenstaande informatie gelezen en begrepen en heb antwoord gekregen op al mijn vragen betreffende deze studie. Ik stem toe om deel te nemen.

Datum:
Naam en handtekening proefpersoon   Naam en handtekening onderzoeker

3

Figure A.2: Informed Consent, approved by the ethical committee of the KU Leuven and signed by all subjects.

Sessie: CS1–CS2–TS1-TS2-TS3-TS4–FUS1–FUS2 Name: ________________

Block: Ear: Left/Right Date: ________________

Stream D: Baba Jaga – De Duivel met de 3 gouden haren – Het gouden slakkenhuisje

Part 1

1. Wat moest het meisje van haar stiefmoeder gaan vragen aan haar tante?
    a. brood
    b. een hemdje
    c. naald&draad
    d. Weet ik niet

2. Wat heeft de kater gekregen van het meisje?
    a. kaas
    b. melk
    c. ham
    d. Weet ik niet

3. Wat heeft ze aan de honden gegeven?
    a. botje
    b. vlees
    c. brood
    d. Weet ik niet

4. Wat kwam NIET voor in dit fragment?
    a. een mes
    b. olie
    c. een handdoek
    d. Weet ik niet

Op een schaal van 1 (heel saai/heel makkelijk) tot 10 (heel boeiend/heel moeilijk):

- Hoe moeilijk vond je het om je te concentreren op het juiste verhaaltje?
- Hoe boeiend vond je het verhaaltje?

Figure A.3: Example of a questionnaire that subjects filled out after each block.

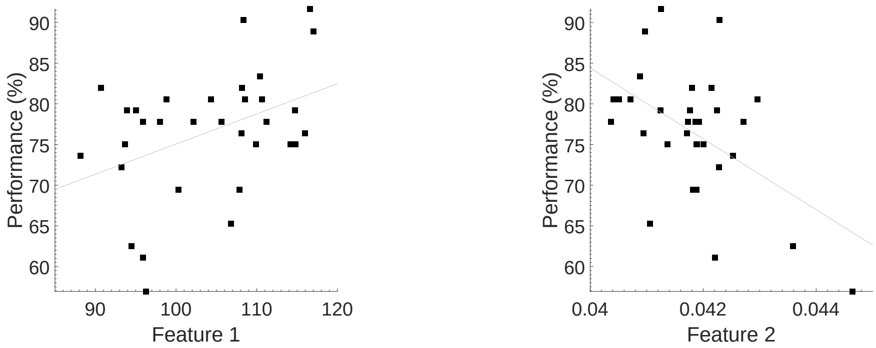
Appendix B

Illustrations to Section 2.5

B.1 The Influence of the Audio Stimulus on AAD Performance

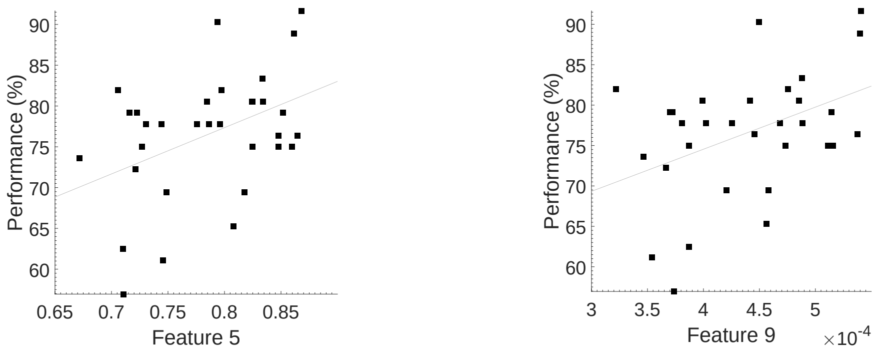
Table B.1: More information about the audio features. For more detailed information, we kindly refer the reader to [46].

N	Feature	Information [46]
1	LineLength	Line length
2	RMSamp	Root Mean Square value
3	Slope	Slope coefficient of a fitted polynomial of 1^{st} degree
4	Activity	Hjorth parameter
5	Mobility	Hjorth parameter
6	Complexity	Hjorth parameter
7	Kurtosis	Kurtosis
8	Skewness	Skewness
9	NonLinEnergy	Teager Energy Operator [47]
10	ZeroCrossings	Number of zero crossings
11	Minima	Minimum value
12	Maxima	Maximum value
13	ARModErr	Autoregressive model of order 1
14	VarFirstDer	Variance of the first derivative
15	VarSecDer	Variance of the second derivative
16	ZCFirstDer	Number of zero crossings of the first derivative
17	ZCSecondDer	Number of zero crossings of the second derivative
18	Mean	Mean value



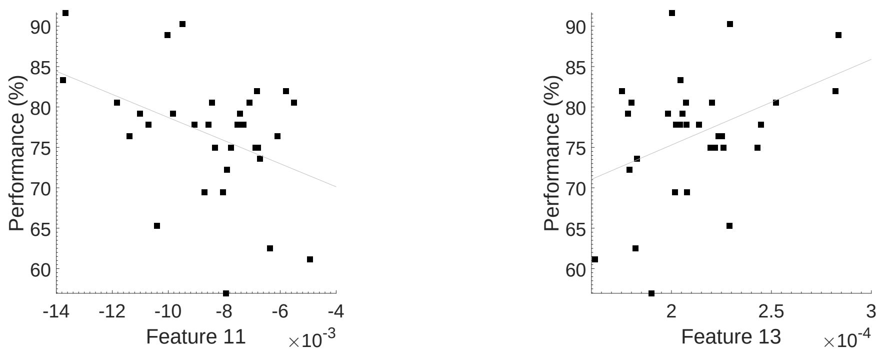
(a) (b)

Figure B.1: Influence of Audio Features 1 and 2 on Performance.



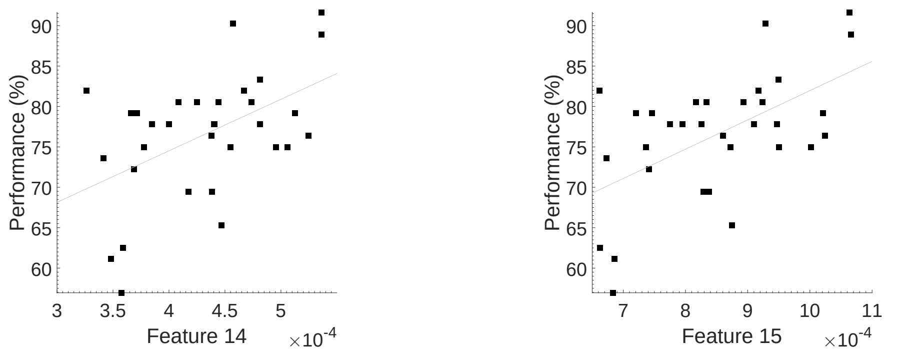
(a) (b)

Figure B.2: Influence of Audio Features 5 and 9 on Performance.



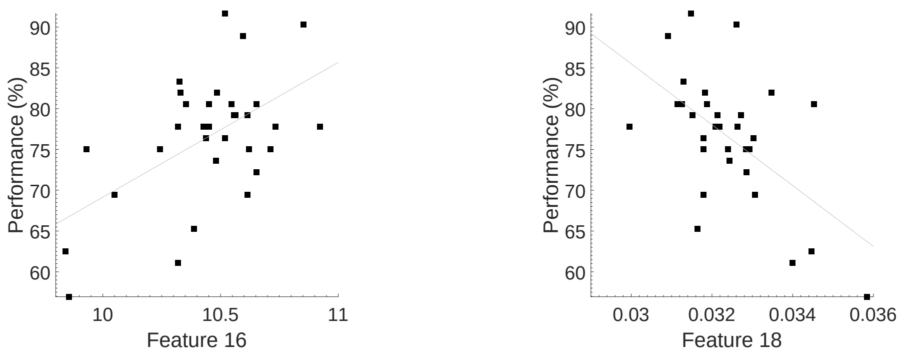
(a) (b)

Figure B.3: Influence of Audio Features 11 and 13 on Performance.



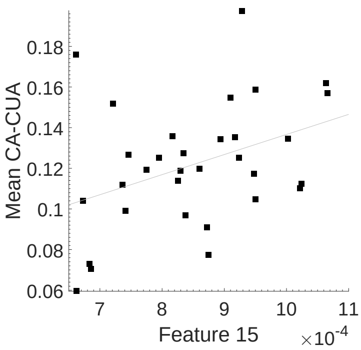
(a) (b)

Figure B.4: Influence of Audio Features 14 and 15 on Performance.

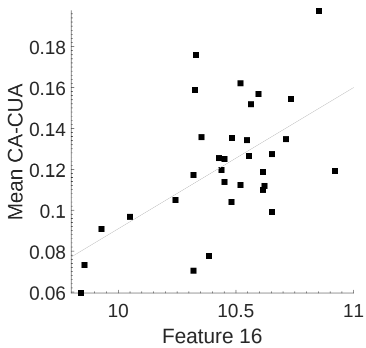


(a) (b)

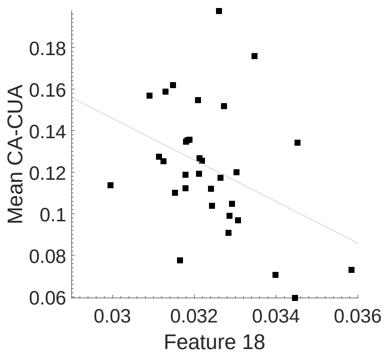
Figure B.5: Influence of Audio Features 16 and 18 on Performance.



(a)



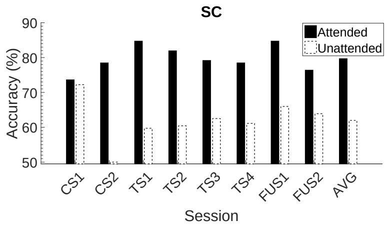
(b)



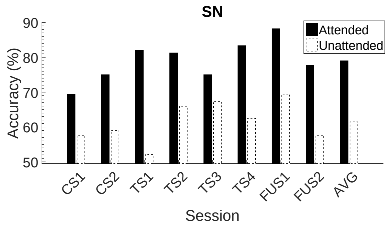
(c)

Figure B.6: Influence of Audio Features 15, 16 and 18 on Mean CA-CUA.

B.2 Unattended Decoders



(a)



(b)

Figure B.7: Session Performance with Attended vs. Unattended decoder for a) SC and b) SN.

Bibliography

[1] 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, 2015.

[2] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos, "Decoding the attended speech stream with multi-channel eeg: implications for online, daily-life applications," *Journal of Neural Engineering*, vol. 12, no. 4, p. 9, 2015.

[3] N. Das, W. Biesmans, A. Bertrand, and T. Francart, "The effect of head-related filtering and ear-specific decoding bias on auditory attention detection," *Journal of Neural Engineering*, vol. 13, no. 5, p. 9, 2016.

[4] W. Biesmans, N. Das, T. Francart, and A. Bertrand, "Auditory-inspired speech envelope extraction methods for improved eeg-based auditory attention detection in a cocktail party scenario," *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 5, pp. 402–412, 2017.

[5] A. Aroudi, B. Mirkovic, M. De Vos, S. Doclo, and Ieee, "Auditory attention decoding with eeg recordings using noisy acoustic reference signals," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, International Conference on Acoustics Speech and Signal Processing ICASSP, (NEW YORK), pp. 694–698, Ieee, 2016.

[6] 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, p. 11, 2016.

[7] 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.

[8] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *NeuroImage*, 2017.

[9] R. Zink, A. Baptist, A. Bertrand, S. Van Huffel, and M. De Vos, "Online detection of auditory attention in a neurofeedback application," in *Proc. 8th International Workshop on Biosignal Interpretation*, no. accepted, 2016.

[10] C. Cherry and J. A. Bowles, "Contribution to a study of the cocktail party problem," *Journal of the Acoustical Society of America*, vol. 32, no. 7, pp. 884–884, 1960.

[11] E. M. Kaya and M. Elhilali, "Modelling auditory attention," *Philosophical Transactions of the Royal Society B-Biological Sciences*, vol. 372, no. 1714, p. 10, 2017.

[12] S. J. Aiken and T. W. Picton, "Human cortical responses to the speech envelope," *Ear and Hearing*, vol. 29, no. 2, pp. 139–157, 2008.

[13] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, pp. 233–U118, 2012.

[14] 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.

[15] J. Z. Simon, "The encoding of auditory objects in auditory cortex: Insights from magnetoencephalography," *International Journal of Psychophysiology*, vol. 95, no. 2, pp. 184–190, 2015.

[16] M. Haghighi, M. Moghadamfalahi, M. Akcakaya, and D. Erdogmus, "Eeg-assisted modulation of sound sources in the auditory scene," *arXiv preprint arXiv:1612.00703*, 2016.

[17] E. Alickovic, T. Lunner, F. Gustafsson, and Ieee, "A system identification approach to determining listening attention from eeg signals," in *24th European Signal Processing Conference (EUSIPCO)*, European Signal Processing Conference, (NEW YORK), pp. 31–35, Ieee, 2016.

[18] S. Akram, A. Presacco, J. Z. Simon, S. A. Shamma, and B. Babadi, "Robust decoding of selective auditory attention from meg in a competing-speaker environment via state-space modeling," *Neuroimage*, vol. 124, pp. 906–917, 2016.

[19] C. Neuper and G. Pfurtscheller, *Neurofeedback Training for BCI Control*, pp. 65–78. Frontiers Collection, Berlin: Springer-Verlag Berlin, 2010.

[20] S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, "Eeg-neurofeedback as a tool to modulate cognition and behavior: A review tutorial," *Frontiers in Human Neuroscience*, vol. 11, p. 19, 2017.

[21] J. Rogala, K. Jurewicz, K. Paluch, E. Kublik, R. Cetnarski, and A. Wrobel, "The do's and don'ts of neurofeedback training: A review of the controlled studies using healthy adults," *Frontiers in Human Neuroscience*, vol. 10, p. 12, 2016.

[22] R. Sitaram, T. Ros, L. Stoeckel, S. Haller, F. Scharnowski, J. Lewis-Peacock, N. Weiskopf, M. L. Blefari, M. Rana, E. Oblak, N. Birbaumer, and J. Sulzer, "Closed-loop brain training: the science of neurofeedback," *Nature Reviews Neuroscience*, vol. 18, no. 2, pp. 86–100, 2017.

[23] H. Marzbani, H. R. Marateb, and M. Mansourian, "Methodological note: Neurofeedback: A comprehensive review on system design, methodology and clinical applications," *Basic and Clinical Neuroscience*, vol. 7, no. 2, pp. 143–158, 2016.

[24] C. Zich, M. De Vos, C. Kranczioch, and S. Debener, "Wireless eeg with individualized channel layout enables efficient motor imagery training," *Clinical Neurophysiology*, vol. 126, no. 4, pp. 698–710, 2015.

[25] C. Zich, C. Schweinitz, S. Debener, C. Kranczioch, and Ieee, "Multimodal evaluation of motor imagery training supported by mobile eeg at home: A case report," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE International Conference on Systems Man and Cybernetics Conference Proceedings, (LOS ALAMITOS), pp. 3181–3186, Ieee Computer Soc, 2015.

[26] M. O. Krucoff, S. Rahimpour, M. W. Slutzky, V. R. Edgerton, and D. A. Turner, "Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation," *Frontiers in Neuroscience*, vol. 10, p. 23, 2016.

[27] M. Arns, S. de Ridder, U. Strehl, M. Breteler, and A. Coenen, "Efficacy of neurofeedback treatment in adhd: the effects on inattention, impulsivity and hyperactivity: a meta-analysis," *Clinical Eeg and Neuroscience*, vol. 40, no. 3, pp. 180–189, 2009.

[28] M. B. Sterman and T. Egner, "Foundation and practice of neurofeedback for the treatment of epilepsy," *Applied Psychophysiology and Biofeedback*, vol. 31, no. 1, pp. 21–35, 2006.

[29] M. S. Shin, H. Jeon, M. Kim, T. Hwang, S. J. Oh, M. Hwangbo, and K. J. Kim, "Effects of smart-tablet-based neurofeedback training on cognitive function in children with attention problems," *Journal of Child Neurology*, vol. 31, no. 6, pp. 750–760, 2016.

[30] H. J. Hwang, K. Kwon, and C. H. Im, "Neurofeedback-based motor imagery training for brain-computer interface (bci)," *Journal of Neuroscience Methods*, vol. 179, no. 1, pp. 150–156, 2009.

[31] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the Ieee*, vol. 89, no. 7, pp. 1123–1134, 2001.

[32] G. Pfurtscheller, C. Neuper, G. R. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and

C. Schrank, "Graz-bci: State of the art and clinical applications," *Ieee Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 177–180, 2003.

[33] G. Pfurtscheller, C. Brunner, R. Leeb, R. Scherer, G. R. Muller-Putz, and C. Neuper, *The Graz Brain-Computer Interface*, pp. 79–96. Frontiers Collection, Berlin: Springer-Verlag Berlin, 2010.

[34] S. S. Yoo, J. H. Lee, H. O'Leary, V. Lee, S. E. Choo, and F. A. Jolesz, "Functional magnetic resonance imaging-mediated learning of increased activity in auditory areas," *Neuroreport*, vol. 18, no. 18, pp. 1915–1920, 2007.

[35] R. Milner, M. Lewandowska, M. Ganc, K. Ciesla, I. Niedzialek, and H. Skarzynski, "Slow cortical potential neurofeedback in chronic tinnitus therapy: A case report," *Applied Psychophysiology and Biofeedback*, vol. 41, no. 2, pp. 225–249, 2016.

[36] C. Jeunet, B. N'Kaoua, and F. Lotte, *Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates*, vol. 228 of *Progress in Brain Research*, pp. 3–35. Amsterdam: Elsevier Science Bv, 2016.

[37] C. Jeunet, B. N'Kaoua, S. Subramanian, M. Hachet, and F. Lotte, "Predicting mental imagery-based bci performance from personality, cognitive profile and neurophysiological patterns," *Plos One*, vol. 10, no. 12, p. 21, 2015.

[38] M. Wostmann, B. Herrmann, B. Maess, and J. Obleser, "Spatiotemporal dynamics of auditory attention synchronize with speech," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 14, pp. 3873–3878, 2016.

[39] Y. Y. Gao, Q. Wang, Y. Ding, C. M. Wang, H. F. Li, X. H. Wu, T. S. Qu, and L. Li, "Selective attention enhances beta-band cortical oscillation to speech under "cocktail-party" listening conditions," *Frontiers in Human Neuroscience*, vol. 11, p. 10, 2017.

[40] W. Noble, N. S. Jensen, G. Naylor, N. Bhullar, and M. A. Akeroyd, "A short form of the speech, spatial and qualities of hearing scale suitable for clinical use: The ssq12," *International Journal of Audiology*, vol. 52, no. 6, pp. 409–412, 2013.

[41] R. Bauer, M. Vukelic, and A. Gharabaghi, "What is the optimal task difficulty for reinforcement learning of brain self-regulation?," *Clinical Neurophysiology*, vol. 127, no. 9, pp. 3033–3041, 2016.

[42] J. M. Gottselig, G. Hofer-Tinguely, A. A. Borbely, S. J. Regel, H. P. Landolt, J. V. Retey, and P. Achermann, "Sleep and rest facilitate auditory learning," *Neuroscience*, vol. 127, no. 3, pp. 557–561, 2004.

[43] M. Hiscock and M. Kinsbourne, “Attention and the right-ear advantage: What is the connection?,” *Brain and Cognition*, vol. 76, no. 2, pp. 263–275, 2011.

[44] E. B. Petersen, M. Wostmann, J. Obleser, and T. Lunner, “Neural tracking of attended versus ignored speech is differentially affected by hearing loss,” *Journal of Neurophysiology*, vol. 117, no. 1, pp. 18–27, 2017.

[45] C. E. Shannon and W. Weaver, “The mathematical theory of communication,” 2002.

[46] N. Koolen, K. Jansen, J. Vervisch, V. Matic, M. De Vos, G. Naulaers, and S. Van Huffel, “Line length as a robust method to detect high-activity events: Automated burst detection in premature eeg recordings,” *Clinical Neurophysiology*, vol. 125, no. 10, pp. 1985–1994, 2014.

[47] I. Antoniadou, G. Manson, N. Dervilis, T. Barszcz, W. J. Staszewski, and K. Worden, “Use of the teager-kaiser energy operator for condition monitoring of a wind turbine gearbox,” *Proceedings of International Conference on Noise and Vibration Engineering (Isma2012) / International Conference on Uncertainty in Structural Dynamics (Usd2012)*, pp. 4255–4268, 2012.

[48] S. Halder, M. Rea, R. Andreoni, F. Nijboer, E. M. Hammer, S. C. Kleih, N. Birbaumer, and A. Kubler, “An auditory oddball brain-computer interface for binary choices,” *Clinical Neurophysiology*, vol. 121, no. 4, pp. 516–523, 2010.

[49] D. W. Kim, H. J. Hwang, J. H. Lim, Y. H. Lee, K. Y. Jung, and C. H. Im, “Classification of selective attention to auditory stimuli: Toward vision-free brain-computer interfacing,” *Journal of Neuroscience Methods*, vol. 197, no. 1, pp. 180–185, 2011.

[50] D. J. Mitchell, N. McNaughton, D. Flanagan, and I. J. Kirk, “Frontal-midline theta from the perspective of hippocampal quot;thetaquot;,” *Progress in Neurobiology*, vol. 86, no. 3, pp. 156–185, 2008.

[51] K. Inanaga, “Frontal midline theta rhythm and mental activity,” *Psychiatry and Clinical Neurosciences*, vol. 52, no. 6, pp. 555–566, 1998.

[52] T. Demiralp and E. Basar, “Theta-rhythmicities following expected visual and auditory targets,” *International Journal of Psychophysiology*, vol. 13, no. 2, pp. 147–160, 1992.