Selective Attention Tracking

Neural Markers in Listeners Across Hearing Conditions

Keding, Oskar

2026

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

– CENTRUM SCIENTIARUM MATHEMATICARUM –

Selective Attention Tracking

Neural Markers in Listeners Across Hearing Conditions

OSKAR KEDING

Lund University
Faculty of Engineering
Centre for Mathematical Sciences
Mathematical Statistics

Selective Attention Tracking
Neural Markers in Listeners Across Hearing Conditions

Selective Attention Tracking
Neural Markers in Listeners Across Hearing Conditions

Oskar Keding

Lund University

To be presented, with the permission of the Faculty of Engineering of Lund University, for public criticism in lecture hall MH:R at the Centre for Mathematical Sciences on Wednesday, the 27th of May 2026 at 09:15.

Organization
LUND UNIVERSITY
Centre for Mathematical Sciences
Mathematical Statistics
Box 118
SE–221 00 Lund, Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2026-05-27

Sponsoring organization
The research in this thesis was supported by the ELLIIT strategic research environment.

Author(s)
Oskar Keding

Title and subtitle
Selective Attention Tracking— Neural Markers in Listeners Across Hearing Conditions

Abstract
Understanding how the human brain selectively attends to a single speaker in multi-speaker environments is central to both auditory neuroscience and the development of hearing technologies. A promising approach for addressing these challenges, and learning how the brain encodes speech, is auditory attention decoding (AAD), which aims to infer a listener's attended speech stream from measurements of neural activity, such as electroencephalography (EEG).

Challenges within the field of AAD includes developing reliable and usable measure for selective attention in wearers. These methods should be both applicable in wearable EEG and robustly decode attention. Additionally, these methods should not bias decisions without necessity. Furthermore, the spectrum of listening scenarios is broad. AAD methods should be applicable to a wide variety of languages, talkers, acoustic listening conditions, and listening tasks such as switching attention between conversations of interest. Solutions should also work for both normal hearing and hearing aid users, as their acoustic experiences differ.

This thesis develops and evaluates methods for tackling some of these challenges. The work investigates how preprocessing choices affect AAD performance, proposes model-based methods for quantifying relationships between continuous speech signals and concurrent neural recordings. The work also evaluates spatial feature extraction approaches for decoding the directional focus of auditory attention. Furthermore, a novel dataset is introduced, capturing more ecologically valid listening conditions including conversational speech, attention switching, and mobile EEG recordings. The generalization of state-of-the-art AAD methods to these settings is examined. Finally, methods for tracking auditory attention dynamically over time are explored as a step toward real-time applicability. The results highlight both the potential and current challenges of integrating neural attention decoding into future hearing aid technology, with the long-term goal of understanding and enabling selective attention in varied acoustic scenes.

Key words
Electroencephalography (EEG), Signal Processing, Spectral Coherence, Auditory Attention Decoding (AAD), Neural Speech Tracking

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
1404-0034
Doctoral Theses in Mathematical Sciences 2026:6

ISBN
978-91-8104-982-4 (print)
978-91-8104-983-1 (pdf)

Recipient's notes

Number of pages
xxi+164

Price

Security classification

Signature ______________________ Date 2026-04-13

Selective Attention Tracking
Neural Markers in Listeners Across Hearing Conditions

Oskar Keding

Doctoral Thesis
Thesis advisor: Maria Sandsten
Thesis co-advisors: Emina Alickovic, Martin A. Skoglund
Faculty opponent: Prof. Edmund C. Lalor

Cover illustration front

Five women stand gathered in a circle, with the central figure listening attentively as the others converse around her. Credits: Adapted from *Samtalande Kvinder* by Christoffer Wilhelm Eckersberg, sourced from the Nationalmuseum collection (`https://collection.nationalmuseum.se/sv/collection/item/214087/`).

Published by

Centre for Mathematical Sciences
Mathematical Statistics
Faculty of Engineering
Lund University
Lund 2026

ISBN: 978-91-8104-982-4 (print)
ISBN: 978-91-8104-983-1 (pdf)
ISSN: 1404-0034

Printed in Sweden by Media-Tryck, Lund University, Lund 2026

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

Printed matter 3041 0903

MADE IN SWEDEN

When people talk, listen completely. Most people never listen.
— Ernest Hemingway

Abstract

Understanding how the human brain selectively attends to a single speaker in multi-speaker environments is central to both auditory neuroscience and the development of hearing technologies. A promising approach for addressing these challenges, and learning how the brain encodes speech, is auditory attention decoding (AAD), which aims to infer a listener's attended speech stream from measurements of neural activity, such as electroencephalography (EEG).

Challenges within the field of AAD includes developing reliable and usable measure for selective attention in wearers. These methods should be both applicable in wearable EEG and robustly decode attention. Additionally, these methods should not bias decisions without necessity. Furthermore, the spectrum of listening scenarios is broad. AAD methods should be applicable to a wide variety of languages, talkers, acoustic listening conditions, and listening tasks such as switching attention between conversations of interest. Solutions should also work for both normal hearing and hearing aid users, as their acoustic experiences differ.

This thesis develops and evaluates methods for tackling some of these challenges. The work investigates how preprocessing choices affect AAD performance, proposes model-based methods for quantifying relationships between continuous speech signals and concurrent neural recordings. The work also evaluates spatial feature extraction approaches for decoding the directional focus of auditory attention. Furthermore, a novel dataset is introduced, capturing more ecologically valid listening conditions including conversational speech, attention switching, and mobile EEG recordings. The generalization of state-of-the-art AAD methods to these settings is examined. Finally, methods for tracking auditory attention dynamically over time are explored as a step toward real-time applicability. The results highlight both the potential and current challenges of integrating neural attention decoding into future hearing aid technology, with the long-term goal of understanding and enabling selective attention in varied acoustic scenes.

Publications

Papers concerning the work of this thesis have been made as follows:

I **Effect of Independent Component Artifact Rejection on EEG-Based Auditory Attention Decoding**
Oskar Keding, Johanna Wilroth, Martin A. Skoglund, Emina Alickovic
32nd European Signal Processing Conference (EUSIPCO), Lyon, France, 2024

II **Novel bias-reduced coherence measure for EEG-based speech tracking in listeners with hearing impairment**
Oskar Keding, Emina Alickovic, Martin A. Skoglund, Maria Sandsten
Front. Neurosci., Volume 18 - 2024

III **Neural Tracking of Sustained Attention, Attention Switching, and Natural Conversation in Audiovisual Environments using Wearable EEG**
[1] Johanna Wilroth, **Oskar Keding**, Martin A. Skoglund, Maria Sandsten, Martin Enqvist, Emina Alickovic
Submitted Manuscript

IV **A Comparison of Common Spatial Patterns Algorithms for Auditory Attention Decoding**
Oskar Keding, Emina Alickovic, Martin A. Skoglund, Maria Sandsten
Submitted Manuscript

V **Dynamic Audiovisual Selective Attention Tracking, with Attention Switching**
Oskar Keding, Martin A. Skoglund, Emina Alickovic, Maria Sandsten
Manuscript

[1] Johanna Wilroth and Oskar Keding contributed equally to the work.

Publications not included in this thesis:

A **Robust Phase Difference Estimation of Transients in High Noise Levels**

Oskar Keding, Maria Sandsten
30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 2022, pp. 2271-2275

B **Coherence Expectation Minimisation and Combining Weighted Multitaper Estimates**
Oskar Keding, Maria Åkesson, Maria Sandsten
31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 2023

C **Highly Accurate and Noise-Robust Phase Delay Estimation using Multitaper Reassignment**
Maria Åkesson, **Oskar Keding**, Maria Sandsten
31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 2023, pp. 1763-1767

D **Coherence Estimation Tracks Auditory Attention in Listeners with Hearing Impairment**
Oskar Keding, Emina Alickovic, Martin A. Skoglund, Maria Sandsten
Proc. INTERSPEECH 2023, 2023, pp. 5162-5166

E **Eye Tracking-Based Speech Label Estimation for Auditory Attention Decoding with Portable EEG**
Johanna Wilroth, **Oskar Keding**, Martin A Skoglund, Emina Alickovic, Martin Enqvist
2025 28th International Conference on Information Fusion (FUSION), 2025

Acknowledgements

During my master thesis, I had no plan to pursue a PhD. As a matter of fact, I even thought it was unlikely that I would ever do so. I was excited to finish my studies and start working in industry. However, after starting my master thesis, I was inspired to give it a try. Maria encouraged me to apply, and what a fun and challenging journey this has been.

After these five years, I am grateful for this experience. I am also proud to have at least managed fairly these years. It has been a transformative experience, both professionally and personally. Now I look forward to the next chapter in my life, hopefully able to give back to others. But before that, I would like to express many thanks to those who have supported me in various ways along the way.

First and foremost, I would like to thank those who have lent me their wisdom and expertise during my PhD studies. Their guidance have been invaluable to me. Thank you Maria Sandsten, my main supervisor, for your unwavering support, calmness and coolness when I have needed it the most. Your insights, both research-wise and in broader perspectives, have helped me grow as a researcher and professional. Thank you Emina Alickovic, for guiding me through the cross-section between auditory neuroscience and signal processing. Your enthusiasm for the topic has been contagious and given experiences that would never have been possible otherwise. Thank you Martin Skoglund, for your guidance and insights into what is important in research. Your ability to see the translational value of research while also keeping a levity in everyday research has been inspiring.

I would also like to thank those who have inspired me along the way, both within and outside my field of research. Johanna Wilroth, who I have worked most closely to through these years. Your passion for the topic of neuroscience is contagious. All the trips and travels together have been very fun and definitely enlightened these years. David Sundström, whom I have shared an office with. Not only have you been a great colleague who is always up for talks about anything between heaven and earth, but also a great friend. Also, you continuously remind me how fun it is to do research. Your outlook on life and work is certainly admirable. Thank you Sofia Kockum, Maria Juhlin and Maria Åkesson for sharing the ups and downs of the PhD studies, certainly making me feel like I fit in in academia. Thank you Satyaki, John, Shokoufa, Rahul, Niloofar, Klara and Jonatan. As PhD students at Mathematical Statistics with very different backgrounds, you show how and why research can be done in so many different ways. You have also brightened up the many long days at the department. Thank you to David Ohlin, for inspiring me to start as a PhD student in the first place, and your continued friendship during these years.

A PhD position is in part job, but also importantly an education. There is an uncountable number of persons who I have learned from, but here is the best list I can muster. Thank you to Martin Enqvist, for the guidance, important insights and broad perspectives within our research project. Thank you to Pex, Martin, and Frida for exciting discussions and idea sharing during our meetings early on in my PhD studies. Thank you to Magnus, Andreas, Johan, Filip, Linda, Erik and other professors at the department for all that you have taught me during these years. Both within statistics and about everything else! Not only that, it has been fun to listen to you in the class room as well as the lunch room. Also, thank you Behtash Babadi, for graciously lending your time and (to me) new research perspectives during my visit to UMD.

A big thanks to those who have helped me during these years. Thank you Susann for always supporting and knowing what to do when I come with questions. Thank you Nils for helping me each time I have come to you with issues. Thank you James and Michael for solving anything related to IT. Thank all other administrative staff at the department for making everything not only run but run smoothly, and making coffee breaks nice.

Also thank you to everybody at Eriksholm Research Centre, for letting Johanna and I into your labs and helping us with everything from experimental design to practicalities of data collection. Thanks especially to Michael, Jette, Hamish, Lorenz and Kasper for lending time and experience.

Lastly a big thanks to all the colleagues at Mathematical Statistics for making a friendly workplace, and to all PhD student colleagues in the math house, who have made the years more fun.

Finally, I must thank those who have supported me emotionally and practically throughout these years. Thank you to my friends outside of academia, for love and beers. Thank you to my mother and father for always believing in me and giving me the opportunities that I have gotten. Your support has meant the world to me. Thank you to Astrid, my sister, for careless kindness regardless of what I do. This is very liberating for me. The largest thanks of all go to Jonna, for your endless patience and support. You, who remind me of what is important in life, have stood by me through thick and thin, and I could not have done this without you.

Funding

Work part of this thesis has been funded by the ELLIIT strategic research environment, as part of project C04 with project name Brain-Based Monitoring of Sound.

Popular summary

As kids we are taught that the ear is the organ that does the hearing in our body. In reality, this is not the case. Our perception of what we hear is a result of cognitive processing in the brain. In everyday life, auditory experiences are highly varied. One particularly straining scene is when multiple speakers are present in the same room as the person you are trying to listen to. This can happen in meetings at work, during morning commutes or at a cocktail-party. The last example has inspired the name of the natural ability of our auditory system to subliminally focus on a particular speaker: the *cocktail party effect*.

Although picking out single speakers in noisy environments can be straining, normal hearing individuals are often able to solve the cocktail party problem. However, for people with deteriorating hearing and hearing aid users this ability can be heavily encumbered. Tiredness and wanting to disconnect is common for hearing aid users, where the user turns the device off after being in a multiple speaker situation for too long. This is because the cocktail party effect does not come as naturally. Perhaps you even have experienced this with an older relative.

This thesis develops different methods for looking at measurements of brain activity when people are attending specific people in multi-speaker scenarios. We can measure brain responses using electroencephalography (EEG), which measures electrical neural activity. The connection between the signals in the brain and speech is different for attended and ignored speech, which makes it possible to work out where a listeners attention is targeted, and this practice is called auditory attention decoding (AAD).

First, in Paper I we look at how much of an influence different variations of EEG cleaning has on results within AAD. The second paper of this thesis, Paper II, a method is developed for detecting connections between the speech we hear and concurrent measured brain signals, without imposing too much prior assumptions about the data. Work in Paper IV focuses on analyzing data from a slightly different perspective, by instead looking at the problem: "to which spatial angle is a person currently attention?" In the paper, we compare performance of spatial feature extraction methodology common in EEG, on different online-available datasets.

In Paper III a novel dataset is introduced, showing that state-of-the-art methods within AAD generalizes to scenarios, including speaker in conversations, attention switching, and mobile brain signal measuring techniques. Finally, in Paper V we attempt to take step further towards application in hearing aid technology, by developing methods for analyzing the auditory attention of a listener over time.

If this methodology becomes both real-time and robust, it could be applied to improve the living experience for hearing aid users, and therefore simplify for these individuals participation in society. However, these AAD methods will never work in listening scenarios where the noise is too loud, or too similar to the sounds one is trying to attend. In these cases, hearing aids will fail, like how older hearing aids fail in an even wider range of cases. Nevertheless, these contributions to AAD research will hopefully widen the range of scenarios where hearing aids are useful.

Looking forward, there are many challenges before these models can be applied to help hearing aid users. For example, there is a further need for methods that can track how auditory attention dynamically changes over time and scenarios, as well as taking into account all the hard-to-predict peculiarities of everyday life.

Populärvetenskaplig sammanfattning

När vi är barn lär vi oss att öronen är de organ som sköter hörseln i vår kropp. I verkligheten är detta inte fallet. Vår uppfattning av vad vi hör är ett resultat av kognitiv bearbetning i hjärnan. I vardagen kan hörselupplevelser vara mycket varierande. Ett särskilt ansträngande scenario är när flera talare är närvarande i samma rum som personen du försöker lyssna på. Detta kan hända under möten på jobbet, under morgonpendling eller vid ett cocktailparty. Det sista exemplet har inspirerat namnet på den naturliga förmågan hos vårt hörselsystem att subliminalt fokusera på en viss talare: *cocktailparty-effekten*. Även om det kan vara ansträngande att välja ut enskilda talare i bullriga miljöer, kan oftast normalhörande personer lösa cocktailpartyproblemet. Men för personer med försämrad hörsel och hörapparatanvändare kan denna förmåga vara kraftigt begränsad. Trötthet och en önskan att koppla bort är vanligt för hörapparatanvändare, där användaren stänger av apparaten efter att ha varit i en situation med flera talare för länge. Detta beror på att cocktailparty-effekten inte kommer lika naturligt. Kanske har du till och med upplevt detta med en äldre släkting.

Denna avhandling utvecklar olika metoder för att undersöka hjärnaktivitet när människor lyssnar på specifika personer i scenarier med flera talare. Vi kan mäta hjärnaktivitet med hjälp av elektroencefalografi (EEG), som mäter elektriska nervaktiveringar. Sambandet mellan hjärnsignaler och tal är olika för tal man lyssnar på jämfört med tal man ej lyssnar på, vilket gör det möjligt att ta reda på var en lyssnares uppmärksamhet riktas. Med nya analysmetoder är det möjligt att spåra, i tid och position, hjärnaktivitet i förhållande till det tal som hörs. Utöver detta är det också möjligt att ta reda på vilken talare en viss person lyssnar på.

Först, i artikel I, undersöks hur stort inflytande olika variationer av EEG-rengöring har på resultaten inom AAD. I artikel II utvecklas en metod för att upptäcka samband mellan det uppfattade talet och samtidiga uppmätta hjärnsignaler, utan att lägga till för många förhandsantaganden om data. I artikel IV analyseras data från ett något annorlunda perspektiv, genom att istället undersöka: vilken rumslig vinkel en person för närvarande är uppmärksam på. I artikeln jämförs prestandan hos metoder, vanliga inom EEG, på olika datamängder, tillgängliga på internet.

Artikel III introducerar en nytt experiment-dataset som visar att metoder inom AAD generaliserar till olika scenarier, med talare i samtal, med uppmärksamhetsväxling och med tekniker för mätning av mobila hjärnsignaler. Slutligen tar vi i artikel V ytterligare ett steg mot tillämpning inom hörapparatteknik, genom att utveckla metoder för att analysera en lyssnares auditiva uppmärksamhet över tid.

Om dessa metoder blir både realtidsbaserade och robusta, skulle de kunna tillämpas för att förbättra livsupplevelsen för hörapparatanvändare och därmed förenkla för dessa individers deltagande i samhället. Dessa AAD-metoder kommer dock aldrig att fungera i lyssningsscenarier där bruset är för högt eller för likt de ljud man försöker uppmärksamma. I dessa fall kommer hörapparater att misslyckas, precis som äldre hörapparater misslyckas i ett ännu större antal fall. Trots detta kommer dessa bidrag till AAD-forskning förhoppningsvis att bredda antalet scenarier där hörapparater är användbara.

Framöver finns det många utmaningar innan dessa modeller kan tillämpas för att hjälpa hörapparatanvändare. Till exempel finns det ytterligare behov av metoder som kan spåra hur hörseluppmärksamhet dynamiskt förändras över tid och scenarier, och som tar hänsyn till alla svårförutsägbara särdrag som uppkommer i våra vardagar.

Contents

List of Abbreviations

AAD auditory attention decoding

AV audiovisual

BCI brain-computer interface

CSP common spatial pattern

EEG electroencephalography

ERP event-related potential

FFT fast Fourier transform

fMRI functional magnetic resonance imaging

fNIRS functional near-infrared spectroscopy

HA hearing aid

HI hearing impaired

ICA independent component analysis

LDA linear discriminant analysis

MEG magnetoencephalography

NH normal hearing

PCA principal component analysis

SNR signal-to-noise ratio

TRF temporal response function

Chapter 1

Introduction to Hearing, Auditory Attention and Neural Decoding

Most people are taught that the ear is the organ that does the hearing in our body. In reality, which has been known for a very long time, this is not the case. Our perception of what we hear is a result of cognitive processing in the brain as well as combination of multiple sensory networks. The biological process from detection of the physical sound to hearing and understanding is complicated to say the least.

In everyday life the auditory environmental experiences are highly varied. One particularly straining scene is when multiple speakers are present in the same room as the person one is trying to listen to. This can happen in meetings at work, during morning commutes or a cocktail party. The last example has inspired the name of the natural ability of our auditory system to subliminally focus on a particular speaker: the *cocktail party effect*, first named by Cherry in the 1950s [1].

Although picking out single speakers in noisy environments can be straining, normal hearing (NH) individuals are often able to solve the cocktail party problem. However, for people with deteriorating hearing and hearing aid (HA) users this ability can be heavily encumbered. To intuitively understand the problem, one can imagine a situation that most people have been in nowadays. In an online video call meeting, it is often the case that multiple people on different ends of the call start talking at the same time. If you try distinguishing a certain speaker from the others, this will require a high amount of effort compared to a normal meeting. This is because all speakers are jumbled in the speakers of the computer, where the visual and auditory scene may not match. At the end of the meeting, you can feel very tired. This tiredness is also quite common for HA users, where the user turns the hearing device off after being in a multiple speaker situation for too long.

HA technology has advanced significantly during the last decades. Modern HAs are able to adaptively filter sounds, attenuating background noise and enhancing speech signals. Looking at the largest HA companies (Sonova, Demant, WS Audiology, GN Store Nord, Starkey), their most modern HAs at the moment leverage techniques such as deep-learning, multi-sensor modalities and processing personalization to enhance user experience. However, with current HAs it is still not known which speaker the wearer is trying to listen to. Thus, if multiple speakers are present, HA can attenuate sounds from certain preset directions, but cannot specifically filter out the speaker attended at the moment. This is where auditory attention decoding (AAD) comes into play.

By working out which sound a HA user is focusing on, one could attenuate other disruptive sounds in the HA, helping the wearer. This requires continuously measuring brain activity, relating this to the sounds of which the individual hears and work out which source to enhance. To decipher this cognitive processing, understanding the processing of sounds in the auditory system is important. There have been many studies that identify and characterize speech-entrainment to natural stimuli, and how this is attenuated/amplified by auditory attention [2–7]. Similar brain responses have been found in hearing impaired (HI) individuals and HA users [8–13]. Furthermore it has been extended to bi-directional models [14, 15], and to non-linear models [16–19].

1 What is the cocktail party effect?

The cocktail party problem was formalized as the task to solve by humans to recognize what one person is saying when others are speaking at the same time. Multiple aspects can alleviate or exacerbate the difficulty of this task. If the attended speech is relatively soft compared to other speakers, then focusing becomes harder. The separation is also harder if speakers are similar, for example in fundamental frequency, accent or talking speed. Although spatial clues and information of the sound is useful for separating speakers, separation is not purely a result of these types of inferences. Interestingly, although it is harder to separate speakers that share the same direction to the listener, it is still very much possible [20].

The auditory cortex is involved in the processing of sounds and speech [21]. In particular this involvement can be measured during listening in multi-talker scenarios [4]. In the primary auditory cortex both attended and ignored sounds can be tracked. Moving from the primary auditory cortex, outwards to other parts of the auditory cortex, the observed connection to sounds is reduced to attended speech alone [22]. In auditory processing of speech in a multi-talker scenario, tracking is stronger in the left hemisphere compared to the right [23]. This is true although there certainly is a bilateral aspect to hearing and listening processes in the brain.

AAD is the process and research field of determining which, out of multiple concurrent speakers, a certain listener is attending. This is done by measuring brain signatures in some way. Multiple analogue approaches are available for AAD. Attention can be spatially decoded, working out the locus of attention, i.e. which direction the person is devoting mental resources to interpret [24–26]. Attention decoding can also focus more about the connection between measured brain responses and the content of what is spoken in each speech source. Significant research has been made relating both acoustic and linguistic speech features to corresponding brain responses [27–30].

2 Hearing, from vibrations to neuronal activity

A brief overview of the biological processes that connect sound to brain activity is given here, although this should by all means not be seen as a complete description of the auditory system. The aim is to convince the reader of the assumption that the connection between sound pressure and electrical signal recorded in the brain is measurable.

Sound is a series of changes in sound pressure that propagate through all mediums [31]. These mechanical pressure waves travel through multiple biological structures before they are converted into signals in the brain. These structures can be seen in the schematic view of the human ear shown in Fig. 1.1. Vibrations in air reach the outer ear and travel into the ear canal. At the end of the ear canal, the sound waves push on the eardrum. The eardrum separates the outer ear and the middle ear. In the middle ear lies three bones, called the *malleus*, *incus* and *stapes*. The eardrum pushes on the malleus and the pressure from sound is propagated through the bones. At the end of the stapes the concentrated pressure (due to a very small surface area) is transferred into the *cochlea*.

When sound enters the fluid-filled cochlea, it enters one side of it [31]. The cochlea is split in two by the *basilar membrane*, which is a tapered wall that separates two tubes that make up the spiral-shaped cochlea. As pressure must be dispersed, sound waves move from one side of the basilar membrane to the other. Because of the tapered structure, different frequencies of sounds will take different paths, separating themselves along the basilar membrane. Within the basilar membrane, these mechanical oscillations of different frequencies are extracted and converted into electrical signals. These electrical signals are then carried by neurons through the auditory nerve to the central auditory pathway. These signals are frequency dependant as well as amplitude dependant of the oscillations caused by sounds.

As the fibres of the auditory nerve leaves each left and right cochlea they enter the corresponding cochlear nucleus, where they bifurcate. Roughly speaking, the nerves travel to a range of different nuclei that make up the cochlear nuclei, before they travel to various parts of the ascending auditory pathway. Signals reach the superior olivary complex, the nuclei of the lateral lemniscus, the inferior colliculus and the medial geniculatae body before arriving to the auditory cortex. The auditory cortex is the most recent evolutionary part of the auditory system and thus is important for processing of sound that is specific to humans [31]. High-level processing of sounds heard is done here in connection to other parts of the brain. It is located

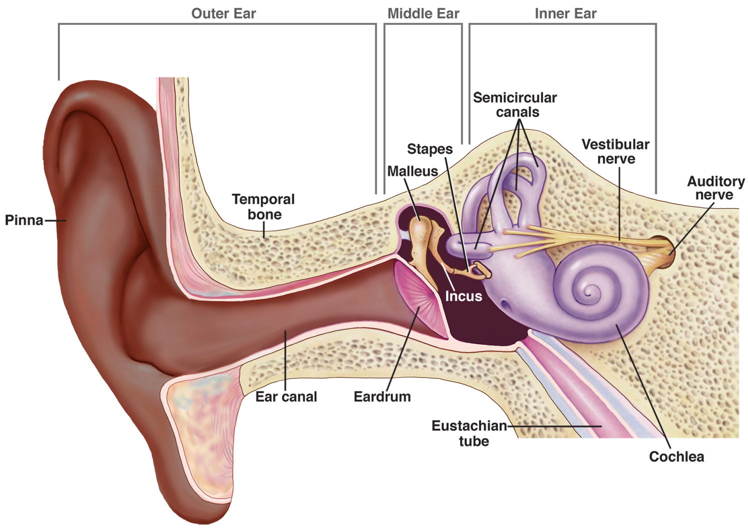

Figure 1.1: Parts that make up the human ear. Taken from the website of the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (public domain).

bilaterally, at the upper sides of the temporal lobes in areas such as the Heschl's gyri and superior temporal gyri. The location of the auditory cortex is shown in Fig. 1.2.

Although this is the main path leading from sound coming into our ears to us actually hearing and listening, other paths and areas are included in the hearing process. These areas can be connected in undefined ways or be connected to specific biological processes, such as a connection from the inferior colliculus to the gaze control centre for reflexive eye movements towards sounds. Additionally, the auditory pathway is not a one way path for sound processing. In reality, feedback loops are incorporated into the biological structures, with descending projections back as far as the cochlea.

Thus, although the entire system is complex, sound is transformed from pressure waves in air, to mechanical vibrations in the ear, to electrical signals in the auditory nerve, to neuronal activity in the brain. Along this path, multiple processing steps are done that extract relevant features of sound for further processing. In AAD, we try to analogously extract features of sounds that are relevant for listening and attention, and relate these to measurable brain activity. Examples of these features are the envelope of speech [5], the fundamental frequency of speech [32], speech onsets or higher-level linguistic features such as phonemes or words [33–35]. Each of these features carry different information about the speech signal, and will therefore have different relations to brain activity. For example, for many of the aforementioned features,

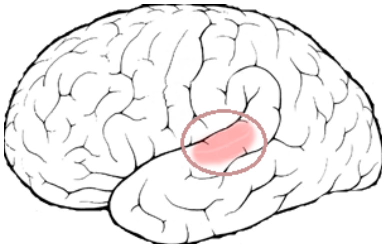

Figure 1.2: Location of the auditory cortex, highlighted in pink, on the surface of the brain. The anterior of the brain is to the left, and the posterior to the right. Source: Author.

they are tracked in the brain less when ignored, compared to when attended. This is the basis for many AAD methods. Features can also be learned with deep learning based approaches, although this is not explored in this thesis.

3 Probing brain activity

Measuring what goes on in the brain is non-trivial, to say the least. Inferring neural activity non-invasively can be done in different ways. Techniques such as functional magnetic resonance imaging (fMRI) measure the blood flow in the brain. Blood flow is coupled to neurological activity, i.e. blood flows to regions that are activated in neuro-processing. The technique of fMRI has a high spatial resolution, but lacks temporal resolution to see fast cortical changes. A more portable technique is functional near-infrared spectroscopy (fNIRS), which measures blood oxygenation changes in the brain, similar to fMRI, but with a wearable set-up. However, fNIRS has a low temporal resolution, and is currently not widely used in AAD research.

Other techniques, such as the electroencephalography (EEG) and magnetoencephalography (MEG), measure electrical activity in the brain. Both techniques have a significantly higher temporal accuracy. Although both techniques are measured outside the head and MEG has better spatial source estimation qualities, EEG is significantly cheaper and easier to set-up. EEG is currently seen as the most probable technique to have use in future HAs. Recently there have also been parallel advancements in EEG to make it more portable, which is suitable for psychophysiological measurements in real-world scenarios [36].

The first human EEG was captured in 1924, by Hans Berger [37]. By placing as many as 32, 64 or 128 electrodes on the scalp, and measuring the electrical potential of each electrode, one can observe a spatial map of electrical potentials at each sample during EEG. An example of an EEG cap is shown in Figure 1.3. The sample rate is in the range of thousands of Hz, although this usually is reduced during analogue to digital conversion. Measuring electrical potential requires setting a reference potential, which is non-trivial. Commonly one designates a reference electrode with minimal contribution from cortical sources or that has a stable

Figure 1.3: In the top, a Biomsei EEG cap (left) with corresponding EEG electrodes (right). The bottom left show cEEGrid arrays with electrode placement around ears (source: `https://commons.wikimedia.org/wiki/File:CeeGRID.jpg`). The bottom right shows an battery driven amplifier for a mobile EEG cap.

baseline. Although in practice no site is fully electrically neutral, but commonly reference electrodes are place at the mastoids, ear lobes or the vertex.

After measurement, data can either be analyzed or visualized. Commonly, EEG activity is interpreted through its spectral signature. Activity is split into the following frequency bands; $\sim [1, 4]$ Hz as delta, $\sim [4, 8]$ Hz as theta, $\sim [8, 12]$ Hz as alpha, $\sim [12, 30]$ Hz as beta, $\sim [30, 64]$ Hz as low gamma, and $\sim [64, 128]$ Hz as high gamma. An an example of an EEG signal and it's decomposition into frequency bands is shown in Fig. 1.4. The EEG caps used for measurements mentioned in this thesis have 32 or 64 channels spread out over the head of the wearer. Generally, they can be grouped into frontal, central, parietal, and occipital, as well as into the left (L) and right (R) hemispheres. In paper IV and V a mobile EEG cap (mBrainTrain Smarting Pro) is used, which allows measurements without being tethered to a stationary amplifier. This allows measurements in more naturalistic scenarios. This mobile set-up is augmented to include cEEgrid arrays around each ear [38], with 10 electrodes each. Example images of both of these technologies are shown in Fig. 1.3. Ideally, similar arrays can be used in future hearing devices to measure brain activity in a compact way.

Although multiple limitations plague EEG analysis, perhaps the main limitation is the poor spatial resolution. EEG cannot sense activation of individual neurons. Instead, EEG captures the synchronized activation of thousands of individual neurons. Thus source estimation is very hard, compared to fMRI. Additionally, EEG is influenced heavily by aperiodic so-called pink noise, $1/f$-shaped in the frequency domain. Recalling Fig. 1.4, the energy of EEG tapers off as the frequency increases. Noise sources originate from ocular activity, heart beats, line

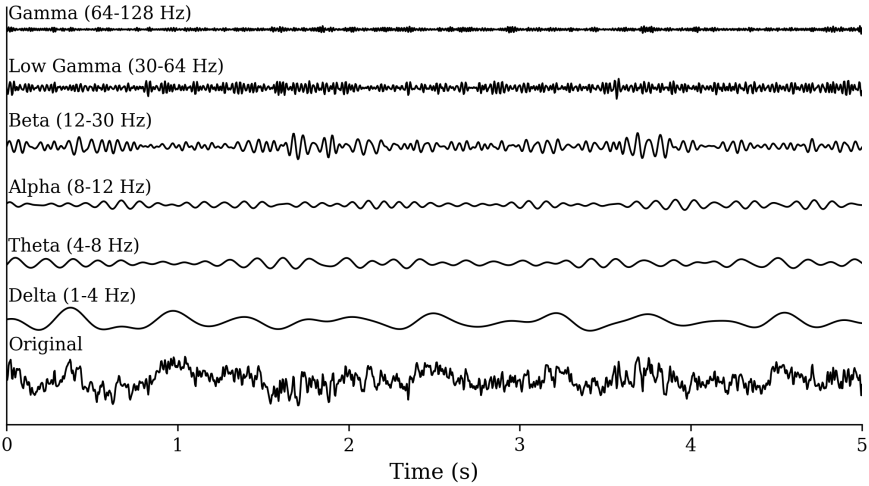

Figure 1.4: An example EEG signal from a single channel is shown (top), together with the same signal filtered to different bands.

noise, channel noise in electrodes, muscles artifacts, as well as brain activity that is not of interest to the particular study or clinical trial [39]. These noise sources disturb measurements both spatially and temporally.

4 Outline of thesis

Within the research field, multiple advancements are needed to understand auditory attention in natural scenarios, and make AAD usable in real-world applications. This thesis presents improvements and answers to some of these challenges.

In Chapter 2, a system perspective is first presented in order to analyze brain responses in relation to auditory stimuli. Mathematical models, widely adopted within the field, are presented, and bases for some different analytical view points are given. An introduction to the stochastic signal processing techniques used in this work follows, showing standard signal processing techniques common for EEG analysis to continuous stimuli.

Subsequently, Chapter 3 presents an outline of research questions and challenges within in the field of AAD. This includes discussions on usability and reliability of measures of attention, comparisons of results within AAD (and neurophysiological signal analysis in general), and the path towards realistic experimental design.

In Chapter 4, the aim and contributions of this thesis is summarized. We share the scientific contributions of each included work, as well as an outline of the author contributions within each paper.

Lastly, each work included in the thesis is appended, together with Appendix A, which is supplementary material to Paper IV.

Chapter 2

System Perspectives on Auditory Attention Decoding

1 Modelling connection between sound and brain responses

Consider recorded EEG with N_c channels, measured during an auditory attention experiment, as $y_c(t)$, during N_{tr} trials lasting T seconds each. During experiments, subjects are presented with speech from multiple concurrent speakers. Analogously to the conversion of sound pressure to electrical signals in the inner ear, transforms of the speech can be calculated. These are denoted speech features, which for instance can be the envelope of speech. The speech feature, denoted $s(t)$, can be for attended but also ignored speakers.

Measured EEG at channel c is denoted

$$y_c(t) = r_c(t) + e_c(t) \tag{2.1}$$

where $e_c(t)$ is noise, including measurement noise and background processing of the brain. The EEG response is assumed to be the output of some processing of the attended speech feature

$$r_c(t) = f_c(s(\tau), \tau \leq t) \tag{2.2}$$

This processing is assumed to be stationary in time, and only $s(t)$ and $y_c(t)$ are observable. Our system of signals can be summarized as

Speech feature	:	$s(t)$
EEG measured, channel c	:	$y_c(t)$
EEG response, channel c	:	$r_c(t)$
EEG noise, channel c	:	$e_c(t)$

2 Linearity

There is no reason to believe f_c is a linear operator. As described in Chapter 1, there is a large amount of different processes that connect sound heard to measurable brain processes. However, with this being said, there are multiple examples where linear modelling have been successful in describing the system in Eq. (2.1) for making meaningful inference.

Restricting the system model to be linear, one can write

$$r_c(t) = f_c(s(t)) = h_c(t) * s(t) = \int_{-\infty}^{\infty} h_c(\tau)s(t-\tau)d\tau \qquad (2.3)$$

where $h_c(t)$ is the impulse response of the system at channel c. The brain does not have an instantaneous response to stimuli, thus $h_c(t)$ has support over multiple time lags. And since the brain has limited memory capacity, $h_c(t)$ has finite support.

As a brief comment on the causality of this system, one may ask whether the brain processes events before they occur. Although predictive processing models that challenge strict causality are prevalent in the literature, we assume a causal framework in general for analysis. Accordingly, $h_c(t) = 0$ for $t < 0$.

This linear assumption has been used widely in the field, and has provided meaningful insights into auditory attention processing as the shape of h changes in different attentional states [2, 5]. Assessing $h_c(t)$ can be done in multiple ways, both parametrically and non-parametrically. Here follows a brief overview of some ways of looking at system in Eq. (2.1).

3 Discretization of the biological system

In practice, signals are measured in discrete time. Thus, the continuous time system in Eq. (2.1) is effectively discretized in all applications and analysis. By denoting discrete time with index n, and sampling period Δt, we have $t = n\Delta t$. The discrete time version of the linear system Eq. (2.3) is then

$$r_c[n] = h_c[n] * s[n] = \sum_{m=-\infty}^{\infty} h_c[m]s[n-m]. \qquad (2.4)$$

In practice, both $s[n]$ and $y_c[n]$ are measured for N samples, and $h_c[n]$ has finite support of M samples. Thus, the summation in Eq. (2.4) is limited to $m = 0, 1, \dots, M-1$.

EEG is sampled at some rate, in this thesis between 250 Hz and 1000 Hz. An important assumption to the set-up of EEG, is that brain responses are band-limited to below half the sampling rate, according to the Nyquist-Shannon sampling theorem. This is generally a valid assumption, as most cortical responses captured in AAD are below 32 Hz [2, 5], with some captured amplitude modulation in the high gamma band [4]. Even in non-linear modelling of information, there has been no observed benefit of including frequencies above 150 Hz. Thus, by sampling at these rates, it is plausible that no essential information is lost when capturing EEG. Assuming a linear system between speech features and EEG, any component in speech features above half the sampling rate of EEG will not be represented in the EEG measured. Therefore, speech features are commonly low-pass filtered and down-sampled to match the sampling rate of EEG before any analysis.

4 Non-parametrical quantification and coherence

A real valued stationary random processes, $x(t)$, and its Fourier transform $X(f)$ is related through

$$X(f) = \mathcal{F}\left[x(t)\right] = \int x(t)e^{-i2\pi ft}dt \tag{2.5}$$

Now assuming a linear system of the form in Eq. (2.3), $y(t) = h(t) * x(t) + e(t)$, the frequency response of the system is defined as

$$H(f) = \mathcal{F}\left[h(t)\right] = \int h(t)e^{-i2\pi ft}dt \tag{2.6}$$

Using the properties of the Fourier transform, the Fourier transform of the output, $Y(f)$ is related to the Fourier transform of the input $X(f)$ through

$$Y(f) = H(f)X(f) + E(f) \tag{2.7}$$

and the auto-spectrum of the output signal is

$$S_{yy}(f) = |H(f)|^2 S_{xx}(f) + S_{ee}(f) \tag{2.8}$$

where $S_{xx}(f)$ is the auto-spectrum of the input and $S_{ee}(f)$ is the auto-spectrum of the noise process, in turn defined as

$$S_{xx}(f) = \mathbb{E}\left[|X(f)|^2\right], \quad S_{ee}(f) = \mathbb{E}\left[|E(f)|^2\right]. \tag{2.9}$$

Furthermore, the cross-spectrum between input and output signals is defined as

$$S_{xy}(f) = H(f)S_{xx}(f). \tag{2.10}$$

The magnitude squared coherence, or simply coherence, between two signals $x(t)$ and $y(t)$ is defined as

$$C(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)} \tag{2.11}$$

where a cross-spectrum $S_{xy}(f)$ is normalized by the two respective auto-spectra of the signals. The common interpretation of the coherence measure is that it measures the relative linearity between the two processes. If $y(t)$ can be explained by a linear filtering of $x(t)$, then $C(f) = 1$ at the frequencies where x and y have power. If there is no linear connection between the processes at a frequency, coherence is zero instead. Another way of putting this is that coherence can detect spectral components in channels $x(t)$ and $y(t)$ that are phase-locked.

Thus, as EEG response to stimulus can be described by a linear filter with usefulness, one can estimate this filter and its relative power non-parametrically. One advantage of this, is that there is no imposed statistical prior on the filter. This in turn can avoid systematic bias in imposing filter models, if proper model priors are hard to derive.

5 Parametric estimation

Commonly, one wants to parametrically estimate the filter $h_c(t)$ of Eq. (2.3). The advantages of this includes using priors and regularization to improve robustness to noise. Parametric estimation can be done in multiple ways. One way is to estimate each time lag of $h_c(t)$ directly, for example through least squares estimation or maximum likelihood estimation. This is commonly done in AAD research, where the filter is called a temporal response function (TRF) [5]. This is further detailed in Section 8.1, where other options are also explained.

If the goal of finding $h_c \in \mathbb{R}^M$, that predicts the response in channel c, r_c from s, one can set up the following optimization problem,

$$\hat{h}_c = \arg\min_{h_c} \sum_{n=1}^{N} \left(y_c[n] - \sum_{m=0}^{M-1} h_c[m]s[n-m] \right)^2 + \lambda \|h_c\|^2 \tag{2.12}$$

where λ is a regularization parameter. This is a ridge regression problem, which has a closed form solution. The regularization term is important to avoid overfitting to noise in the data, especially when M is comparatively large compared to N. The filter found by Eq. (2.12) seeks to maximize the correlation between estimated response and measured EEG, which is called *forward modelling*, as one tries to model the forward path of the system in Eq. (2.1).

Alternatively, one can try to find a filter that estimates the speech features s from the EEG $y_1, \dots, y_C$. This is called *backward modelling*, as one tries to model the inverse path of the system in Eq. (2.1). The optimization problem is then

$$\hat{g} = \arg\min_{g_1,\dots,g_C} \sum_{n=1}^{N} \left(s[n] - \sum_{c=1}^{N_c} \sum_{m=0}^{M-1} g_c[m]y_c[n-m] \right)^2 + \lambda \|g\|^2 \tag{2.13}$$

where $g_c[m]$ is the filter element for channel c at lag m. The solution to Eq. (2.13) also has a closed form solution. Backward modelling has been shown to perform better in AAD tasks compared to forward modelling [27], as it leverages information from all EEG channels simultaneously. A drawback of backward modelling is that the filter g is less interpretable compared to h_c in forward modelling, and it requires more computational resources to estimate, as all channels are estimated simultaneously.

It is also possible to combine forward and backward modelling into a hybrid model. For example, this can be done through canonical correlation analysis applied on lagged data, where both speech features and EEG from multiple time lags are projected into subspaces where their correlation is maximized [14].

6 Non-linear modelling of the system

Although linear models have been shown to be useful in describing the system in Eq. (2.1), non-linear models have the potential capture more complex behaviour. For the past 10 years, this has been an active area of research in AAD. One path of modelling the non-linear system is to use co-called neural networks, to approximate f_c. The universal approximation theorem states that a neural network exists that can approximate any continuous function to arbitrary accuracy. In it simplest form, this holds for a feed-forward neural network [40]. This is caveated by assuming models have wide or deep enough hidden layers, and that the activation functions within used is non-polynomial [41]. Similar results exist for other deep learning architectures, such as convolutional neural networks [42] and recurrent neural networks [43].

Thus, in theory, neural networks can approximate the system in Eq. (2.1) to arbitrary accuracy. Note that this only states the existence of such a neural network. The largest issue comes in finding proper weights. In practice, data amounts, overfitting, and computational costs limit the performance of neural networks. Also, f_c is not necessarily continuous as neural spiking are discrete events. Furthermore, f_c is not time-invariant, as the brain changes its processing strategies depending on context, effort, fatigue, and other cognitive processes. Thus, neural networks may not be able to approximate f_c perfectly in practice, as data labelling is insufficient. As a final point, stochasticity in brain processing and brain measuring limits the performance of any model, linear or non-linear.

After these caveats and limitations have been stated, it still remains that neural networks have been used both in forward and backward modelling of AAD systems successfully, often beating performance compared to linear models [18, 26, 44]. In [26], authors showed that including non-linearities in stimulus reconstruction through neural networks improved reconstruction metrics compared to linear models. Although this can be partly explained by the neural network learning the structure of speech features better (for example common lengths of words and phonemes), it still shows that non-linear models can capture more complex relations between speech and brain activity. Using these non-linear model to classify attention between candidate speakers, by looking at which speaker is better reconstructed, has also been shown to improve performance compared to linear models.

7 Analysis and classification of the output of the system

Again, the linear model of h is a considerable simplification, but has proven useful. However, the model has limits to what behaviour it can capture. As it also may be hard to define the non-linear f_c, one can analyse the output alone to circumvent these problems. In this case, one is essentially performing pattern recognition on the output $y(t)$ of the system. Typically, data is labelled depending on current attentional state. For example, this can be attention towards left or right. This view-point can simplify the process of decoding attention to these specific directions. However, the output-only viewpoint introduces other problems instead. Classification models may overfit to other patterns in the data, and care is therefore needed when defining metrics and labels on the output alone.

Additionally, if one rejects to use the speech signals at all, how does one gauge where in time EEG data is interesting and carries information? Commonly the application of labels is done on a macro level, in a top-down manner, where labels are set as the instructed attentive state (such as ”attend left speaker!”). In reality, things are not as stable or binary. But as continuous true labels are hard to impossible to capture in real data, one is left with the binary class labels. Changes to experimental design are needed to get closer to continuous labels, but this is out of scope for the thesis.

To do classification, multiple approaches are again possible. Any type of classifier can be used on the output data. Commonly, data is first commonly pre-processed to extract relevant features. Although there is no exhaustive list of methods used, examples include neural networks [45], linear discriminant analysis [46], and common spatial patterns [47].

Key challenges in classification is to avoid overfitting to noise and other patterns in the data. These issues are mitigated through regularization, cross-validation, as well as careful design of training and test sets. Additionally, interpretability of models is important to understand what patterns in the data that are used for classification. With this being said, overfitting remains a key challenge in AAD research, as auditory attention can be heavily connected with other cognitive processes in the brain. An example of this is that a listener will intuitively look towards the attended speaker. Thus, eye-movements and auditory attention are confounded in many experimental set-ups.

Employed in the work is primarily a common spatial pattern (CSP)-based classification paradigm, with linear classification of output features. This is described in more detail in Section 8.

8 Signal processing methods in AAD

This section presents methodological considerations, when performing different types of inference on the system of Eq. (2.1), Eq. (2.3) and Eq. (2.4).

8.1 Temporal filter estimation

As explained previously, parametrical estimations of the forward or backward filter between speech features and EEG responses, entails solving the optimization problem of Eq. (2.12) or Eq. (2.13). In practice, this problem can be underdetermined, or at least very susceptible to noisy measurements. Commonly, regularization is used to mitigate these issues.

There are many options for applying regularization to the optimization problem at hand. Each regularization infers some prior knowledge about the filter sought. Utilizing L2 regularization (ridge regression) infers that filters should be smooth and distributed between times (and channels in the case of backwards models). L1 regularization (lasso regression) infers that filters should be sparse, and thus only a few time lags (and channels) should be used in the filter. Tikhonov regularization (generalized ridge regression), where the regularization term can be defined as $\lambda\|Lh\|^2$, where L is a matrix that defines the structure of the regularization. For example, if L is the identity matrix, it reduces to ridge regression. If L is a matrix that computes the second derivative of the filter, it infers that filters should be smooth in terms of their second derivative. The choice of regularization depends on the desired properties of the filter and the amount of data available for estimation. In the case of AAD, L can be defined to infer spatial covariance between channels, or temporal smoothness. A combination of L2 and L1 regularization, often called elastic net regression, can also be used to combine the properties of both regularization methods. Grouped lasso, where the regularization term is defined as $\lambda \sum_g \|h_g\|_2$, where h_g is a group of parameters, can also be used to infer that certain groups of parameters should be selected together. For example, in the case of backwards models, one could define groups of parameters corresponding to each channel, and use grouped lasso to select which channels are important for the filter. Each of these regularization methods has its own advantages and disadvantages, and the choice of regularization should be based on the specific problem at hand and the desired properties of the filter as it infers priors assumptions about the connection between speech and brain responses.

Another way to explicitly assume a certain prior, is to introduce a parametric form of $h_c(t)$, for example as a sum of basis functions, and estimate the parameters of these basis functions. Both forward and backward modelling can utilize basis functions instead of estimating each time lag directly. For example, $h_c[n]$ can be modelled as a sum of P basis functions $\phi_p[n]$ in discrete time, such that

$$h_c[n] = \sum_{p=1}^{P} a_{c,p}\phi_p[n] \tag{2.14}$$

where $a_{c,p}$ are the weights of each basis function for channel c. This reduces the number of parameters to estimate, and can improve robustness to noise. Depending of the basis function used, the implicit prior implied on $h_c(t)$ changes. For example, using Gaussian basis functions imposes smoothness on $h_(t)$.

8.2 Spectral density estimation

A continuous signal $x(t)$ can be sampled to $x[n]$ at some sample rate f_s, resulting in N samples. In this case the Fourier transform, formulated in Eq. (2.5), instead becomes

$$X[f] = \sum_{n=0}^{N-1} x[n]e^{-i2\pi fn} \tag{2.15}$$

where $0 \leq f < f_s/2$. The Fourier transform can be computed efficiently through the fast Fourier transform (FFT) algorithm [48], making spectral analysis a viable option in real-time applications such as HA technology.

To estimate the spectral density of a signal, multiple methods exists. The most simple is the periodogram method, where the Fourier transform of a signal is calculated, and the auto-spectrum is estimated as

$$\hat{S}_{xx}[f] = \frac{1}{N}|X[f]|^2 \tag{2.16}$$

However, this method is known to be a biased and inconsistent estimator of the true spectral density [49]. Thus, other methods are commonly used instead. The modified periodogram is one such method, where the signal is windowed before calculating the Fourier transform,

$$X_w[f] = \sum_{n=0}^{N-1} w[n]x[n]e^{-i2\pi fn} \tag{2.17}$$

where $w[n]$ is a window function, such as a Hamming or Hann window. The modified periodogram is then estimated as Eq. (2.16). This reduces spectral leakage, at the cost of some narrow-band bias in the estimate.

Reducing noise and bias of spectral density estimates

The periodogram and modified periodogram methods have high variance, making them noisy estimates of the spectral density. To reduce this variance, multiple averaging methods exists. The perhaps most common methods are the Welch method [50] and the Thomson method [51]. Both these methods seek to reduce the variance of the spectral density estimate, at the cost of some increased bias. We shall first describe the two methods and then discuss their properties in variance reduction and bias.

The Welch method divides the signal into L overlapping segments of length M, where each segment is windowed with a window function $w[n]$. The Fourier transform of each segment is calculated, and the modified periodogram is estimated for each segment as

$$\hat{S}_{xx}^{(l)}[f] = \frac{1}{U}|X_w^{(l)}[f]|^2 \tag{2.18}$$

where U is a normalization factor depending on the window function. The windowing functions used in the Welch affect the bias and variance of the spectral density estimate.

Common window functions include the Hamming, Hann, and sine windows. The choice of window function depends on the desired trade-off between bias and variance. The final spectral density estimate is then obtained by averaging the modified periodograms of each segment,

$$\hat{S}_{xx}[f] = \frac{1}{L} \sum_{l=1}^{L} \hat{S}_{xx}^{(l)}[f] \tag{2.19}$$

The Thomson method, also known as the multitaper method, uses multiple orthogonal window functions of length N, called tapers, to reduce the variance of the spectral density estimate. The tapers are designed to have optimal spectral concentration within a certain bandwidth. The Fourier transform of the signal is calculated for each taper, and the spectral density estimate is obtained by averaging the modified periodograms of each taper,

$$\hat{S}_{xx}[f] = \frac{1}{K} \sum_{k=1}^{K} \hat{S}_{xx}^{(k)}[f] \tag{2.20}$$

where $\hat{S}_{xx}^{(k)}[f]$ is the modified periodogram for taper k. The tapers are found by solving an eigenvalue problem that maximizes the spectral concentration within a certain bandwidth,

$$\sum_{m=0}^{N-1} \frac{\sin(2\pi W(n-m))}{\pi(n-m)} v_k[m] = \lambda_k v_k[n] \tag{2.21}$$

where W is the half-bandwidth parameter, and $v_k[n]$ are the tapers. The eigenvalues λ_k indicate the spectral concentration of each taper, and only the tapers with the highest eigenvalues are used in the spectral density estimate.

The variance reduction of the Welch and Thomson methods comes from averaging multiple modified periodograms, which reduces the variance of the estimate. So, for white noise the variance is reduced by a factor of L or K, in the Welch method or Thomson method correspondingly.

The bias of each estimate is increased due to the windowing and tapering of the signal, which introduces spectral leakage. The bias increases as the effective bandwidth increases. The effective bandwidth for the Welch method with a Hann window is $4f_s/M$, where M is the length of each segment [50]. The effective bandwidth for the Thomson method is $2NWf_s/N$, where T is the signal duration and W is the half-bandwidth parameter [51]. Increasing the segment length in the Welch method reduces the bandwidth and therefore reduces bias, while decreasing the half-bandwidth parameter in the Thomson method reduces the bandwidth and bias of the estimate.

The choice of window function and tapers affects the trade-off between bias and variance, and should be chosen based on the desired properties of the spectral density estimate, and the characteristics of the signals being analyzed.

Metrics for comparing signals

Many parts of this thesis will use correlation between different signals as a metric for comparing similarity. Correlation can be computed in both time and frequency domain. In time domain, the Pearson correlation coefficient is commonly used, defined as

$$\hat{r}_{xy} = \frac{\sum_{n=1}^{N}(x[n]-\bar{x})(y[n]-\bar{y})}{\sqrt{\sum_{n=1}^{N}(x[n]-\bar{x})^2}\sqrt{\sum_{n=1}^{N}(y[n]-\bar{y})^2}} \tag{2.22}$$

where $\bar{x}$ and $\bar{y}$ are the means of signals $x[n]$ and $y[n]$, respectively.

8.3 Spatial techniques

In AAD, multiple EEG channels are measured. Channels measure the electrical activity at different locations on the scalp, but also contain the same information. In this thesis the methods that utilize spatial information and patterns between the electrodes, are denoted spatial techniques. These techniques can be used for various purposes, such as improving signal-to-noise ratio, reducing dimensionality, and extracting relevant features, or performing classification. Two common spatial techniques used in AAD research are the CSP algorithm and independent component analysis (ICA).

Common spatial pattern

The CSP algorithm is a widely used technique in brain-computer interface (BCI) research for spatial filtering of multichannel EEG data. The goal of CSP is to find spatial filters that maximize the variance of the filtered signals for one class while minimizing it for another class. This is particularly useful in binary classification tasks. The CSP algorithm works by solving a generalized eigenvalue problem. Given two classes of EEG data, $X_1, X_2 \in \mathbb{R}^{C\times N}$, and the covariance matrices for each class, $R_1, R_2 \in \mathbb{R}^{C\times C}$, the CSP are found by solving the optimization problem

$$\max_{w} \frac{||w^T X_1||^2}{||w^T X_2||^2} \tag{2.23}$$

or equivalently,

$$\max_{w} \frac{w^T R_1 w}{w^T R_2 w} \tag{2.24}$$

where R_1 and R_2 are the covariance matrices for class 1 and class 2, respectively. Similar to how the eigenvalue problem $R_1 w = \lambda w$ finds directions that maximize variance in data, Eq. (2.24) is solved by

$$R_1 w = \lambda R_2 w \tag{2.25}$$

where w is the spatial filter, and λ is the corresponding eigenvalue. The eigenvectors corresponding to the largest eigenvalues maximize the variance for class 1 while minimizing it for class 2, and vice versa for the smallest eigenvalues. This problem can be reformulated by exchanging R_2 to $R_1 + R_2$ in order to have better conditioned matrices in optimization. On this form the eigenvalue problem becomes

$$R_1 w = \lambda(R_1 + R_2)w. \tag{2.26}$$

This formulation will produce the same eigenvectors, but the eigenvalues will be scaled. Again, the eigenvectors corresponding to the largest eigenvalues maximize the variance for class 1 while minimizing it for class 2, and vice versa for the smallest eigenvalues. Thus, in classification pipelines, the eigenvectors corresponding to the largest M and smallest M eigenvalues are commonly used as spatial filters, where M is a hyperparameter.

Note that Eq. (2.24) assumes that the covariance matrices are full rank. If they are not, some eigenvalues can be zero or infinity, leading to numerical instability. To mitigate this, regularization techniques can be applied, such as adding a small constant to the diagonal of the covariance matrices. Also, the dimension of data can be reduced through methods such as principal component analysis (PCA) before applying CSP to ensure that the covariance matrices are full rank.

Independent component analysis

ICA is a computational method for separating a multivariate signal into additive, independent components. It is commonly used in EEG signal processing to separate brain activity from artifacts such as eye blinks and muscle movements. The basic assumption of ICA is that the observed signals are linear mixtures of statistically independent source signals. The goal of ICA is to find a demixing matrix A such that

$$S = AX \tag{2.27}$$

where $S \in \mathbb{R}^{n_{\text{ica}},N}$ is the matrix of n_{ica} independent source signals, and $X \in \mathbb{R}^{C,N}$ is the matrix of observed signals. The demixing matrix A is estimated by maximizing the statistical independence of the source signals. This is commonly done by maximizing the non-Gaussianity of the source signals, as independent signals are assumed to be non-Gaussian (since mixing increase Gaussianity according to central limit theorem). Alternatively, the entropy can be maximized. Various algorithms exist for estimating the demixing matrix, and is a whole field of itself, but in this work two very commonly used algorithms are applied: the FastICA algorithm [52] and the Infomax algorithm [53].

Application to EEG cleaning ICA is commonly used in EEG signal processing to separate brain activity from artifacts such as eye blinks and muscle movements. The independent components are estimated from the observed EEG signals, and components corresponding to artifacts are identified and removed. The cleaned EEG signals are then reconstructed by multiplying the remaining components with the mixing matrix. The idea is that by removing components corresponding to artifacts, the relevant EEG components will have a higher signal-to-noise ratio (SNR) and be more suitable for further analysis.

Solving Eq. (2.27) only finds independent spatial components that make up the observed EEG data. No information about identifing the sources of the component is included at this stage. Thus, the components found by ICA are not automatically labeled as brain activity or artifacts, but left to manual interpretation. Packages for artifact identification and removal, such as MNE and EEGLAB for Python and Matlab respectively, have built-in ICA artifact analysis applications. In these applications the components found by the ICA algorithm can be visualized and manually inspected. Traditionally, components corresponding to artifacts can be identified by their spatial topography, time course, and frequency content. Once identified, these components can be removed from the data, and the cleaned EEG signals can be reconstructed. In Fig. 2.1, examples of identified artifact components are shown. The step of identifying which component to remove requires technical expertise, as misidentifying components can lead to loss of important brain activity information.

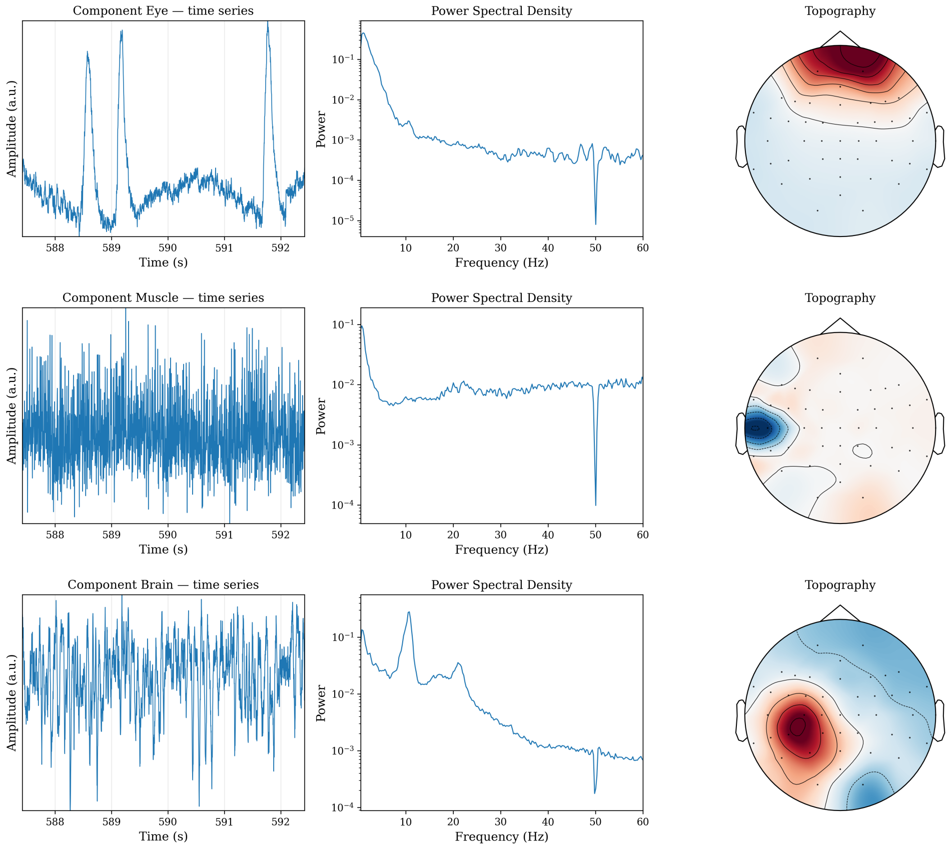

Figure 2.1: Some examples of independent components found in EEG data using ICA. Each component is represented by its time course (left column) and power spectral density (middle column) and topographic pattern (right column). Components corresponding to artifacts such as eye blinks and muscle movements can be identified by their characteristic patterns.

Chapter 3

Practical Challenges and Advancements of Auditory Attention Decoding

With these perspectives on the system in mind, a plethora of challenges and research questions exist in order to understand auditory attention in naturalistic scenarios, and to make AAD usable in real-world applications. Here follows an outline of some of these challenges, which we have focused on in this thesis, as well as some relevant advancements.

1 Usable and reliable measures of attention

One aspect of designing methods for assessing auditory attention, is to have reliable and objective measures of behaviour in relation to the technology. If these measures are to be used to make decisions about human behaviour, they need to actually measure what clinicians and designers think they do. This is important both for research purposes, but also for clinical applications. Two examples follow, to illustrate this need.

The first example, assume one wants to assess how well a listener is able to attend a certain speaker in a noisy environment. This could be useful for tuning the settings of a HA. One can use an AAD measure to give objective feedback for this. However, if the measure is not reliable, it may give different results depending on unknown factors. For example, if the measure has overfitted to ocular activity during training, it may overestimate how well the listener is performing, as the listener will probably look the correct way.

As a second example, assume one wants to analyze the shape of the filters found in AAD experiments, to understand how the brain processes attended and ignored speech. If the researcher goes in with the assumption that filters should be smooth and literally look like the known auditory evoked potentials, they may impose priors on the filter estimation that bias the results. For example, if the researcher uses a basis function set that only allows smooth filters, any non-smooth behaviour in the data will be ignored. Thus, the researcher may miss important aspects of auditory attention processing. As SNR is high, one may have to impose strong priors to get any results at all, but care is needed to not bias results too much.

Development of reliable measures is thus important in both these aspects. Designing models that do not overfit to noise and other confounding factors is vital. Additionally, proper validation of methods is needed to ensure that they measure what they are intended to measure. Only then, when methods are trustworthy, can they be used in evaluating hearing conditions for individuals.

Examples of work advancing the field of objective evaluation of hearing conditions have had various focus. In [54], authors show that speech-in-noise tests show effects from different HA functions. In a series of works, [55–58] focus has been on how pupilometry can be used as an objective measure of listening effort and other metrics in different hearing conditions. In [59] authors show increased speech neural representation with decreased sound quality and in a series of papers, [12], [13], authors show effects of HA noise reduction on the neural representations of speech.

Further advancements in objective evaluation would be valuble and a progression in signal processing techniques to improve reliability would make that easier. Making analysis methods that are based on solid statistical, model-based foundations is essential.

2 How can one compare data and results between datasets and labs?

Research is a collective effort. A single person or lab cannot and should not solve all problems on their own. Instead, sharing data, methods, and results is fundamental to advance the field. Thus, being able to compare results between different labs and datasets is important. However, this is not trivial. Different experimental set-ups, different measurement devices, as well as different analysis methods can lead to results that are not comparable. In other fields, such as computer vision and automatic speech recognition, the research community relies heavily on common datasets and benchmarks for comparison of methodologies.

There is strife and ambition within the field to increase the amount of consensus and harmonization in regards to data. Recent work that have helped in this aspect include the publication of public datasets for AAD research, [60–62]. Additionally, challenges such as the *Auditory EEG decoding challenge* at ICASSP 2023 and 2024 have been useful in comparing methods between labs [63, 64]. The challenges have also released a comparatively large corpus of data for EEG decoding of audio. A broad application of novel methods between datasets is vital, to see how well they generalize. Also, more insights into each dataset can be gained by applying different methods to the same dataset.

With this being said, one can run into data sharing problems. EEG data is considered sensitive personal data in many countries, and sharing it can be legally complicated. Thus it is unavoidable that the field needs to compare results and methods between datasets, without sharing data. Different labs have different experimental set-ups, making direct comparison of results hard. For example, different EEG caps have different electrode placements, making it hard to apply spatial filters between datasets. Different stimulus material and experimental paradigms can make comparison of results hard. Also, the researchers' choices, such as in preprocessing of data can vary between labs. Thus, it is important to understand how each of these factors affect results, to be able to compare results from different labs and datasets.

3 Moving towards varied and naturalistic scenarios

Eventually, AAD methods should be usable to understand auditory attention in the real world, no matter if it is for clinical or research purposes. However, most research in AAD is done in controlled laboratory settings, with limited variation in set-up compared to what is possible in every-day life. Moving towards more naturalistic scenarios is thus important to make AAD usable in real-world applications. Here follows some areas to where this thesis aims bringing the lab closer to the world out side. Please remember the caveat that moving towards naturalistic scenarios introduces new degrees of freedom in experimental set-up which can be problematic in drawing scientific conclusions.

3.1 Varied talkers and conversations

Natural progression of auditory experiments have gone from short clicks, to words, to sentences, to continuous speech; not necessarily in such a straight path as the list might suggest. Additionally, moving from single speaker scenarios to multiple speaker scenarios have been done. However, most experiments still use a limited set of speakers, commonly only two monologuing speakers at $\pm 30^\circ$ or in earphones. This is far from real-world scenarios, where multiple speakers are present, and many of them are engaged in the same conversation. Thus moving towards including conversation data has been recently done in [65], where two and three sets of conversations, each presented on one screen, were used in classic auditory attention decoding. One of the conversations was to be attended at a time, and some trials included attention switching within trials. Other conversational behavioral studies have been performed, but nothing else where neural responses are captured, to my knowledge. Multiple challenges arise during analysis of conversations. The characteristic of the speech from each speaker changes compared to the characteristics of a single monologuing speaker. Also, attention switches between the participants of the conversations are expected and happen often. Multiple ways of looking at the conversations are available, where either each conversation is

Figure 3.1: Examples of stimuli presented to participants of AAD studies. To the left four single AV speakers are shown. To the right an example of conversation AV stimuli is shown. (Images taken videos from *Spørg Direkte* and *HEARTBEATS DK* Youtube channels, both under CC-BY)

considered a target or each participant of all conversations. Examples of single speaker AV stimuli and conversation AV stimuli is shown in Fig. 3.1. One can see how sticking to the left single speaker corpus may not capture the full range of variability in speech and attention that is perhaps more common in real-world scenarios.

Moving towards more speakers and more varied talkers is also important to understand how auditory attention works in different real-world scenarios. Speech entails diverse sets of languages, speaking styles, accents, dialects and voice characteristics. It is important that models generalize between these conditions, and do not overfit to certain patterns that are the most common in experimental conditions.

3.2 Varied acoustic conditions

Listening changes depending on the conditions where it is done. As an example, if it is easy to hear someone, little to no effort is needed for the task. Comparing this to listening in highly noisy environments, which be can be very taxing, one immediately realizes that neural signatures can differ. This is also backed up by many studies on listening effort.

Arugably this is the area where most progress have already been made. Within the field, scientists have varied the stimuli presentation methods from different type of headphones to free field presentation through speakers. Signals have been presented alone, with one to multiple competing speakers. It has been done in quiet rooms and in rooms with background noise. The type of this background noise has also been varied. Experiments have been performed in anaechoic chambers, normal rooms, and outside [66, 67]. All this is important as it makes it possible to gauge the large and small changes to AAD behaviour and performance. Recently, work has been done to create computational models that directly relate peripheral auditory processing, by emulating human behaviour in hearing tasks, such as [68, 69]. Further evolutions of this could be used to pick out interesting acoustic settings to run specific experiments on, as well as to be used in AAD method development.

3.3 Varied listening tasks

Designing effective EEG experiments is a challenge, where leveraging different aspects is challenging. One needs to make sure the neurological mechanisms of interest are measurable, while at the same time not making the laboratory setting so unnatural that results are irrelevant. Generally, these are opposing each other, where the extremes at either end make experiments inadequate.

In most AAD experimental set-ups, the listening task has been to sustain attention to one out of multiple speakers for the entirety of experimental trials. In reality, listening could be to listen to one talker and switch, or to try to have considerable attention to both to pick up important information. There are multiple works that have advanced the field in these directions. In [70] method development for dynamic tracking of attention was presented. In [71], authors verified the commonly AAD methods on experiments spatially moving speakers. The effects of attention switches on pupilometry and alpha power of EEG was investigated in [72]. Furthermore, in [65], author investigated auditory attention and attention switches between sets of conversations. In [73], authors showed asymmetric disengagement and engagement processes for TRF models during attention swtiches, indicating that the newly attended speaker is encoded before the previously attended speaker is disengaged fully.

In real-life one often look at the speaker whom one is listening to. By processing congruent visual input enhances auditory, perception is improved. As evidence for this, studies have shown that this is objectively the case. Recent advancements include [74], where neural representations of audio is increasing, even when measuring EEG in experiments with unscripted, natural speech audiovisual (AV)-stimuli from untrained speakers within a virtual acoustic environment. The same study even specifically show the importance of lip movements in adverse listening conditions, by including experimental conditions where lip movement is masker to the participants. This yields similar neural representations to the audio-only condition. This clarifies the importance of including multiple sensory modalities in AAD research, which in the past have seldom taken into account during auditory research.

3.4 Including hearing impaired

For individuals with HI, the cocktail party problem is harder. Due to HI, the auditory pathway deteriorates in its ability of processing sounds in noisy environments [75]. Exactly what mechanisms triggers this decline is not fully understood. Both the ability of individuals with HI to discern sounds, as well as the binaural fusing of sound information can play part. Hearing loss is also in and of itself an early indicator for both depression and dementia [76, 77], another reason to alleviate the perceptual barriers for people with HI.

HAs help individuals in selective attention in noisy environments. However, the mechanisms in play can be affected. Since HAs themselves augment the listening experience for the user, this adds a degree of freedom in estimation. The same algorithms that can decode attention

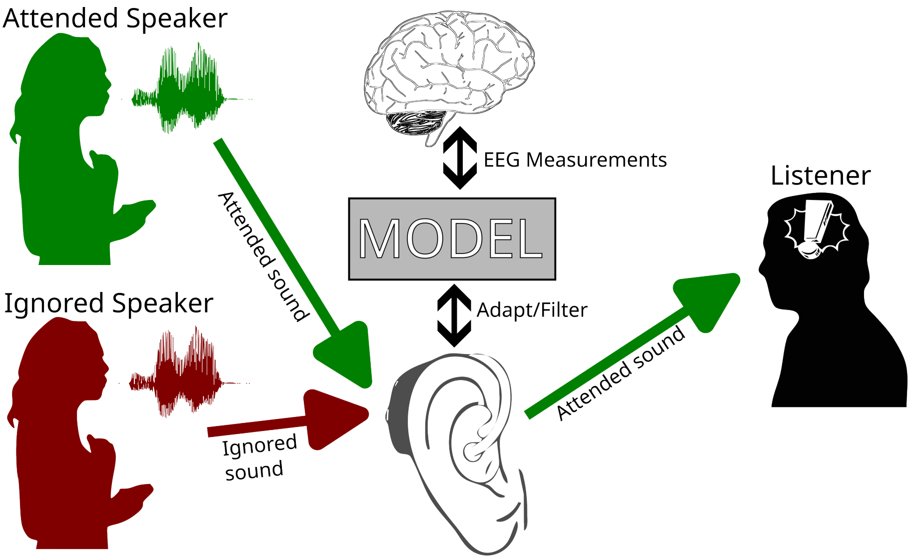

Figure 3.2: A conceptual future HA that concurrently measures brain activity as a wearer (in black) listens to an attended speaker (in green) when a disturbing speaker (in red) is present. By analyzing brain signals, the HA adapts and only lets through sound from the speaker the listener is paying attention to.

of NH individuals, may fail when applied to HA users.

Therefore it is of utmost importance to include this group into the development of methods and scientific progression, so the path from research to translation is short. It is also HA users, or at least hearing device users, that will make the most use of the results from this research field.

Ideally, future HAs fused with some brain monitoring system can decode the attention of a listener and attenuate disturbing sounds. A schematic illustration of a simplified potential system is shown in Figure 3.2. For example, in any real application one has to be able to let enough of other speakers through so users can change attention naturally in a conversion. Also, the HA needs to let through salient or potentially threatening sounds.

As a side note, measuring how HAs help users in cocktail party scenarios is important, as this can be used clinically. HAs have to be fitted which means setting appropriate enhancement levels of sounds at different frequencies. Additionally, modern HAs are employed with noise reduction schemes and different settings for different audio environments. Objective evaluation of performance of settings and algorithms in HAs are constantly sought after to help users in systematic ways [78], which also relates back to the challenges described in Sec. 1.

4 Trusting labels

An important factor to keep in mind when interpreting results in this thesis, and an inherent problem in AAD research, is the problem of trusting labels too much. In most AAD research, labels are used to define which parts of the data are interesting and what parts are not. For example, in a classic AAD experiment, one may have two speakers, and the label is which speaker is attended at a certain time window. However, these labels are not always reliable. For example, if a participant is supposed to attend to one speaker, but their attention drifts to the other speaker for a moment, the label will be wrong for that time window. Additionally, if a participant is supposed to attend to one speaker, but they are not able to do so due to HI or other factors, the label will also be wrong. Furthermore, there is again a conflict of interest moving towards realistic scenarios, which is opposed to having reliable and specific labels and knowing what subjects are actually doing in the lab. Moving towards more naturalistic scenarios increases the degrees of freedom in what participants can do, and thus makes it harder to define reliable labels. The extreme would be to have an experiment where participants are free to switch attention whenever they want, and there is no way to know what they are attending to at any given time.

One recent work, that have touched upon this problem is [79], where the authors show that in a classic AAD experiment, the performance of a model can be explained by confounding factors in experiments such as eye-movements, rather than actual auditory attention. This is a clear example of how trusting labels too much can lead to wrong conclusions about what is actually being measured.

In order to mitigate this problem, researchers often use behavioural measures to validate the labels. For example, they may ask participants to answer questions about the content of the attended speaker after the experiment, to ensure that they were paying attention and managed to perform the listening task. However, it is not possible to ask questions about all time points in a given trial, and thus questions must be asked about things said by the speaker at some points in time within the trial. This means that there might not be any behavioural information wether the participant was able to understand the attended speaker on trial segments far away from these time points. Furthermore, asking questions is not always possible, especially if one wants to move towards naturalistic scenarios where attention can switch frequently whenever the participant wants to.

One might naively ask, why not just ask participants more questions to cover the entire trial? However, even if it would be possible, this is not practical. It would make the experiment too long and tedious for participants. It would decrease the amount of time one can actually capture neural data, and thus decrease the amount of data available for analysis. Therefore, different approaches have been proposed to get more reliable labels. One example is to give participants a button and tell them to press it whenever they hear a certain word or phrase in the attended speaker. The trial ends when the button is pressed, and restarted with a new word to listen for. This can give more information about whether the participant is paying attention during the continuity of the trial, as a trial probably should be discarded if the participant fail to press the button at all. However, this approach is not perfect either, as data may be thrown away in cases where the participant is paying attention but just fails to press the button, or if they are not paying attention but just happen to press the button at the right time. Additionally, this approach can be problematic in naturalistic scenarios, where one cannot define specific words or phrases to listen for. It may also be challenging for participants to design experiments around this type of task, in regards to factors such as randomization of conditions. A technique that could be used together with this is extending quality estimation metrics, which is standard practice in event-related potential (ERP) analysis [80], to TRF analysis.

Another way is to use eye-tracking data to infer where the participant is looking, and thus which speaker they are visually attending to. However, this means that eye artifacts will be present in data, and additionally sometimes one may look at one speaker but have auditory attention to another speaker, which becomes problematic. Additionally, eye-tracking data can be noisy and loss of data can make entire experiment session unusable.

To mitigate the problem of trusting labels too much, further development on the experimental design side is needed, but also on the analysis side. Further developments include unsupervised methods that do not rely on labels, such as [81, 82]. Developing methods that can detect when labels are unreliable, and either discard those data points or give them less weight in analysis, could be useful. Additionally, developing methods that can leverage other sources of information, such as eye-tracking data, to improve the reliability of labels could also be useful. In the end, it will always be impossible to know the true, instantaneous intention of humans.

Chapter 4

Summary and Contributions

Co-authors are abbreviated as follows:
Oskar Keding (OK), Johanna Wilroth (JW), Emina Alickovic (EA), Martin A. Skoglund (MAS), Maria Sandsten (MS), Martin Enqvist (ME).

I **Effect of Independent Component Artifact Rejection on EEG-Based Auditory Attention Decoding**
Oskar Keding, Johanna Wilroth, Martin A. Skoglund, Emina Alickovic
32nd European Signal Processing Conference (EUSIPCO), Lyon, France, 2024

Scientific contributions: While the rest of this thesis focuses on analysis of EEG data in a AAD context, this paper instead anlyzes the rather important preprocessing steps, that are common practice in all EEG analysis. Understanding the influence of EEG preprocessing is critical to make conclusions in analysis. One step in this preprocessing step is performing artifact cleaning of EEG data, where parts of the signals that are known to not be of interest are removed, which is commonly done with ICA. In the specific field of AAD, there is a research gap in how reliable results are in relation to the methodology used when rejecting EEG artifacts with ICA. There is a level of subjectivity in assesment of each component, since the spectral, temporal and spatial characteristics are visually analyzed. Also, this task is done recording-wise, so an increased risk of mis-identification when fatigue of the researcher increases. In this work, multiple automatic labellers of artifacts are analyzed and compared to traditional manual inspection, for specifically an AAD experiment. If automatic artifact rejection schemes are satisfactory, this would save a lot of time. We show that there are only small deviations in the performance of TRF-based decoders on the investigated dataset, but we are able to extrapolate some patterns. For example, omitting artifact rejection can actually increase performance in backward modelling, but manual artifact rejection was superior compared to automatic artifact rejection schemes for finding encoding of attention in frontal

EEG channels.

Authors' contributions: OK had the first idea for the paper, but the idea was then developed by all authors (JW, EA, MAS and OK). JW and OK wrote the code for the investigation and produced results. Result analysis was done by all authors (JW, EA, MAS and OK). The majority of the paper was written by OK, but with substantial parts written by JW, and review from EA and MAS.

II **Novel bias-reduced coherence measure for EEG-based speech tracking in listeners with hearing impairment**
Oskar Keding, Emina Alickovic, Martin A. Skoglund, Maria Sandsten
Front. Neurosci., Volume 18 - 2024

Scientific contributions: Due to success of linear input-to-output methods in AAD, the measure of spectral coherence should be a suitable method to detect cognitive connections between speech features and EEG responses. Compared to parametric methodology, the advantage of this approach is that the risk of overfitting and misspecified statistical priors are significantly lower, both important aspects in gathering interpretable quantitive results. Spectral coherence measures have previously been shown to differentiate attended and ignored speech on group levels for NH individuals, in [83]. Firstly, this paper shows that this observation extends to a HI population, where coherence is applied on a free-field HI experimental dataset. In parallel, methodology developments are made, specifically in relation to AAD research. High amounts of measurement noise in EEG and non-linearities cause very low values of coherence in these applications, even though there is a linear connection. We present a methodological improvement in coherence estimation compared to previous applied coherence methods in the field, which increases performance. Thirdly, a EEG-specific characterisation of the coherence estimation method, Thomson's multi-window method, is made. We present an analytical bias expression to show an inherent bias of coherence peak towards higher frequencies, due to $1/f-$ shaped noise. This needs to be taken into account when interpreting possible peaks in analysis. Additionally, the use of improved coherence is shown advantageous for decoding auditory attention of hearing impaired. Finally, an application of coherence in evaluation of hearing conditions is also presented, showing possible use in objective evaluation of listening conditions.

Authors' contributions: The idea for the paper was realized by all authors (OK, EA, MAS and MS). OK developed the method changes to multitaper coherence estimation presented in the paper. OK identified the bias caused by $1/f$-noise. OK made the application of methods to data and was main contributor to the writing process. MS, EA and MAS contributed to method development and writing.

III **Neural Tracking of Sustained Attention, Attention Switching, and Natural Conversation in Audiovisual Environments using Wearable EEG**
Johanna Wilroth, **Oskar Keding**, Martin A. Skoglund, Maria Sandsten, Martin En-

qvist, Emina Alickovic
Submitted Manuscript

Scientific contributions: This work advances AAD by introducing a comprehensive wearable EEG framework for tracking selective attention in dynamic, more naturalistic audiovisual listening environments. Unlike most prior studies that often rely on controlled, audio-only paradigms with sustained attention to a single talker, we present a novel dataset and subsequent analysis encompassing sustained attention, attention switching, and unscripted two-talker conversation. By using TRF modelling on all these conditions, this study significantly improves the ecological validity of neural attention tracking and is a step towards bridging the gap between laboratory paradigms and everyday communication scenarios.

We demonstrate that attention-related neural signatures, particularly modulations of the P2 and N400 components, are robustly preserved across listening tasks and stimulus types, enabling reliable decoding even during active attention switches and conversational listening. Our results show that models trained under sustained attention conditions generalize effectively to switching and conversational tasks, supporting the feasibility of task-invariant attention decoding. Furthermore, this work provides systematic evaluations of scalp EEG and around-the-ear cEEGrid electrodes in complex audiovisual environments, highlighting both the promise and current limitations of wearable EEG for real-world applications. Together, these contributions offer methodological guidance and empirical evidence toward the development of practical, neuro-steered hearing technologies capable of operating in realistic, multisensory settings.

Authors' contributions: OK and JW jointly share first authorship. OK, JW, EA and MAS jointly developed the study concept and experimental design. OK and JW prepared AV stimuli, prepared and tested the experimental set-up. EA performed the majority of data collection, with substantial contribution from OK and JW. OK and JW jointly led data preprocessing and analysis, with methodological input from remaining authors (EA, MAS, MS, ME). Interpretation of the quantitive analysis was done by OK, JW, EA and MAS. The manuscript was primarily written by OK and JW, with significant contributions and revisions from EA, MAS, ME and MS.

IV **A Comparison of Common Spatial Patterns Algorithms for Auditory Attention Decoding**
Oskar Keding, Emina Alickovic, Martin A. Skoglund, Maria Sandsten
Submitted Manuscript

Scientific contributions: Classification methods utilizing CSP are proven in EEG classification, in certain fields such as motor imagery and BCI. However, applications of these methods within the field of AAD have been scarse. This paper employs and reports performance of a variety of CSP methods from other field of EEG classification to three different AAD datasets. These datasets are from EEG of listeners that are both NH and hearing impaired, enabling us to compare performance of CSP between the two subsets. Classification of left or right locus of attention has higher accuracy for the dataset with NH participants, while it is still possible to predict locus of attention for individuals with hearing impairment as well. Additionally, the paper continues a discussion from [46], and calculates how much a linear discriminant analysis (LDA) classifier overestimates performance if train and test splits have not been properly performed, such as when ”leave-one-trial-out” is not employed. Specifically, the bias is calculated for the case when test samples share dependence on one or some training samples. This could be employed to estimate true performance of classifiers if papers report performance with improper training and testing split.

Authors' contributions: OK developed the idea for the paper, with support from MS, EA and MAS. OK gathered the CSP methodologies and conducted the application on the different datasets. OK, with support from MS developed the analysis of performance bias. Results analysis and writing was done by OK, in discussion with the other authors. MS, EA and MAS provided substantial feedback on the manuscript.

V **Dynamic Audiovisual Selective Attention Tracking, with Attention Switching**
Oskar Keding, Martin A. Skoglund, Emina Alickovic, Maria Sandsten
Manuscript

Scientific contributions: Expanding the analysis on the dataset presented in Paper III, this paper focuses on the feasibillity of real-time applications of AAD algorithms. By recursively updating model estimates, it is demonstrated that tracking attention robustly in real-time is subject, with high performance in a subset of the subjects. Both stimulus ccorelation-based methods and CSP methods for locus of attention decoding are investigated, showing that locus of attention decoding is more robust and has faster response times during attention switches. Furthermore, it is shown that these locus of attention decoding methods generalize, firstly between conditions, but also between subjects. The models also generalize between conditions from different subjects, which lends support to the promise of a task-invariant attention decoding model, even in real-time applications.

We also discuss the possible combination of different types of information, such as locus of attention estimates and speech reconstruction scores and other EEG based metrics, to make filter based state estimates of the whole attentional state. This incorporates the idea that attention is not a binary state, but rather a continuous states that has more dimensions than just attention to one or the other talker. This is an important step towards practical applications of AAD in hearing devices.

Authors' contributions: OK, MAS and EA developed the idea for the paper. OK performed the majority of analysis, with significant contribution from MAS. Method development was done by OK and results interpretation was done by all authors. OK wrote the manuscript, with contribution and revisions from EA, MAS and MS.

1 Conclusions and Outlook

In this thesis, we have investigated neural markers of selective attention in listeners across various hearing conditions. We have shown that spectral coherence can be used as a measure of cognitive connections between speech features and EEG responses, applicable to the objective evaluation of hearing conditions. We have also shown performance generalization of both stimulus correlation-based methods and locus of attention decoding methods across various conditions and datasets. Furthermore, we showed that both stimulus correlation-based methods and locus of attention decoding methods can be used to track selective attention in real time.

With this being said, there are still many open questions and challenges to be addressed in future research. Although there are many more, we describe here some of the key areas for future investigation.

Understanding the reason for the high variability in performance across subjects is an important step towards developing applicable neural decoding models and reliable, fair technologies. It is common in EEG for some subjects to have little to no measurable neural response to stimuli, and this is also the case in AAD research. However, a subset of subjects thought to be in this category can in fact reach meaningful levels of performance with other types of modelling than then ones implemented in current frameworks. Discerning and understanding the reasons for poor performance is of utmost importance to make sure that the technologies we develop are not only applicable to a subset of the population, but also to make sure that we can make the best use of the data we have and understand the underlying neural processes in more detail.

Additionally, moving away from a classification approach to attention and instead estimating current neural states would be valuable for more information-dense AAD in hearing devices. Incorporating listening effort, engagement, and similar metrics would enhance the understanding of the cognitive processes involved in auditory attention. This, of course, requires both development in the types of experiments conducted and a deeper understanding of how neural processes interact in the brain during listening in everyday scenarios. This is a very exciting area of research, and there are many open questions to be addressed in the future.

References

[1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears", en, *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

[2] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers", en, *Proceedings of the National Academy of Sciences*, vol. 109, no. 29, pp. 11 854–11 859, Jul. 2012. DOI: `10.1073/pnas.1205381109`

[3] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening", en, *Journal of Neurophysiology*, vol. 107, no. 1, pp. 78–89, Jan. 2012. DOI: `10.1152/jn.00297.2011`

[4] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception", en, *Nature*, vol. 485, no. 7397, pp. 233–236, May 2012.

[5] J. A. O'Sullivan et al., "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG", en, *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, Jul. 2015.

[6] B. N. Pasley et al., "Reconstructing Speech from Human Auditory Cortex", en, *PLoS Biology*, vol. 10, no. 1, R. Zatorre, Ed., e1001251, Jan. 2012. DOI: `10.1371/journal.pbio.1001251`

[7] P. Das, C. Brodbeck, J. Z. Simon, and B. Babadi, "Neuro-current response functions: A unified approach to MEG source analysis under the continuous stimuli paradigm", en, *NeuroImage*, vol. 211, p. 116 528, May 2020. DOI: `10.1016/j.neuroimage.2020.116528`

[8] S. Anderson, A. Parbery-Clark, T. White-Schwoch, S. Drehobl, and N. Kraus, "Effects of hearing loss on the subcortical representation of speech cues", en, *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 3030–3038, May 2013. DOI: `10.1121/1.4799804`

[9] E. B. Petersen, M. Wöstmann, J. Obleser, and T. Lunner, "Neural tracking of attended versus ignored speech is differentially affected by hearing loss", en, *Journal of Neurophysiology*, vol. 117, no. 1, pp. 18–27, Jan. 2017. DOI: `10.1152/jn.00527.2016`

[10] R. E. Millman, S. L. Mattys, A. D. Gouws, and G. Prendergast, "Magnified Neural Envelope Coding Predicts Deficits in Speech Perception in Noise", en, *The Journal of Neuroscience*, vol. 37, no. 32, pp. 7727–7736, Aug. 2017. DOI: 10.1523/JNEUROSCI.2722-16.2017

[11] T. Goossens, C. Vercammen, J. Wouters, and A. Van Wieringen, "Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults", en, *Hearing Research*, vol. 370, pp. 189–200, Dec. 2018. DOI: 10.1016/j.heares.2018.07.012

[12] E. Alickovic et al., "Neural Representation Enhanced for Speech and Reduced for Background Noise With a Hearing Aid Noise Reduction Scheme During a Selective Attention Task", *Frontiers in Neuroscience*, vol. 14, p. 846, Sep. 2020. DOI: 10.3389/fnins.2020.00846

[13] E. Alickovic, E. H. N. Ng, L. Fiedler, S. Santurette, H. Innes-Brown, and C. Graversen, "Effects of Hearing Aid Noise Reduction on Early and Late Cortical Representations of Competing Talkers in Noise", *Frontiers in Neuroscience*, vol. 15, p. 636 060, Mar. 2021. DOI: 10.3389/fnins.2021.636060

[14] A. De Cheveigné, D. D. Wong, G. M. Di Liberto, J. Hjortkjær, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis", en, *NeuroImage*, vol. 172, pp. 206–216, May 2018. DOI: 10.1016/j.neuroimage.2018.01.033

[15] J. P. Dmochowski, J. J. Ki, P. DeGuzman, P. Sajda, and L. C. Parra, "Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity", en, *NeuroImage*, vol. 180, pp. 134–146, Oct. 2018. DOI: 10.1016/j.neuroimage.2017.05.037

[16] T. D. Taillez, F. Denk, B. Mirkovic, B. Kollmeier, and B. T. Meyer, "Modeling Nonlinear Transfer Functions from Speech Envelopes to Encephalography with Neural Networks", *International Journal of Psychological Studies*, vol. 11, no. 4, p. 1, Aug. 2019. DOI: 10.5539/ijps.v11n4p1

[17] T. De Taillez, B. Kollmeier, and B. T. Meyer, "Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech", en, *European Journal of Neuroscience*, vol. 51, no. 5, pp. 1234–1241, Mar. 2020. DOI: 10.1111/ejn.13790

[18] M. Thornton, D. Mandic, and T. Reichenbach, "Robust decoding of the speech envelope from EEG recordings through deep neural networks", *Journal of Neural Engineering*, vol. 19, no. 4, p. 046 007, Aug. 2022. DOI: 10.1088/1741-2552/ac7976

[19] C. Puffay et al., "Relating EEG to continuous speech using deep neural networks: A review", *Journal of Neural Engineering*, vol. 20, no. 4, p. 041 003, Aug. 2023. DOI: 10.1088/1741-2552/ace73f

[20] M. L. Hawley, R. Y. Litovsky, and H. S. Colburn, "Speech intelligibility and localization in a multi-source environment", en, *The Journal of the Acoustical Society of America*, vol. 105, no. 6, pp. 3436–3448, Jun. 1999.

[21] S. K. Scott and I. S. Johnsrude, "The neuroanatomical and functional organization of speech perception", en, *Trends in Neurosciences*, vol. 26, no. 2, pp. 100–107, Feb. 2003.
[22] E. M. Zion Golumbic et al., "Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party"", en, *Neuron*, vol. 77, no. 5, pp. 980–991, Mar. 2013.
[23] S. K. Scott, "Identification of a pathway for intelligible speech in the left temporal lobe", *Brain*, vol. 123, no. 12, pp. 2400–2406, Dec. 2000.
[24] S. Geirnaert, T. Francart, and A. Bertrand, "Fast EEG-Based Decoding Of The Directional Focus Of Auditory Attention Using Common Spatial Patterns", *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1557–1568, May 2021.
[25] S. Vandecappelle, L. Deckers, N. Das, A. H. Ansari, A. Bertrand, and T. Francart, "EEG-based detection of the locus of auditory attention with convolutional neural networks", en, *eLife*, vol. 10, e56481, Apr. 2021.
[26] G. Ciccarelli et al., "Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods", en, *Scientific Reports*, vol. 9, no. 1, p. 11 538, Aug. 2019.
[27] E. Alickovic, T. Lunner, F. Gustafsson, and L. Ljung, "A Tutorial on Auditory Attention Identification Methods", *Frontiers in Neuroscience*, vol. 13, p. 153, Mar. 2019.
[28] E. C. Lalor and J. J. Foxe, "Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution", en, *European Journal of Neuroscience*, vol. 31, no. 1, pp. 189–193, Jan. 2010.
[29] C. Brodbeck, L. E. Hong, and J. Z. Simon, "Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech", en, *Current Biology*, vol. 28, no. 24, 3976–3983.e5, Dec. 2018.
[30] C. Daube, R. A. Ince, and J. Gross, "Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech", en, *Current Biology*, vol. 29, no. 12, 1924–1937.e9, Jun. 2019.
[31] J. Schnupp, I. Nelken, and A. J. King, *Auditory Neuroscience: Making Sense of Sound*, en. The MIT Press, 2010.
[32] M. Kegler, H. Weissbart, and T. Reichenbach, "The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information", *Frontiers in Neuroscience*, vol. 16, p. 915 744, Jul. 2022. DOI: `10.3389/fnins.2022.915744`
[33] J. O'Sullivan et al., "Hierarchical Encoding of Attended Auditory Objects in Multi-talker Speech Perception", en, *Neuron*, vol. 104, no. 6, 1195–1209.e3, Dec. 2019. DOI: `10.1016/j.neuron.2019.09.007`
[34] E. S. Teoh, F. Ahmed, and E. C. Lalor, "Attention Differentially Affects Acoustic and Phonetic Feature Encoding in a Multispeaker Environment", en, *The Journal of Neuroscience*, vol. 42, no. 4, pp. 682–691, Jan. 2022. DOI: `10.1523/JNEUROSCI.1455-20.2021`

[35] S. Carta, E. Aličković, J. Zaar, A. L. Valdés, and G. M. Di Liberto, “Cortical encoding of phonetic onsets of both attended and ignored speech in hearing impaired individuals”, en, *PLOS ONE*, vol. 19, no. 11, A. R. Dykstra, Ed., e0308554, Nov. 2024. DOI: 10.1371/journal.pone.0308554

[36] A. D. Bateson, H. A. Baseler, K. S. Paulson, F. Ahmed, and A. U. R. Asghar, “Categorisation of Mobile EEG: A Researcher's Perspective”, en, *BioMed Research International*, vol. 2017, pp. 1–15, 2017. DOI: 10.1155/2017/5496196

[37] L. F. Haas, “Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography”, *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 9–9, Jan. 2003.

[38] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, “Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear”, en, *Scientific Reports*, vol. 5, no. 1, p. 16 743, Nov. 2015. DOI: 10.1038/srep16743

[39] J. A. Urigüen and B. Garcia-Zapirain, “EEG artifact removal—state-of-the-art and guidelines”, *Journal of Neural Engineering*, vol. 12, no. 3, p. 031 001, Jun. 2015.

[40] G. Cybenko, “Approximation by superpositions of a sigmoidal function”, en, *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, Dec. 1989. DOI: 10.1007/BF02551274

[41] A. Pinkus, “Approximation theory of the MLP model in neural networks”, en, *Acta Numerica*, vol. 8, pp. 143–195, Jan. 1999. DOI: 10.1017/S0962492900002919

[42] D.-X. Zhou, “Universality of deep convolutional neural networks”, en, *Applied and Computational Harmonic Analysis*, vol. 48, no. 2, pp. 787–794, Mar. 2020. DOI: 10.1016/j.acha.2019.06.004

[43] A. M. Schäfer and H. G. Zimmermann, “Recurrent Neural Networks Are Universal Approximators”, in *Artificial Neural Networks – ICANN 2006*, D. Hutchison et al., Eds., Series Title: Lecture Notes in Computer Science, vol. 4131, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 632–640. DOI: 10.1007/11840817_66

[44] B. Accou, J. Vanthornhout, H. V. Hamme, and T. Francart, “Decoding of the speech envelope from EEG using the VLAAI deep neural network”, en, *Scientific Reports*, vol. 13, no. 1, p. 812, Jan. 2023. DOI: 10.1038/s41598-022-27332-2

[45] M. A. Tanveer, M. A. Skoglund, B. Bernhardsson, and E. Alickovic, “Deep learning-based auditory attention decoding in listeners with hearing impairment*”, *Journal of Neural Engineering*, vol. 21, no. 3, p. 036 022, Jun. 2024. DOI: 10.1088/1741-2552/ad49d7

[46] I. Rotaru, S. Geirnaert, N. Heintz, I. Van De Ryck, A. Bertrand, and T. Francart, “What are we really decoding? Unveiling biases in EEG-based decoding of the spatial focus of auditory attention”, en, Neuroscience, preprint, Jul. 2023. DOI: 10.1101/2023.07.13.548824

[47] S. Cai, E. Su, Y. Song, L. Xie, and H. Li, “Low Latency Auditory Attention Detection with Common Spatial Pattern Analysis of EEG Signals”, en, in *Interspeech 2020*, ISCA, Oct. 2020, pp. 2772–2776.

[48] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series", *Mathematics of computation*, vol. 19, no. 90, pp. 297–301, 1965.

[49] M. S. Bartlett, "Periodogram Analysis and Continuous Spectra", *Biometrika*, vol. 37, no. 1/2, p. 1, Jun. 1950. DOI: 10.2307/2332141

[50] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms", en, *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[51] D. Thomson, "Spectrum estimation and harmonic analysis", *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[52] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis", *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999. DOI: 10.1109/72.761722

[53] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution", en, *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995. DOI: 10.1162/neco.1995.7.6.1129

[54] H. Peeters, F. Kuk, C.-c. Lau, and D. Keenan, "Subjective and Objective Evaluation of Noise Management Algorithms", en, *Journal of the American Academy of Audiology*, vol. 20, no. 2, pp. 89–98, Feb. 2009. DOI: 10.3766/jaaa.20.2.2

[55] D. Wendt, R. K. Hietkamp, and T. Lunner, "Impact of Noise and Noise Reduction on Processing Effort: A Pupillometry Study", en, *Ear & Hearing*, vol. 38, no. 6, pp. 690–700, Nov. 2017. DOI: 10.1097/AUD.0000000000000454

[56] D. Wendt, T. Koelewijn, P. Książek, S. E. Kramer, and T. Lunner, "Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test", en, *Hearing Research*, vol. 369, pp. 67–78, Nov. 2018. DOI: 10.1016/j.heares.2018.05.006

[57] B. Ohlenforst, D. Wendt, S. E. Kramer, G. Naylor, A. A. Zekveld, and T. Lunner, "Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response", en, *Hearing Research*, vol. 365, pp. 90–99, Aug. 2018. DOI: 10.1016/j.heares.2018.05.003

[58] L. Fiedler, T. Seifi Ala, C. Graversen, E. Alickovic, T. Lunner, and D. Wendt, "Hearing Aid Noise Reduction Lowers the Sustained Listening Effort During Continuous Speech in Noise—A Combined Pupillometry and EEG Study", en, *Ear & Hearing*, vol. 42, no. 6, pp. 1590–1601, Nov. 2021. DOI: 10.1097/AUD.0000000000001050

[59] J. Vanthornhout, L. Decruy, and T. Francart, "Effect of Task and Attention on Neural Tracking of Speech", *Frontiers in Neuroscience*, vol. 13, p. 977, Sep. 2019. DOI: 10.3389/fnins.2019.00977

[60] N. Das, *Auditory Attention Detection Dataset KULeuven*, 2019.

[61] S. A. Fuglsang, D. D. Wong, and J. Hjortkjær, *EEG and audio dataset for auditory attention decoding*, Mar. 2018. DOI: 10.5281/ZENODO.1199011

[62] I. Rotaru, S. Geirnaert, A. Bertrand, and T. Francart, *Audiovisual, Gaze-controlled Auditory Attention Decoding Dataset KU Leuven (AV-GC-AAD)*, Apr. 2024. DOI: 10.5281/ZENODO.11058711

[63] L. Bollens, C. Puffay, B. Accou, J. Vanthornhout, H. V. Hamme, and T. Francart, "Auditory EEG Decoding Challenge for ICASSP 2024", *IEEE Open Journal of Signal Processing*, vol. 6, pp. 478–488, 2025. DOI: 10.1109/OJSP.2025.3534122

[64] M. J. Monesi, L. Bollens, B. Accou, J. Vanthornhout, H. Van Hamme, and T. Francart, "Auditory EEG Decoding Challenge for ICASSP 2023", *IEEE Open Journal of Signal Processing*, vol. 5, pp. 652–661, 2024. DOI: 10.1109/OJSP.2024.3376296

[65] I. Van De Ryck, N. Heintz, I. Rotaru, S. Geirnaert, A. Bertrand, and T. Francart, "EEG-based decoding of auditory attention to conversations with turn-taking speakers", en, *Hearing Research*, vol. 471, p. 109 539, Feb. 2026. DOI: 10.1016/j.heares.2026.109539

[66] L. Straetmans, B. Holtze, S. Debener, M. Jaeger, and B. Mirkovic, "Neural tracking to go: Auditory attention decoding and saliency detection with mobile EEG", *Journal of Neural Engineering*, vol. 18, no. 6, p. 066 054, Dec. 2021. DOI: 10.1088/1741-2552/ac42b5

[67] L. Straetmans, K. Adiloglu, and S. Debener, "Neural speech tracking and auditory attention decoding in everyday life", *Frontiers in Human Neuroscience*, vol. 18, p. 1 483 024, Nov. 2024. DOI: 10.3389/fnhum.2024.1483024

[68] A. J. Kell, D. L. Yamins, E. N. Shook, S. V. Norman-Haignere, and J. H. McDermott, "A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy", en, *Neuron*, vol. 98, no. 3, 630–644.e16, May 2018. DOI: 10.1016/j.neuron.2018.03.044

[69] A. Francl and J. H. McDermott, "Deep neural network models of sound localization reveal how perception is adapted to real-world environments", en, *Nature Human Behaviour*, vol. 6, no. 1, pp. 111–133, Jan. 2022. DOI: 10.1038/s41562-021-01244-z

[70] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-Time Tracking of Selective Auditory Attention From M/EEG: A Bayesian Filtering Approach", *Frontiers in Neuroscience*, vol. 12, p. 262, May 2018. DOI: 10.3389/fnins.2018.00262

[71] E. S. Teoh and E. C. Lalor, "EEG decoding of the target speaker in a cocktail party scenario: Considerations regarding dynamic switching of talker location", *Journal of Neural Engineering*, vol. 16, no. 3, p. 036 017, Jun. 2019. DOI: 10.1088/1741-2552/ab0cf1

[72] S. Haro, H. M. Rao, T. F. Quatieri, and C. J. Smalt, "EEG alpha and pupil diameter reflect endogenous auditory attention switching and listening effort", en, *European Journal of Neuroscience*, vol. 55, no. 5, pp. 1262–1277, Mar. 2022. DOI: 10.1111/ejn.15616

[73] S. Carta, E. Aličković, J. Zaar, A. L. Valdés, and G. M. Di Liberto, *Simultaneous cortical tracking of competing speech streams during attention switching*, en, Jul. 2025. DOI: 10.1101/2025.07.02.662762

[74] M. Daeglau, J. Otten, G. Grimm, B. Mirkovic, V. Hohmann, and S. Debener, "Neural speech tracking in a virtual acoustic environment: Audio-visual benefit for unscripted continuous speech", *Frontiers in Human Neuroscience*, vol. 19, p. 1 560 558, Apr. 2025. DOI: 10.3389/fnhum.2025.1560558

[75] N. Marrone, C. R. Mason, and G. Kidd, "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms", en, *The Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 3064–3075, Nov. 2008.

[76] B. J. Lawrence, D. M. P. Jayakody, R. J. Bennett, R. H. Eikelboom, N. Gasson, and P. L. Friedland, "Hearing Loss and Depression in Older Adults: A Systematic Review and Meta-analysis", en, *The Gerontologist*, vol. 60, no. 3, P. C. Heyn, Ed., e137–e154, Apr. 2020.

[77] D. G. Loughrey, M. E. Kelly, G. A. Kelley, S. Brennan, and B. A. Lawlor, "Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis", en, *JAMA Otolaryngology–Head & Neck Surgery*, vol. 144, no. 2, p. 115, Feb. 2018.

[78] H. Dillon, J. A. Zakis, H. McDermott, G. Keidser, W. Dreschler, and E. Convery, "The trainable hearing aid: What will it do for clients and clinicians?", en, *The Hearing Journal*, vol. 59, no. 4, p. 30, Apr. 2006.

[79] I. Rotaru, S. Geirnaert, N. Heintz, I. Van De Ryck, A. Bertrand, and T. Francart, "What are we really decoding? Unveiling biases in EEG-based decoding of the spatial focus of auditory attention", *Journal of Neural Engineering*, vol. 21, no. 1, p. 016 017, Feb. 2024. DOI: 10.1088/1741-2552/ad2214

[80] C. Elberling and M. Don, "Quality Estimation of Averaged Auditory Brainstem Responses", en, *Scandinavian Audiology*, vol. 13, no. 3, pp. 187–197, Jan. 1984. DOI: 10.3109/01050398409043059

[81] S. Geirnaert, T. Francart, and A. Bertrand, "Unsupervised Self-Adaptive Auditory Attention Decoding", *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3955–3966, Oct. 2021. DOI: 10.1109/JBHI.2021.3075631

[82] S. Geirnaert, T. Francart, and A. Bertrand, "Time-Adaptive Unsupervised Auditory Attention Decoding Using EEG-Based Stimulus Reconstruction", *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3767–3778, Aug. 2022. DOI: 10.1109/JBHI.2022.3162760

[83] V. Viswanathan, H. M. Bharadwaj, and B. G. Shinn-Cunningham, "Electroencephalographic Signatures of the Neural Representation of Speech during Selective Attention", en, *eneuro*, vol. 6, no. 5, ENEURO.0057–19.2019, Sep. 2019.

Scientific Publications

Printed by Media-Tryck, Lund 2026 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Doctoral Theses in Mathematical Sciences 2026:6
ISBN 978-91-8104-982-4
ISSN 1404-0034