
# Investigating cortical encoding of auditory space and motion in humans using EEG

Adam Bednar, MSc.

Under the supervision of

Dr Edmund C. Lalor

A dissertation submitted to the

University of Dublin, Trinity College

In fulfilment of the requirements for the degree of

Doctor of Philosophy

April 2020


Department of Electronic and Electrical Engineering

University of Dublin, Trinity College

# Declaration

I, Adam Bednar, confirm that this thesis has not been submitted as an exercise for a degree at this or any other university and is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed,

_____

Adam Bednar

April 2020

# Summary

We can clearly perceive the location of moving auditory objects in space. However, the cortical representation of auditory space, and auditory motion in particular, is not well characterized. This work uses a novel linear regression-based framework together with scalp recorded electroencephalography (EEG) to study various aspects of spatial hearing in humans.

Previously, it has been shown that the sound position is cortically represented by a population of broadly tuned neurons. However, it is unclear how this cortical activity reflects the time-varying location of a moving sound. In our first study, we showed that in an acoustic scene with one sound source, auditory cortex tracks the time-varying location of a continuously moving sound. Specifically, we identified two distinct frequency components, namely, delta (0-2Hz) and the alpha power (8-12Hz) of EEG that track the sound location. The delta and the power of alpha EEG encoding had different spatio-temporal characteristics, which suggested that they potentially reflect different aspects of auditory motion processing. Importantly, we also showed that the trajectory tracking is not specific to a particular type of spatial acoustic cue and is independent from the well-known sound envelope tracking of the cortex.

In natural settings, we are almost exclusively presented with multiple competing sounds and so we typically focus our attention on the relevant source in order to segregate it from the competing sources. While many studies have examined this in the context of sound envelope tracking by the cortex, it is unclear how we process and utilize spatial information in such scenarios. In our second study we created an experiment where subjects listened to two concurrent sound stimuli that were moving independently within the horizontal plane and were tasked with paying attention to one of them. We showed that the attended sound source trajectory can be reliably reconstructed from EEG, even in the presence of other competing sources and demonstrated that the trajectory tracking works for noise as well as more complex speech stimuli. We also observed weak tracking of the unattended source location for the speech stimuli, however, this applied only to delta but not to the alpha power EEG component. This further suggests that location tracking by delta and alpha power EEG possibly represent different neural mechanisms. Finally, with more practical applications in mind, we demonstrated that the trajectory reconstruction approach can be used to decode selective attention.

Following on from our study on multiple moving sounds, we were then interested to study how higher orders of motion are processed in the brain. Psychophysical studies have shown

that we are sensitive to sound velocity as well as acceleration. Although cortical sensitivity to motion has been demonstrated by contrasting static and moving sounds, it is unclear to which particular aspects of motion neurons are sensitive. This has been tested in our third study, where we investigated cortical sensitivity to varying position, velocity, speed and acceleration. We found that sound speed but not velocity can be reconstructed from EEG independently from sound position. Surprisingly, our results also indicated that sound acceleration might be independently represented at the cortical level, which has not been reported before.

Finally, in the last study, we deployed our reconstruction method in a naturalistic scenario where subjects were allowed to move their heads and received visual input over a virtual reality headset. We were primarily interested in whether sound location is cortically encoded using cranio- or allo-centric coordinates. Although our initial analysis indicated a cranio-centric representation of sound location, we were not able to reconstruct the trajectory from EEG after we removed the head motion-related artefacts. Therefore, we were unable to find strong evidence for cortical encoding in either frame of reference. Our secondary goal was to test the feasibility of using the Oculus Rift headset together with EEG recording. Although we found it is possible to use this setup, we have encountered several practical issues, such as subject discomfort during longer recordings that need to be addressed.

Here in this thesis we demonstrated that the stimulus reconstruction framework can be used to study human spatial hearing. Using this framework that enables relating a continuous representation of stimulus spatial properties to un-epoched and un-averaged EEG data, we have addressed several important questions related to the dynamics of cortical spatial encoding and the neural representation of sound motion. Moreover, we hope that this work will help future studies on sound localization to employ a wider variety of sound stimuli and to create more naturalistic acoustic scenarios.

# Acknowledgements

This work would not have been possible without the help and support of many others. I would like to thank:

Ed, for trusting me and for your guidance. You have been a great supervisor and I have learnt a lot from you.

Richard, for all your kind advice and help you gave me.

Aisling, thank you also for making my time in Rochester exciting and for all the wild trips we went on. I also thank you for all the help with proofreading my work - sorry that your English is never going to be the same.

I would also like to give special thanks to my MSc and PhD mates Emily and Shruti. Thank you for all the fun and crazy times in the lab and elsewhere.

Past and current members of the Lalor Lab: Mick, James, Gio (eooo!), Denis, Michael, Andy, Kevin, Nate, Aaron, Shy, Lauren and members of the Reilly Lab: Céline, Isabelle, Martin, Alejandro, Brendan, Niamh, Saskia, Ciara, Surbhi, Shruti and Clodagh.

I would also like to thank those who participated as test subjects in the experiments. Your brains gave me good data!

I would also like to thank my family: Rád bych poděkoval rodičům Lence a Karlovi. Bez vaší podpory a rodičovské lásky by tato práce nikdy nevznikla. Dále děkuji babičkám Alence a Babišákovi. A v neposlední řadě má moje díky brácha Bobo.

This work could not have been completed without the support of the Irish Research Council for Science, Engineering and Technology (GOIPG/2015/1656).

# Table of Contents

# Publications Arising from this Thesis

## Journal Articles

- **Bednar, A.**, & Lalor, E. C. (2018). Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *NeuroImage*, 181, 683-691.

- Crosse, M. J., Di Liberto, G. M., **Bednar, A**., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in human neuroscience*, 10, 604.

## Journal Articles in Review

- **Bednar, A.**, & Lalor, E. C. (in review). Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. *NeuroImage*

## Journal Articles in Preparation

- **Bednar, A.**, & Lalor, E. C. (in prep). Cortical representation of higher orders of auditory motion: Decoding speed and acceleration of moving sound source

## Conference Poster Presentations

- **Bednar, A.**, & Lalor, E. C. Decoding locations of concurrent moving sound sources using EEG. *Society for Neuroscience 2018*, San Diego, USA

- **Bednar, A.**, & Lalor, E. C. Decoding locations of concurrent moving sound sources using EEG. *Advances and Perspectives in Auditory Neuroscience 2018*, San Diego, USA

- **Bednar, A.**, & Lalor, E. C. Decoding the cortical representation of auditory motion using EEG. *6th International Conference on Auditory Cortex 2017*, Banff, Canada

- **Bednar, A.**, & Lalor, E. C. Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *Society for Neuroscience 2017*, Washington DC, USA

- **Bednar, A.**, & Lalor, E. C. Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *Advances and Perspectives in Auditory Neuroscience 2017*, Washington DC, USA

# Glossary of Acronyms

| | |
|---|---|
| AC | Auditory Cortex |
| CN | Cochlear Nucleus |
| DCN | Dorsal Cochlear Nucleus |
| DLPC | Dorsolateral Prefrontal Cortex |
| DNLL | Dorsal Nuclei of Lateral Lemniscus |
| EEG | Electroencephalography |
| ERP | Event-Related Potential |
| fMRI | functional Magnetic Resonance |
| HRTF | Head-Related Transfer Function |
| IC | Inferior Colliculus |
| IFC | Inferior Frontal Cortex |
| ILD | Interaural Level Difference |
| IPL | Inferior Parietal Lobule |
| ITD | Interaural Time Difference |
| MAA | Minimal Audible Angle |
| MAMA | Minimum Audible Movement Angle |
| MEG | Magnetoencephalography |
| PMC | Premotor Cortex |
| pSTG | Posterior Superior Temporal Gyrus |
| PT | Planum Temporale |
| SOC | Superior Olivary Complex |
| VAS | Virtual Acoustic Space |
| VCN | Ventral Cochlear Nucleus |

# Chapter 1. Introduction

## 1.1. Background

Spatial hearing, the ability to localize sounds in the environment, is crucial in our everyday life. This applies particularly to moving sources, which from an evolutionary perspective, often present a potential threat. For example, spatial hearing is important for detecting a predator moving within the bushes, and in modern life, realising a car is approaching so that we can avoid collision or localizing a mosquito that is buzzing around our bed during the night so that we can swat it.

At a first glance, when comparing our spatial hearing ability with other senses – particularly vision, one might consider spatial hearing to be inferior due to its low spatial acuity. That is, using hearing without the help of other senses, we can only estimate the approximate location of the object that makes sound. In contrast, vision allows us to pinpoint the location of every miniscule object within our visual field. However, one should not forget that our field of view is limited, and for example, we cannot see objects behind us. On the other hand, hearing allows localization of sounds all around us, and due to diffraction, we can even monitor sound sources behind obstacles and around corners, which is something that we would not be able to accomplish with vision. Therefore, human spatial hearing and vision can be considered complementary for localizing objects in our environment.

Importantly, in contrast to visual or somatosensory systems, auditory space does not have an explicit topographic representation on the receptor surface as the hair cells within the cochlea are spatially organized according to the frequency content of the incoming sound. Instead, the central auditory system calculates the sound location position based on so called spatial acoustic cues. The horizontal angle, azimuth, of the sound is mainly determined based on differences in intensity and arrival time between the ears, which are caused by the shape and shadowing effects of our head and body. The vertical sound angle, elevation, can be calculated by analyzing the spectral information of the sound as the interaction between the incoming sound and the outer ear as well as our bodies which cause an elevation-dependent filtering effect. Finally, the distance of the sound source can be estimated from the change in overall sound intensity as well as characteristic spectral changes.

Clearly, localizing a sound source is not a trivial task, and despite the importance of spatial hearing and the long line of research, which dates back to 1876 when Lord Raleigh investigated hearing acuity, the exact mechanisms of how we process and localize sounds

are still unknown. The central auditory system consists of a highly complex and hierarchical neural network. There are multiple parallel ascending and descending neuronal streams, some of them crossing over and projecting to the contralateral pathways. While the subcortical regions play an important role in translating the spatial acoustic cues into a neural spiking signal, the synthesis of auditory space is thought to be done at the higher neural levels in auditory cortex. Here, it appears that the sound location is represented by populations of non-topographically organized neurons, which have relatively broad tuning and respond primarily to sounds from the contralateral part of the auditory space. However, how exactly the positional information is encoded using these neurons is unclear.

Even less is known about how we encode dynamic spatial properties of a sound such as motion direction, velocity and acceleration. In vision, it has been shown that there are low-level motion detectors that respond to motion velocity and direction. However, in the auditory domain, it is unknown whether we have specialized motion detectors, or whether we analyze motion at higher levels of the auditory pathway using the same networks that are involved in localizing static sources. A related interesting and unanswered question is how we keep stable perception of our auditory environment despite our constant head motion. Every time we move our head, the environment moves in the opposite direction relative to our ears. Therefore, in order to account for that, the auditory system must incorporate the proprioceptive information of the body.

In real-world scenarios we are often surrounded by multiple overlapping sound sources. In such situations, we are often trying to attend only the most relevant source while ignoring the others. This has been investigated in the so-called cocktail party effect, which describes our remarkable ability to selectively attend a single conversation in a complex acoustic environment such as a noisy bar. Although it might not be obvious, spatial hearing plays an important role in this as it helps us to separate sources and form auditory objects in complex acoustic scenes with multiple concurrent sounds. This is well demonstrated in people with spatial hearing loss, which is a relatively common impairment that manifests by the inability to distinguish spatial cues. As a result of decreased spatial hearing ability, people with this impairment have issues with hearing and processing speech in the presence of background noise. All this suggests that we can process overlapping spatial cues corresponding to several sound sources and that the spatial representations of multiple sound sources exists somewhere within the ascending auditory pathway. Nevertheless, the neural underpinnings of spatial coding in more complex multi-source scenarios when the auditory system needs to localize several overlapping sounds is unknown.

Above, we have shown that spatial hearing is important in our everyday life and we have discussed that many of the neural mechanisms that are driving our ability to localize sound are not well understood. Besides the large complexity of these processes, another important reason that prevents us from better understanding the neurophysiology of the central auditory system and sound localization is methodological. One issue with hearing research is the requirement for an imaging method with high temporal resolution, as most of the information is encoded within the fine structure of the sound waveform. Another issue that is more specific to spatial hearing is how to deliver high-fidelity spatialized sound to the listener. For free-field sound presentation, experiments usually employ large loudspeaker arrays. This solution is expensive and not possible in studies utilizing functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG) for technical reasons. Only a relatively recent development of virtual auditory environments brought about an affordable way to present spatial sound over headphones. This method is based on application of a head-related transfer function (HRTF), which describes the location-dependent filtering properties of the head and pinna, on the sound.

The work in this thesis is centred on investigating various aspects of auditory motion using non-invasive electroencephalography (EEG) and multivariate linear regression data analysis. The traditional method used to analyze the EEG data is to average a large number of neural responses to repeated stimulus presentations and extract the so-called evoked related potential (ERP) waveform. However, the main drawback of this approach is the requirement for hundreds of repetitions of a discrete stimulus, which limits its experimental applications. Here, we used multivariate linear regression as our primary method for EEG data analysis. This was done for several reasons: Firstly, it allows the use of continuous stimuli that do not need to be repeated, which is particularly important for studying motion perception. Secondly, as a multivariate measure, it combines information from all EEG electrodes, and so has the potential to be more sensitive than the traditional univariate ERP analysis.

## 1.2. Aims of This Work

The overarching aim of this work is to investigate the neural signatures of spatial hearing with emphasis on the localization of moving sounds. Specifically, it examines whether the cortical activity reflects the time-varying location of a moving sound and whether the brain tracks the position of multiple concurrent sound sources. We also aim to examine if the cortical activity reflects other dynamic measures of auditory motion such as velocity and acceleration. Finally, we aim to investigate the spatial encoding of sound when listeners

move their heads.

## 1.3. Thesis Outline

Chapter 2 introduces different acoustic cues that are used in spatial hearing and describes basic psychophysical measures that are used to evaluate spatial hearing ability. The following section of this chapter describes the basic principles of electroencephalography (EEG) and discusses frameworks that are used to process the EEG data. Finally, the chapter discusses the neural anatomy of sound localization. It briefly describes the subcortical processing of spatial acoustic cues and mainly focuses on the cortical encoding of sound location and cortical sensitivity to sound motion.

In chapter 3, a novel approach that allows investigation of the cortical correlates of auditory motion is described over two experiments. The first experiment shows that the time-varying sound azimuth of a moving sound can be reconstructed from the scalp recorded EEG. Specifically, it demonstrates that two frequency components of EEG, namely delta and the power of alpha, track the sound location. The second experiment shows that this cortical tracking of sound azimuth is present even for impoverished sound stimuli, which contain either interaural time- or level differences (ITD and ILD).

Chapter 4 investigates whether the cortical activity reflects the locations of multiple sounds in complex acoustic environments. Specifically, it tests this in an environment with two moving concurrent noise stimuli, which are describable based on their frequencies, and two speech stimuli that are dynamically changing in their sound source location. It shows that the attended sound source can be reliably decoded even in the presence of distracting non-stationary auditory stimulus. It also demonstrates that the cortical representation of the unattended sound location is much weaker.

Chapter 5 explores whether higher-order measures of sound motion, namely velocity, speed and acceleration, are encoded within the cortex. In this chapter, it is demonstrated that sound speed and acceleration, but not velocity, can be reconstructed from EEG independently on sound azimuth.

In chapter 6, it is investigated whether it is possible to reconstruct a moving sound trajectory using both cranio- and allo-centric coordinates from EEG while subjects listen to a moving sound and rotate their heads at the same time independently on the sound position. The experiment shows that the sound trajectory can be successfully reconstructed using craniocentric coordinates only when motion artefacts are not controlled for and that the reconstruction accuracy drops below the chance level when we partial-out the head

trajectory. The study also demonstrates the feasibility of using Oculus Rift VR headset for stimulus presentation together with high-density EEG recording.

# Chapter 2. Background

## 2.1. Sound Localization Cues

The basic acoustic cues driving sound localization were discovered during early psychophysical experiments (Rayleigh, 1875; Thompson, 1878; Rayleigh, 1907). Binaural cues such as the interaural level difference (ILD) and the interaural time difference (ITD) are based on the differences in sound perceived by the left and the right ear and are mainly used for sound localization in the horizontal plane. Monaural cues, which are also known as spectral cues, are mainly used to determine the sound elevation and are caused by location-sensitive filtering by the listeners head. See figure 2-1 for an overview of acoustic cues. The individual acoustic cues are described in sections 2.1.1-2.1.3 in more detail.

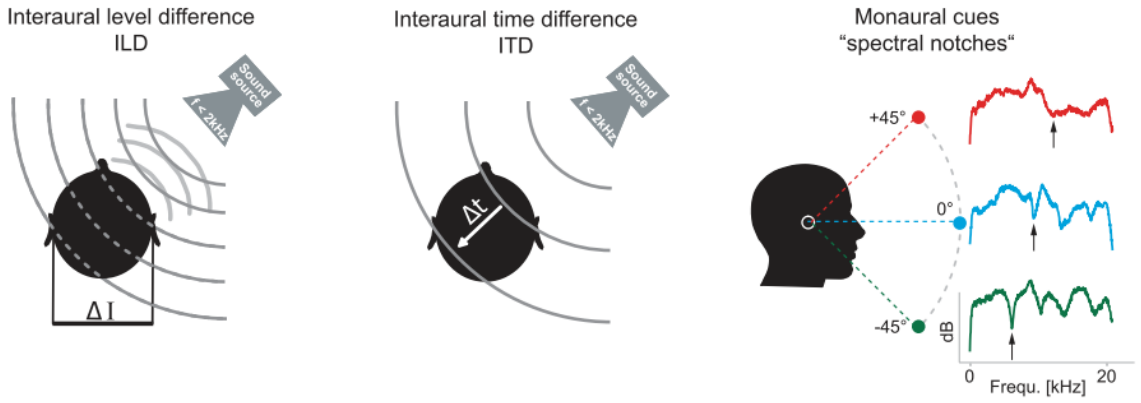



*Figure 2-1. Acoustic sound localization cues. Binaural cues interaural time and level difference are mainly used to localize sound in the horizontal plane. Monaural cues are predominantly used for vertical sound localization and in situations where binaural cues are ambiguous. Adapted from Grothe et al. (2010a)*

According to the well-established "duplex theory", ILD and ITD have complementary roles (Rayleigh, 1875). This theory states that ILD cues are mainly used for sound localization of high frequency sounds, while ITD is used for determining the location of low frequency sounds. Although it has its limitations, e.g. it is possible to extract ITD from the envelope of a rapidly moving signal (Henning, 1974), psychophysical experiments have shown that this theory is still valid (Macpherson & Middlebrooks, 2002). Figure 2-2 shows sensitivity to ILD and ITD cues for different sound frequencies.

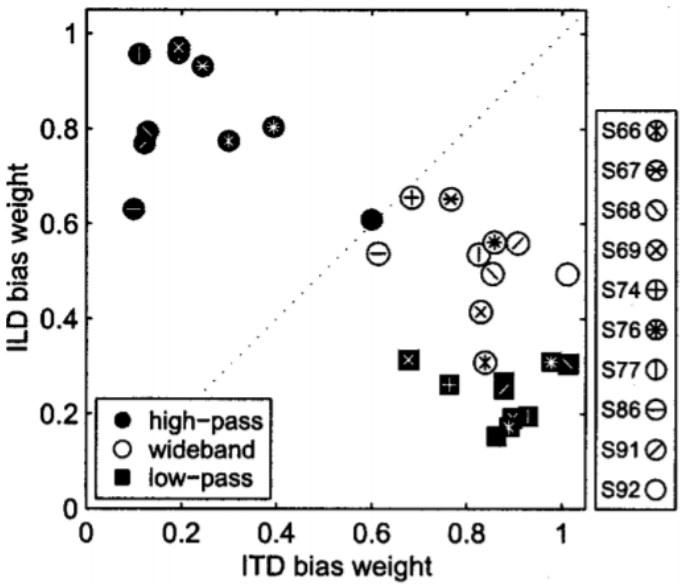



*Figure 2-2. Measured ITD- and ILD-bias weights for individual listeners shown for the wideband (0.5–16 kHz), low-pass (0.5–2 kHz) and high-pass (4–16 kHz) sound stimuli. Listeners gave high weight to ITD and low weight to ILD for low-pass stimuli. For high-pass stimuli, listeners gave high weight to ILD and most listeners gave low weight to ITD. Adapted from Macpherson and Middlebrooks (2002)*

The spectral acoustic cues are monaural and are primarily used for sound localization in the vertical plane, or in situations where the binaural cues are ambiguous. This ambiguity arises when one cue, ITD or ILD, can encode for more than one spatial location. Such situations often occur for sound sources placed within the median plane, where one cannot identify whether the sound is coming from the front or behind us, as ITD and ILD are zero in both cases. A similar scenario of front-back confusion occurs when sources are positioned symmetrically with respect to the interaural axis and lie on a so-called "cone of confusion". As mentioned above, spectral cues can help in resolving this ambiguity. However, in natural conditions, a listener can move his or her head, which is an obvious solution enabling the exploitation of all available cues in these scenarios. Using this mechanism, also known as "Wallach cue", the listener can adjust the position of sound sources relative to the head so that the binaural differences become more favourable (Wallach, 1939). In the following section, the acoustic cues are described in more detail. Further information can be found in reviews (Middlebrooks & Green, 1991; Blauert, 1997; Stecker & Gallun, 2012).

### 2.1.1. Interaural Level Difference

The interaural level difference (ILD), which is sometimes referred to as interaural intensity difference (IID), is caused by the attenuation, reflection and refraction of the sound when interacting with the head and outer ear ("head shadowing effect"). As a result of this

interaction, the sound coming to the closer ear will be louder, whereas the one arriving to the further ear will be softer. When a sound is positioned at the midline relative to the listener, the resultant ILD will be 0 dB. On the contrary, the largest ILD, which is around 20 dB, is achieved when the sources are placed at the most lateral positions.

The availability of ILD depends on the stimulus frequency. For low frequency sounds with long wavelengths, the sound diffracts around the head with little attenuation. For frequencies below 500 Hz, the ILD is close to 0 dB and therefore is not informative for sound localization. For a sound with a frequency of 6 kHz, the ILD can be larger than 20 dB (Feddersen *et al.*, 1957). The ILD as a function of stimulus frequency is shown below in figure 2-3. The estimation of ILD is more complicated for sound sources near (<1 m) to the listener. For such sources, the ILDs can reach up to 30 dB and unlike for distant sources (>1m) the ILDs are also present for low frequencies (Brungart & Rabinowitz, 1999).

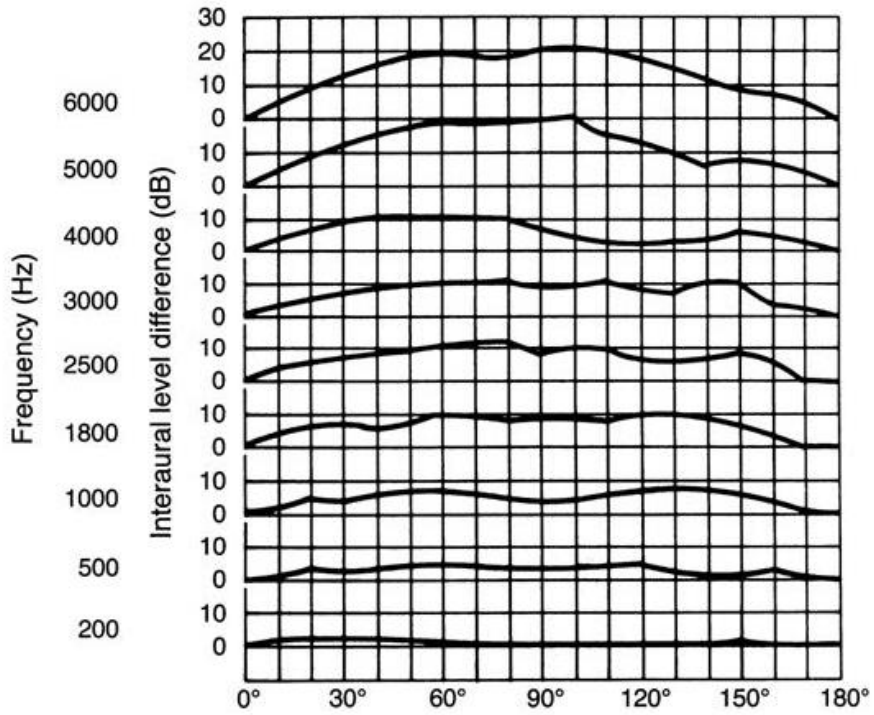



*Figure 2-3. Interaural level differences (ILDs) shown for different azimuth angles and stimulus frequencies. Adapted from Feddersen et al. (1957).*

### 2.1.2. Interaural Time Difference

Originally described by (Thompson, 1878), the interaural time difference (ITD) describes the different timing of sound waveform arrival or phase difference between two ears. For lateral stimuli, the sound emitted by the source arrives earlier to the closer ear and later to the ear that is further away. This timing information can be used analogically to ILD to

determine the azimuth of the sound source. For adults, the range of possible ITDs over all azimuths span approximately within the interval of 0-750 µs (Kuhn, 1977). The temporal information can be extracted from the signal in several ways: (1) from the temporal fine structure of a changing signal, or (2) from larger-scale temporal fluctuations of the sound e.g. amplitude envelope or sound energy onset (Blauert, 1997).

As in the case of ILD, the ITD availability depends on the sound stimulus frequency content. For pure tones and higher frequencies (~ >700 Hz) the information carried by sound phase becomes ambiguous and diminishes further at around 1.6 kHz (see figure 2-4). For complex (non-tonal) stimuli, the ITD information can be obtained even for higher frequencies from the signal envelope (Henning, 1974).

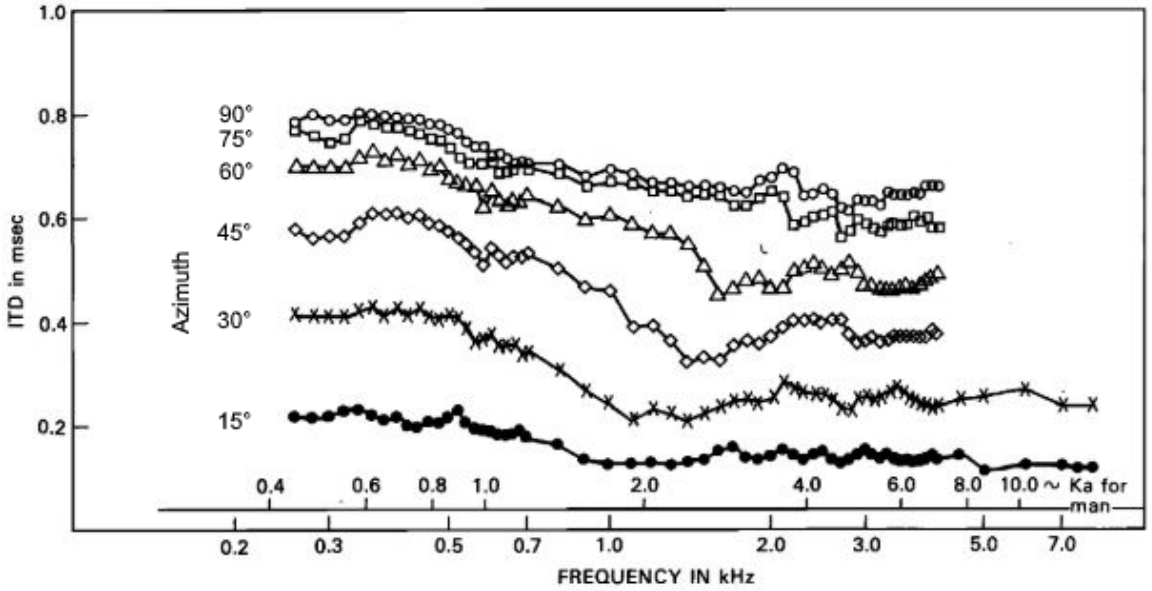



*Figure 2-4. Interaural time differences (ITDs) shown as function of stimulus frequency and sound angle. Adapted from Kuhn (1977).*

### 2.1.3. Spectral Cues

As mentioned above, relying solely on binaural cues for sound localization in the horizontal plane causes front-back confusion for sound sources placed within the midplane and the cone of confusion. Moreover, binaural cues do not carry information about the sound source elevation. In these situations, one employs the spectral (monaural) cues. The spectral cues are, as the name suggests, caused by the location-dependent spectral changes introduced in the incoming sound, which are mainly caused by the interaction of the acoustic signal with the outer ear, head and torso of the listener (Blauert, 1997). As can be seen in figure 2-5, the sound is attenuated differently at different stimulus frequencies and, importantly for sound localization purposes, changes as a function of sound position. This complex relationship

between the sound direction and frequency filtering is characterized by the head-related transfer function (HRTF). Interestingly, the HRTFs exhibit relatively large differences between individuals, which is due to large variability in the outer ear shape within population. An important implication is that when one presents sound via headphones, without incorporating individualized spectral cues, the sounds appear "internalized" i.e. are perceived to originate from the inside of the head. In the case of virtual acoustic spaces (VAS), where sound is presented via headphones and modified using appropriate (individualized) HRTFs, the sound is perceived more "externalized" i.e. coming from the outside of the head.

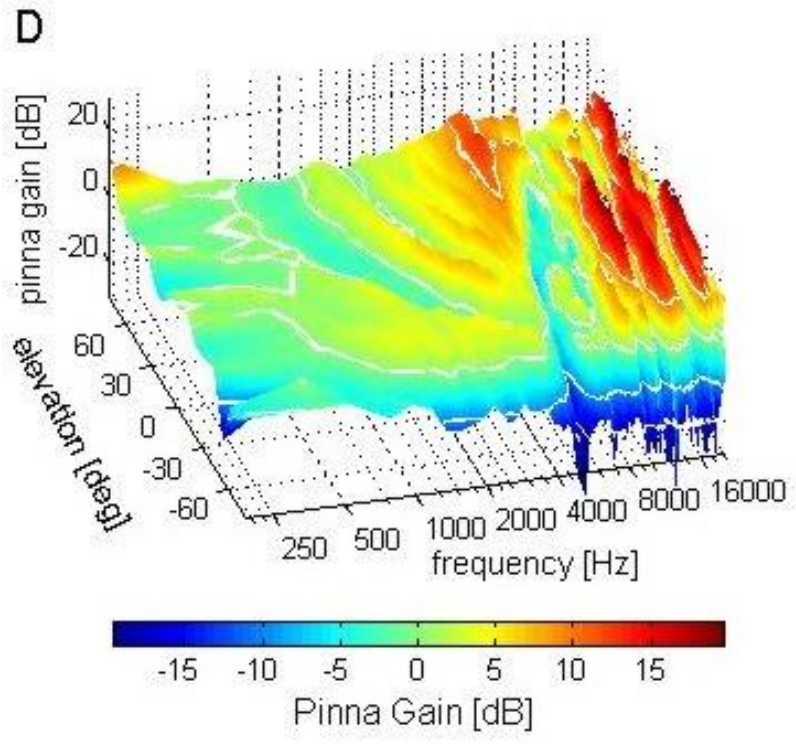



*Figure 2-5. Spectral (monaural cues) are caused by sound location-dependent filtering of outer ear, head and body. The sound attenuation is shown as a function of stimulus frequency and sound source azimuth. Adapted from Schnupp et al. (2011)*

## 2.2. Psychophysical Measures of Spatial Sensitivity

Our ability to localize sound depends on the position of the sound source, as our spatial sensitivity is not uniform across the whole auditory space. Below we introduce the most common measures to quantify spatial sensitivity and discuss how this sensitivity changes for different spatial locations.

For stationary sound sources, the common relative measure of spatial sensitivity is minimal

audible angle (MAA), which is described as the smallest perceivable spatial separation of two subsequent sounds (Mills, 1958). Typically, subjects are asked whether the stimulus was on the left or right from the preceding sound. For the sources positioned on a frontal midline, the detectable azimuthal distance could be as low as 1°. For lateral locations, the MAA is around 7°. For sources varying in elevation the MAA is larger than for azimuth, and is larger than 4° (Perrott & Saberi, 1990). However, the MAA values are only approximate as the localization resolution depends on the spectro-temporal properties of the sound stimulus and acoustic properties of the test environment e.g. psychophysical studies have shown that it is easier to localize broadband than tonal stimuli (Recanzone *et al.*, 1998).

A related measure of spatial sensitivity is minimum audible movement angle (MAMA), which is the minimum distance that the sound needs to travel to be perceived as moving. For slowly moving sources the MAMA is around 9° and it increases with velocity up to approximately 20°. Similarly, to MAA, the MAMA strongly depends on the stimulus characteristics (Chandler & Grantham, 1992; Carlile & Leung, 2016). One alternative to the relative measures described above, is to measure sound localization ability in absolute terms. This is usually done by asking the listeners to point or otherwise indicate the perceived origin of the sound source. The range of source estimates span roughly between 2-30° based on source location and shows relatively large differences between studies (Wightman & Kistler, 1989; Makous & Middlebrooks, 1990; Recanzone *et al.*, 1998).

## 2.3. Neuroimaging

There are several different neuroimaging methods that are used to study the human auditory nervous system. Here in this section, the most common techniques are introduced, and the next sections of this chapter focus specifically on electroencephalography, which is the imaging method that was employed in this work.

Electroencephalography (EEG) and magnetoencephalography (MEG) are both non-invasive methods that directly measure neural activity. EEG measures the electrical field that is caused by synchronous activation of large groups of cortical neurons. MEG is principally like EEG; however, it measures the induced magnetic field. Both imaging methods have the capability to capture fast neural dynamics, but their spatial resolution is worse than the below mentioned imaging methods. Although MEG has superior spatial resolution to EEG, the required acquisition system is relatively expensive and therefore is less common.

Both EEG as well as MEG techniques have been extensively used to study cortical encoding of static sounds (e.g. Palomaki *et al.*,2000, 2005; Briley *et al.*, 2013 and others). Also, due to its ability to capture fast dynamics of cortical activity, these methods seem particularly

suitable for studying location encoding of moving sound sources (e.g. Xiang et al., 2002; Getzmann *et al.*, 2012; Magezi *et al.*, 2013). Although these imaging techniques appear ideal to study location encoding of moving sound sources in humans, challenges exist in how to process such signals and how to improve their low signal-to-noise ratio (see section 2.3.2 below). Specifically, the traditional EEG analysis requires many repetitions of short and discrete stimuli. As a result, sound localization studies typically employed some type of transient sound stimuli (typically noise bursts) that were repeatedly presented from different spatial locations or were moving along short trajectories.

Besides this being arguably a non-naturalistic acoustic scenario, it is difficult to use such methods to answer a question on whether cortical activity reflects the on-going dynamics of a continuously moving sound beyond motion- onset responses. In section 2.3.3 we propose a novel method to study sound localization using EEG that allows to use continuous and unrepeated stimuli.

Electrocorticography (ECoG), is an invasive method that measures electrical activity of neurons directly on the exposed brain surface. This recording method requires a craniotomy and therefore it is usually limited to patients that undergo brain surgery for clinical reasons such as epilepsy. Also, ECoG electrodes typically cover only a limited brain area of a patient, which further limits this method. Nevertheless, this imaging method has superior signal-to-noise ratio as well as high spatial and temporal resolution. To our best knowledge, ECoG has not been used to study sound localization in humans.

Functional magnetic resonance (fMRI) and positron emission tomography (PET) measure brain activity indirectly by detecting local changes in blood flow and metabolic activity. fMRI does this by detecting magnetic responses of tissue to externally applied magnetic fields that are delivered by coils of the MRI scanner. Specifically, it measures cognitive activity-induced changes in blood oxygenation level, which are reflected by the amount of oxyhemoglobin and its form without oxygen, deoxyhemoglobin, in red blood cells that differ in their magnetic properties. PET measures radioactive emission caused by the decay of positron-emitting isotopes that are introduced to the blood stream of a patient. An increase in brain activity causes increased uptake of a radioactive tracer, which is reflected by a change in emitted radiation (Webster, 2009). These two methods have good spatial resolution; however, their temporal resolution is relatively poor.

The use of fMRI/PET in spatial hearing research is relatively common and has been useful to identify brain regions that are active during sound localization process (see extensive meta-study of human imaging studies Arnott *et al.*, 2004). A typical experiment in such

studies is contrasting brain activity during spatial and non-spatial tasks e.g. determining stimulus location and pitch discrimination (e.g. Alain *et al*., 2001). Another example of using fMRI in sound localization studies can be determining cortical sensitivities to ITD and ILD cues (e.g. Higgins *et al*., 2017). However, as the temporal resolution of these methods is relatively poor and these methods are not very suitable for studying highly dynamic stimuli, they have been mostly used to study sound localization using stationary sound sources. Although there are several fMRI studies investigating sound motion, they have usually restricted the analysis to contrasting responses to moving sound and 'corresponding' stationary sounds, which is often problematic from a methodological point of view (Smith *et al*, 2007, Poirier 2017). Nevertheless, novel paradigms to study encoding of auditory motion using fMRI have started to emerge (e.g. Ortiz-Rios *et al*, 2017). Finally, another limitation of the fMRI studies is that it is not possible to use loudspeaker arrays for sound delivery and one is restricted to sound virtualization techniques and fMRI compatible headphones.

### 2.3.1. Electroencephalography

Electroencephalography (EEG) is a non-invasive imaging method that measures the electrical activity of the brain using scalp electrodes. The electrical potential on the cortical surface is generated mainly by the net effect of local postsynaptic potentials of many cortical neurons. A figure of a typical pyramidal cell neuron is shown in figure 2-6. The pyramidal cells are the most common type of cortical neuron and are predominantly oriented perpendicular to the scalp surface. Each neuron consists of a dendritic tree with many branching dendrites that receive input from other neurons, cell body (soma) and an axon, which act as an output to other neurons. In the case of excitatory input, the dendritic membrane depolarizes and causes current flow from the soma, which acts as a current source, to the dendritic part that behaves like a current sink. In the case of inhibitory input, the polarity is reversed and the current flows in the opposite direction. In effect, each of the pyramidal cells act as a radially oriented electrical dipole and its polarity depends on whether it receives excitatory or inhibitory input. When a large number of these neurons synchronously activates it generates an electrical field which is measurable as a unipolar potential on the scalp surface (Webster, 2009).

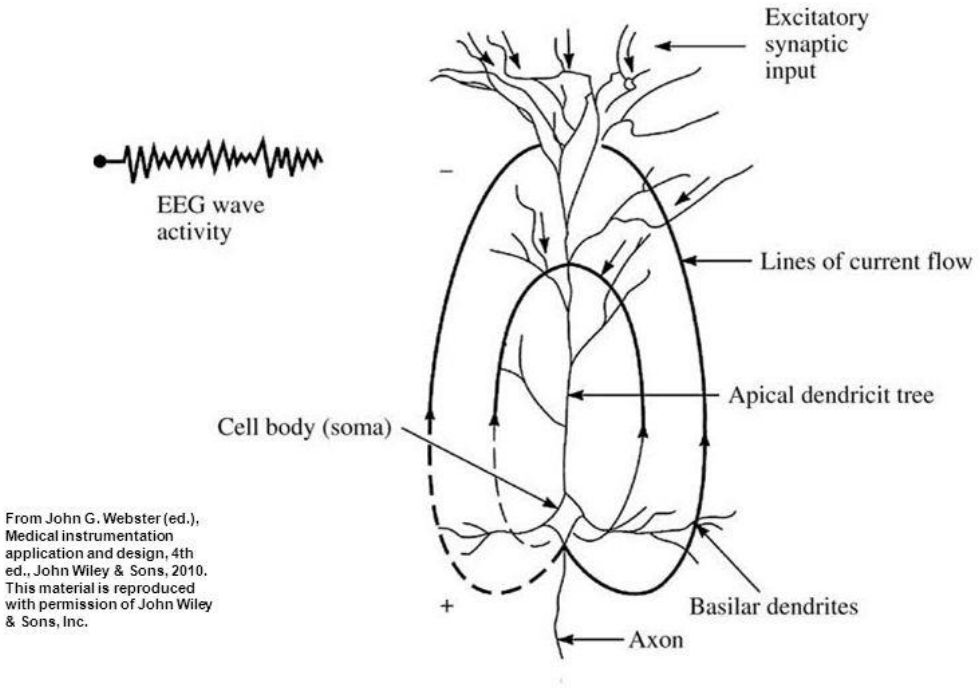



*Figure 2-6. Generation of cortical field potential in the case of an excitatory input. The dendritic tree acts as a current sink and the neuron body (soma) acts as a current source. The generated electrical field can be approximated as an electrical dipole. Adapted from Webster (2009)*

Since the measuring electrodes are placed on the scalp, the electric field generated by neurons is attenuated when passing through several layers of tissue i.e. the meninges covering the brain, the cerebrospinal fluid, the bone and skin. This also leads to volume conduction which causes spatial smearing of the signal across the scalp. As result, EEG has two major weaknesses. The first is low signal-to-noise ratio (SNR) of the EEG signal. The EEG signal is of very low amplitude (range of μV) and is often contaminated by signals which are generated by muscles during eye movements, jaw movement and other muscle activity. The second weakness of EEG imaging is its low spatial resolution, caused by the volume conduction of the brain and by the limited coverage of the brain by EEG electrodes. Nevertheless, the EEG has superb temporal resolution and allows to study responses to rapidly changing stimuli such as sound.

Even when at rest, the brain shows ongoing oscillatory activity. Historically, this rhythmic spontaneous activity was divided into the following frequency bands: Delta (0-4Hz), Theta (4-8Hz), Alpha (8-12Hz), Beta (12-30Hz) and Gamma (30Hz~70Hz) (van Drongelen, 2006). The strength of these oscillations were found to be modulated by different behavioural states such as different stages of sleep, level of attention and measuring spontaneous activity has a role in clinical applications such as epilepsy treatment. Another class of EEG signals is evoked activity or event-related potentials (ERPs) by internal or

external stimuli and is discussed in the next section.

### 2.3.2. Event-Related analysis

An event-related potential (ERP) is a brain response to a standard auditory, visual or somatosensory stimulus such as beep, flash etc. However, these event-related responses are approximately ten-times smaller than the spontaneous EEG activity. To overcome this a method of signal time-averaging is usually applied to the EEG signal. This method assumes that the evoked responses to repeated stimuli are consistent, the noise is truly random and has zero mean and the noise is not correlated with the stimulus (van Drongelen, 2006). Figure 2-7 shows the principle of the averaging method: First, we present many repetitions of the same stimulus and average all the EEG responses, which include the "true" response signal with superimposed noise (spontaneous EEG activity). Then we average the EEG responses to repeated stimuli together. As the random noise component is not time-locked to stimulus, it averages out while the evoked response signal stays intact. If we have enough trial repetitions, this approach increases the SNR and allows us to obtain a good estimate of the neural evoked response signal.

However, despite the effectiveness of the averaging method in increasing SNR, it has two major shortcomings: (1) it requires for the stimuli to be discrete, and (2) it requires many repetitions of the stimuli. An alternative method of extracting evoked responses that overcomes these limitations is described below in section 2.3

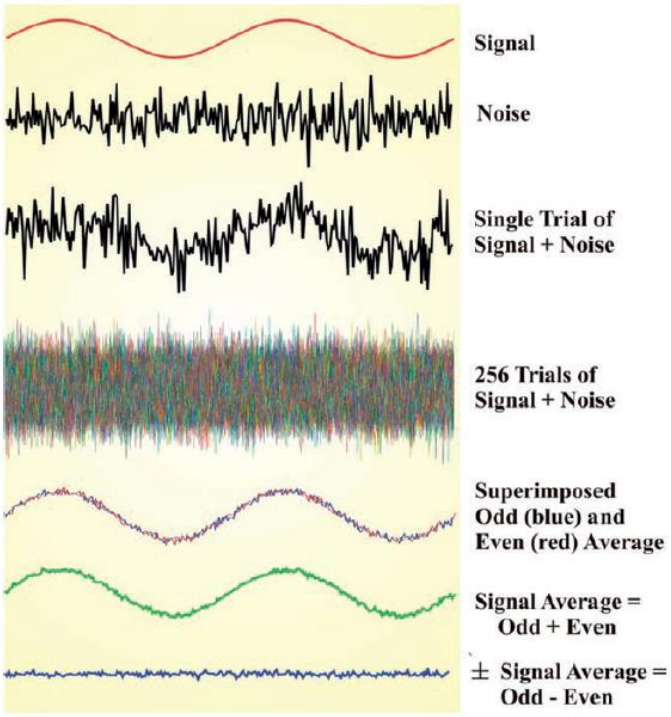



*Figure 2-7. Extracting evoked responses using time-averaging. The first and second trace show the signal and noise separately. How the signal look with superimposed noise is shown for one trial (trace 3) and multiple trials (trace 4). The recovered signal by averaging trials is shown for odd (blue, trace 5), even (red, trace 5) and all trials (green, trace 6). The last trace shows the residual noise, which was calculated by subtracting averaged signal of odd and even trials. Adapted from van Drongelen (2006).*

Typical ERPs to auditory stimuli are shown in figure 2-8. Based on their latency, they can be categorized into three groups: (1) fast responses of the auditory brainstem response (ABR), (2) middle-latency responses from initial activation of the auditory cortex and (3) slow responses from the activation of primary and higher auditory cortices (Picton, 2013). Each evoked response consists of several prominent components (deflections). The components of ABRs are named in order using roman numerals. The middle and slow responses are labelled using P or N, which indicates their polarity i.e. whether they have positive (P) or negative (N) amplitude, and a letter or number which specifies their order. Alternatively, the polarity letter is followed by a number that specifies the latency of the component in milliseconds e.g. N100. The most commonly identified components are described below:

P1

The P1 component, also referred to as the P50, is the smallest ERP component of the late evoked response, appearing around 50 ms post-stimulus. It is thought to originate in primary auditory cortex in Heschl's gyrus and possibly also has generators in the hippocampus, planum temporale, lateral temporal cortex and neocortical areas. The auditory processing

associated with P1 appears to be pre-attentive orienting to new sound stimuli. The P1 also changes during auditory system development and is also modulated by a number of disorders such as posttraumatic stress disorder, autism, Alzheimer's disease and many others (Luck & Kappenman, 2011).

N1

The N1 component is a negative ERP component, peaking between 80 ms and 120 ms after an acoustic change in the environment. The N1 peak is composed of three overlapping, independent subcomponents (Naatanen & Picton, 1987): a fronto-central subcomponent with a negative peak at 100 ms and generators in a wide area around A1 on Heschl's gyrus, a T subcomponent with positive and negative peaks at 100 ms and 150 ms respectively, both originating in association areas in the superior temporal gyrus; and a vertex-negative subcomponent peaking at around 100 ms associated with attention to stimuli (Luck & Kappenman, 2011).

P2

The P2 component, which usually occurs in pair with N1, is a positive deflection with latency in the range 150–250 ms. As the N1 response, P2 is also sensitive to physical parameters of the stimuli, however the exact function is unclear. The source of auditory P2 is thought to be auditory cortices, planum temporale and Brodmann's area 22 (Crowley et al., 2004; Godey et al., 2001).

MMN

The mismatch negativity (MMN) response is seen as a negative displacement in the difference wave obtained by subtracting the ERP to frequent stimuli (standard), from that to less-frequent stimuli (deviant). Although the MMN is not an ERP component in its strict definition, it is frequently used in the literature.

The concept of mismatch negativity was introduced and recently reviewed by (Naatanen *et al.*, 2007). The MMN is an automatically-generated brain response to any change in auditory stimulation exceeding a certain limit roughly corresponding to the behavioural discrimination threshold. The change could be in various aspects of the stimuli, for example, pitch or sound direction. Moreover, the MMN is also elicited by higher-level abstract changes such as grammar violations in mother-tongue sentences.

The MMN response is seen as a negative displacement in the difference wave obtained by subtracting. The scalp-recorded MMN has a fronto-centrally predominant scalp distribution and is mainly explained by the sum of the activity generated in the supratemporal cortices.

The MMN usually peaks at 150–250 ms from change onset. In contrast to the N1 or P2, the MMN is mainly the outcome of a discrimination process where the deviant stimuli are found to mismatch the memory representation of the preceding stimuli in the auditory cortex. Importantly, the MMN is thought to be elicited irrespective of the subject or patient's direction of attention.

How some of these auditory evoked components are modulated by the location and other spatial features of sound is discussed in the section 2.4.2.

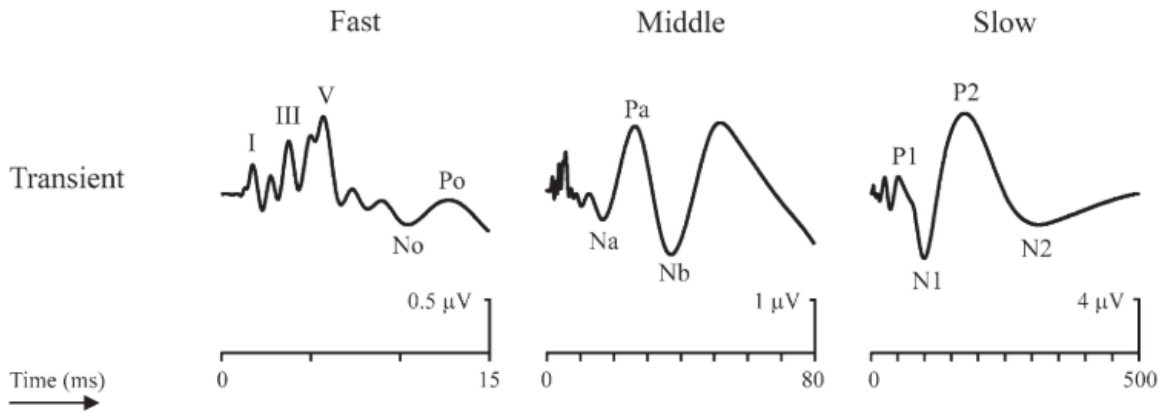



*Figure 2-8. Auditory evoked potentials. Typical evoked responses to a click train showed on three time-scales. Left: fast responses of the auditory brainstem response, Center: Middle-latency responses caused by initial activation of the auditory cortex, Right: slow responses of the auditory and association cortices. Adapted from Picton (2013).*

### 2.3.3. Model-based TRF analysis

As mentioned in the previous section, the time-averaging method for estimating the neural evoked response requires for the stimuli to be discrete and needs many stimulus repetitions. Recently, a model-based approach called temporal-response function (TRF) modelling has been proposed that allows to estimate the neural response using continuous, unrepeated stimuli (Lalor *et al.*, 2006; Lalor *et al.*, 2009; Crosse *et al.*, 2016).

The idea behind this method is that the relationship between the stimulus (input) and corresponding neural response measured by EEG (output) can be described by a linear time-invariant (LTI) system. Certainly, the brain responses violate the assumptions of the LTI, which requires for the stimulus-response relationship to be linear and consistent in time. However, by approximating the brain to LTI, it allows us to employ continuous and unrepeated stimuli in experimental paradigms as well as to model the brain-response relationship by an impulse response (or as referred in this work as TRF). An impulse response is a function that fully characterizes an LTI and is calculated e.g. using deconvolution by measuring the output of the system to a known input signal.

Importantly, when the impulse response of a system is known, it allows one to predict the output of the system to a given input signal (forward mapping). Similarly, the impulse response can also be used to reconstruct the input signal to the system when we know the output (backward mapping). In the context of this work, an example of forward mapping can be predicting the EEG signal from the known sound source azimuth time-series. An example of backward mapping is when we reconstruct the sound azimuth from the recorded EEG signal.

### 2.3.4. Forward TRF mapping

The aim of forward mapping is to predict a neural response (EEG) to a known stimulus feature time-series such as sound azimuth or sound envelope amplitude. See diagram in figure 2-10. A model, also called encoder, $w(\tau, n)$ represents the linear mapping from the stimulus feature time-series $s(t)$ to the multi-channel EEG response $\hat{r}(t, n)$.

$$\hat{r}(t,n) = \sum_{\tau} w(\tau,n)s(t-\tau,n) \qquad \textit{Eq 2-1}$$

where $t = 1 \ldots T$ is the time sample, $\tau = \tau_{min} \ldots \tau_{max}$ is the relative time lag in samples, and $n = 1..N$ is the number of EEG channels. The model weights $w$ can be solved by ridge regression using the equation (Tikhonov and Arsenin, 1977):

$$w = (S^T S + \lambda I)^{-1} S^T r \qquad \textit{Eq 2-2}$$

where $\lambda$ is the ridge regression parameter, $I$ is the identity matrix and the matrix S is the lagged time series of the stimulus feature:

$$S = \begin{bmatrix} s(1-\tau_{min}) & s(-\tau_{min}) & \cdots & s(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & s(1) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & s(1) \\ s(T) & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & s(T) & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & s(T) & s(T-1) & \cdots & s(T-\tau_{max}) \end{bmatrix} \qquad \textit{Eq 2-3}$$

The model weights $w$ are two-dimensional as the model is a function of time (lags) and space (EEG channels) and can be interpreted in a similar way as an ERP. The time lags represent a time-shift between the EEG and stimulus time-series. Negative time-lag means that the stimulus is lagging behind the EEG, while positive time-lag means that that the stimulus is

preceding before the EEG. Essentially, the TRF model can be viewed as an impulse response and the process of predicting the neural response from the stimulus time series is analogous to convolution. Figure 2-9 illustrates the structure of the TRF model.

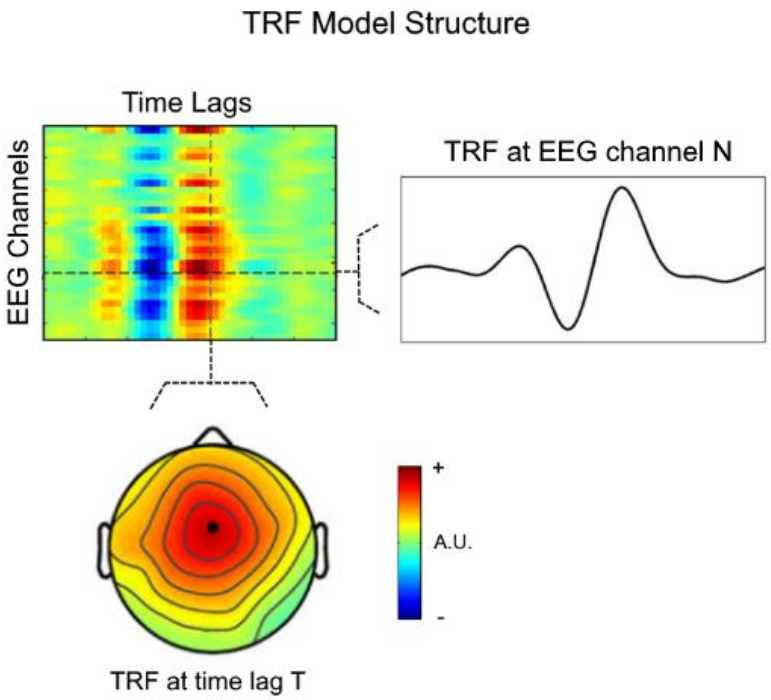



*Figure 2-9. TRF model weights. The TRF model is a function of EEG electrode channels (spatial dimension) and a time-lag between EEG and stimulus (temporal dimension).*

The TRF model is trained and validated on different data in order to prevent overfitting. In practice, this is usually implemented using leave-one-out cross-validation approach. The data is first divided into a number of parts (folds) and the model is trained on all folds except one, which is left out. Then, the model is evaluated on the left-out fold. The model evaluation is performed by correlating the predicted and the actual neural response using Pearson's r. This process is repeated for all folds and then the resultant correlation values are averaged together. See figure 2-10 for the diagram of the TRF fitting process.

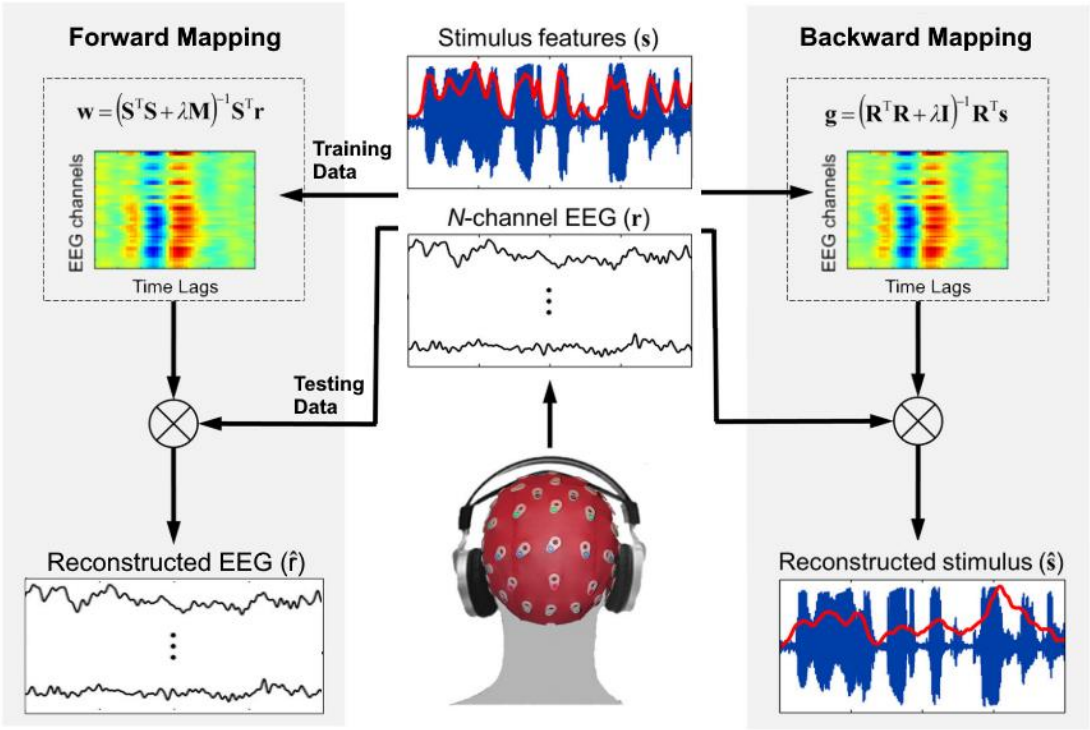



*Figure 2-10. Schematic of forward and backward TRF mapping. The aim of forward mapping is to find an encoding model that allows to predict EEG data from known stimulus time-series. Backward mapping is trying to reconstruct the stimulus time-series from the known EEG data. To prevent model-overfitting, a cross-validation approach is used. The data is split into a training data set, which is used to calculate the model weights, and a validation set that is used for evaluating the model performance.*

### 2.3.5. Backward TRF mapping

In backward mapping, we are interested in reconstructing a stimulus feature time-series from a neural response EEG. See diagram in figure 2-10. The process is mathematically analogical to the forward mapping that is described above. A model, also called decoder, $g(\tau, n)$ represents the linear mapping from the continuous multi-channel EEG response $r(t, n)$ back to the stimulus time-series $\hat{s}(t)$.

$$\hat{s}(t) = \sum_{n=1}^{N} \sum_{\tau} g(\tau, n) r(t + \tau, n) \qquad \textit{Eq 2-4}$$

where $t = 1 \ldots T$ is the time sample, $\tau = \tau_{min} \ldots \tau_{max}$ is the relative time lag in samples, and $n = 1..N$ is the number of EEG channels.

As in the forward mapping, the model $g$ can be solved by performing ridge regression:

$$g = (R^T R + \lambda I)^{-1} R^T s \qquad \textit{Eq 2-5}$$

where $\lambda$ is the ridge regression parameter, $I$ is the identity matrix and the matrix R is the

lagged time series of EEG data:

$$R = \begin{bmatrix} r(1-\tau_{min},1) & r(-\tau_{min},1) & \cdots & r(1,1) & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & r(1,1) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & r(1,1) \\ r(T,1) & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & r(T,1) & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & r(T,1) & r(T-1,1) & \cdots & r(T-\tau_{max},1) \end{bmatrix} \quad Eq\ 2\text{-}6$$

Importantly, there are two features of the backward mapping that are worth mentioning: (1) The method is "multi-variate" in a sense that the stimulus is reconstructed based on information from all available EEG channels across number of time lags. Therefore, it has the potential to be more sensitive than traditional methods of EEG analysis e.g. time-averaging. (2) The backward models, unlike their forward mapping counterparts, are not easily interpretable in a neurophysiological sense due to its sensitivity to noise i.e. the spatial distribution of the decoder weights does not necessary reflect the underlying EEG activity. However, it is possible to transform a backward model weights into forward mapping "activation patterns", which are more physiologically interpretable (Haufe *et al.*, 2014):

$$A = \Sigma_r g \Sigma_{\hat{s}}^{-1} \qquad Eq\ 2\text{-}7$$

where $A$ is the activation pattern, $\Sigma_r$ and $\Sigma_{\hat{s}}$ are covariance matrices corresponding to neural response $r$ and reconstructed stimulus $\hat{s}$, and $g$ are the decoder model weights.

## 2.4. Neural Representation of Auditory Space

In this section we describe the main parts of the central auditory system that are involved in sound localization. First, we briefly discuss the subcortical regions, where many interaural interactions occur and where spatial acoustic cues are first processed. Then we describe which cortical regions are involved in spatial hearing, and lastly we describe current models of the neural representation of acoustic space. Most of this section will focus on the more often studied topic of encoding of azimuthal (horizontal) location and the section 2.4.3 will briefly discuss the neural encoding of sound elevation (vertical position). Sections 2.4.4 and 2.4.5 discuss spatial hearing in complex environments with multiple sources and the effect of attention. Finally, the last two sections describe the neural response to sound motion and the neural sensitivity to sound source velocity and acceleration.

### 2.4.1. Subcortical Processing of Auditory Space

After the incoming sound waves are transduced into neural activity by the peripheral auditory system, they travel to the cochlear nucleus (CN) via the auditory nerve. There the signal diverges into a number of parallel upstream pathways. The schematic diagram of the central auditory pathways is shown in figure 2-11. It appears that the interaural time difference (ITD), interaural level difference (ILD), and monaural spectral cues are processed in separate subcortical streams (Grothe *et al.*, 2010a). There are several subcortical structures that are involved in binaural cue processing: ventral cochlear nucleus (VCN), superior olivary complex (SOC), dorsal nuclei of lateral lemniscus (DNLL) and inferior colliculus (IC) (Stecker & Gallun, 2012).

Initially, the temporal information of the sound is encoded by bushy cells in the VCN, which phase lock to the fine temporal structure or the envelope of the sound. The neural signals then target the nuclei in SOC, medial and lateral superior olives (MSO and LSO), the primary sites of binaural interaction. Here, the neurons were shown to be sensitive both to ITD and ILD acoustic cues (Yin 2002). The next structure along the pathway, DNLL, is thought to play a role in response sharpening and for hearing enhancement in reverberant environments (Pecka *et al.*, 2007). Another major structure that is involved in binaural sound localization is the IC, which receives many inputs from the abovementioned subcortical structures. The responses of IC neurons show similar characteristics to SOC; however, they have some additional properties such as further response sharpening, envelope-ITD sensitivity (Joris, 2003) or reduced frequency dependence of ITD tuning (Fitzpatrick & Kuwada, 2001).

There is evidence that spectral cues are mainly processed in the dorsal part of dorsal cochlear nucleus (DCN). Here, the neurons are responsive to sound location-sensitive notches in the frequency spectrum of the incoming sound. Further, the spectral information is processed by the IC, which is likely involved in response sharpening (Grothe *et al.*, 2010a). More information on subcortical processing can be found in detailed review articles by (Yin, 2002; Grothe *et al.*, 2010a; Stecker & Gallun, 2012).

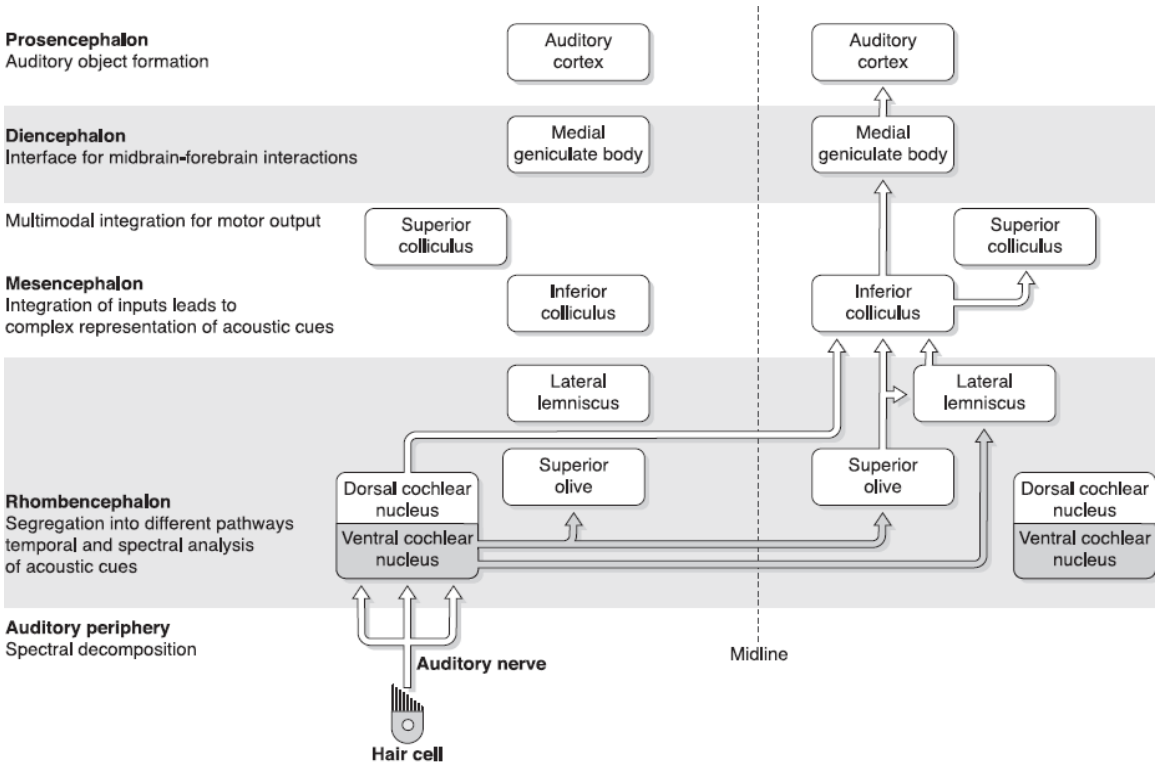



*Figure 2-11. The mammalian ascending auditory pathways involved in sound localisation. Figure adapted from Grothe et al. (2010b).*

### 2.4.2. Cortical Representation of Sound Azimuth

After spatial auditory cues are extracted and pre-processed in subcortical regions, the auditory neural pathways enter the auditory cortex (AC) for further processing. It has been shown that the AC plays an important role in sound localization and lesion studies demonstrated that an intact auditory cortex is necessary for normal spatial hearing (Jenkins & Masterton, 1982; Zatorre & Penhune, 2001; Bizley *et al.*, 2007). Also, an interesting question then arises as to whether the AC encodes sound source position independent of the type of acoustic cue or if the cues are represented separately. As mentioned in the previous section, the ITD and ILD acoustic cues are processed in separate pathways at the brainstem level. However, at the level of cortex it is unclear how these cues are encoded and as concluded by (Higgins *et al.*, 2017), it appears that the ILD and ITD cues are neither fully separated or fully integrated. For further discussion see studies by (Ungan *et al.*, 2001; Altmann *et al.*, 2007; Johnson & Hautus, 2010; Edmonds & Krumbholz, 2014; Salminen *et al.*, 2015; Higgins *et al.*, 2017; Wood *et al.*, 2018).

Which regions of AC are selective specifically to sound location has been discussed. Based on animal studies, it was proposed that sound location is processed in the spatially-selective dorsal "where" pathway, which is separate from the ventral "what" pathway that is involved in sound object recognition (Romanski *et al.*, 1999; Kaas & Hackett, 2000; Rauschecker &

Tian, 2000; Tian *et al.*, 2001). This functional separation of auditory streams is also supported by human imaging studies (Arnott *et al.*, 2004). The "where" stream originates in posterior belt and parabelt regions of AC, targets planum temporale (PT) and posterior superior temporal area (pST). Further, it connects via inferior parietal lobule (IPL) to premotor cortex (PMC), dorsolateral prefrontal cortex (DLPC) and finally reaches to the inferior frontal cortex (IFC). A model diagram of this dual pathway is shown in figure 2-12. For further information on the functional role of the dorsal and ventral auditory pathways see reviews (Ahveninen *et al.*, 2014; Clarke & Geiser, 2015; Rauschecker, 2017).

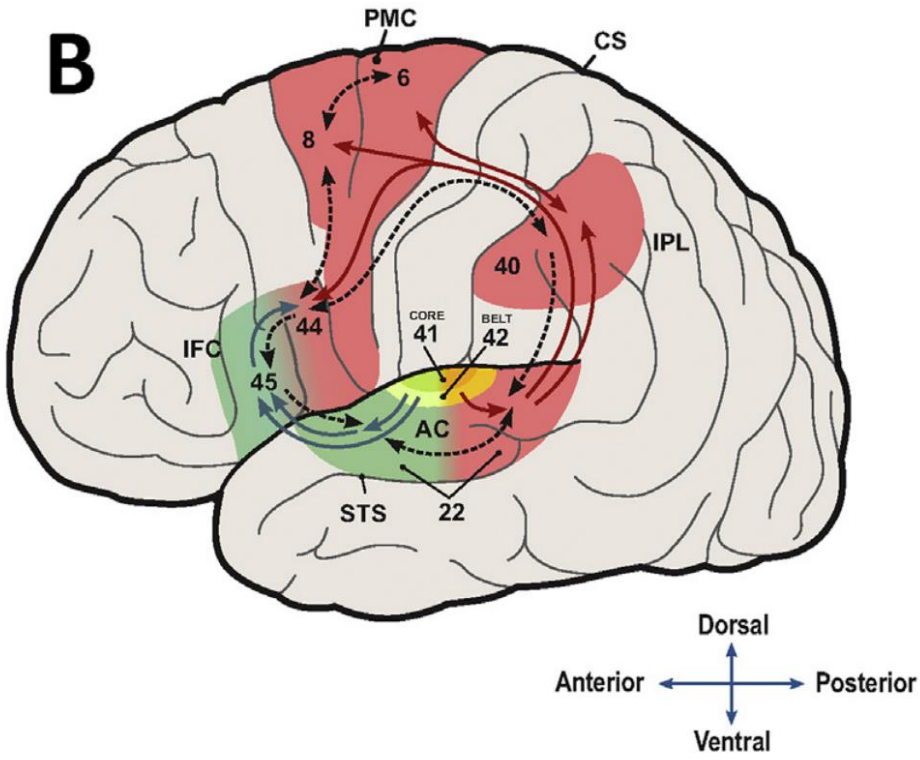



*Figure 2-12. Recent model of the auditory dual-route pathway which separates the neural streams into the "what" pathway (green) that is used for object recognition and the "where" pathway (red) that is involved in sound localization. Marked structures: 41-core auditory cortex (AC), 42-belt areas of AC, 22- dorsal superior temporal sulcus (STS), 40- inferior parietal lobule (IPL), 6-premotor cortex (PMC), 8-dorsolateral prefrontal cortex, 44, 45- inferior frontal cortex (IFC). Adapted from Rauschecker (2017).*

Although we know which regions are likely involved in sound localization, it is unclear how auditory space is encoded in these structures. In contrast to human visual or somatosensory systems, there is no evidence that auditory space is encoded using a local code in a form of cortical topographic map. Instead, animal studies suggested that the neurons in AC are broadly tuned to spatial location and their receptive field mostly covers the contralateral

hemifield of the auditory space (Middlebrooks & Pettigrew, 1981; Stecker *et al.*, 2003; Harrington *et al.*, 2008; Middlebrooks & Bremen, 2013). However, a minority of spatially sensitive AC neurons respond maximally to ipsilateral hemifield or are tuned to the midline. See an example of firing activity of single neuron as function of sound azimuth in figure 2-13.

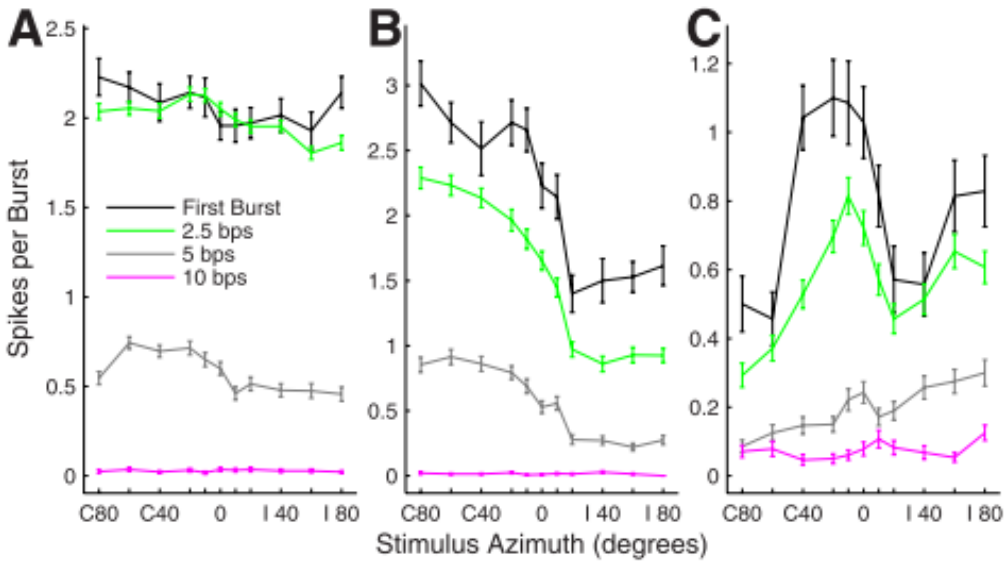



*Figure 2-13. Neuronal spiking activity measured in the primary auditory cortex of a cat expressed as function of sound azimuth. On x-axis, the values on the left (labelled "C") indicates contralateral locations with respect to the listener, while values on the right of the centre (labelled "I") show ipsilateral angles. Different colours show different presentation rates of the stimulus noise bursts. Adapted from Middlebrooks and Bremen (2013)*

Compatible results with animal intracortical recordings have been also obtained using non-invasive imaging in humans. The lateralized evoked responses have been obtained to spatialized static sound sources as shown by MEG (Palomaki *et al.*, 2000; Palomaki *et al.*, 2005) as well as EEG (Magezi & Krumbholz, 2010; Briley *et al.*, 2013) studies. From these studies, it appears that it is the N1 and P2 components of the evoked response, which peak around 100 and 200 ms post-stimulus respectively, that are modulated by horizontal sound location. See figures 2-14 and 2-15. The spatial sensitivity of human cortex to spatial sound have also been investigated using fMRI (Derey *et al.*, 2016), which also showed mostly broad contralateral responses within the AC to sound. See figure 2-16. Interestingly, this study also showed that PT, unlike other spatially sensitive AC regions, is invariant to sound level variations.

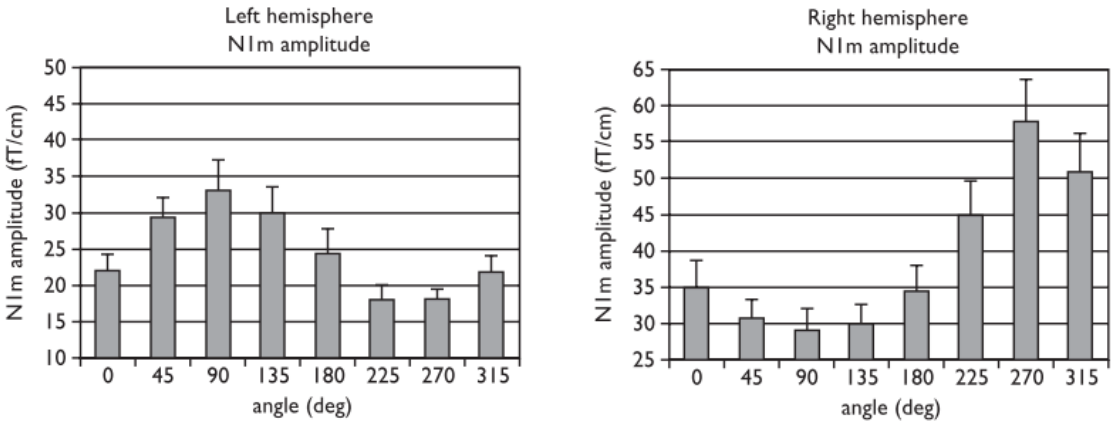



*Figure 2-14. The amplitudes of MEG N1m evoked component shown as function of sound azimuth for the left and right hemispheres. A clockwise convention was used to indicate the sound angles, where 0, 90, 180, 270 deg correspond to front, right, rear and left locations relatively to the listener. Adapted from Palomaki et al. (2000)*

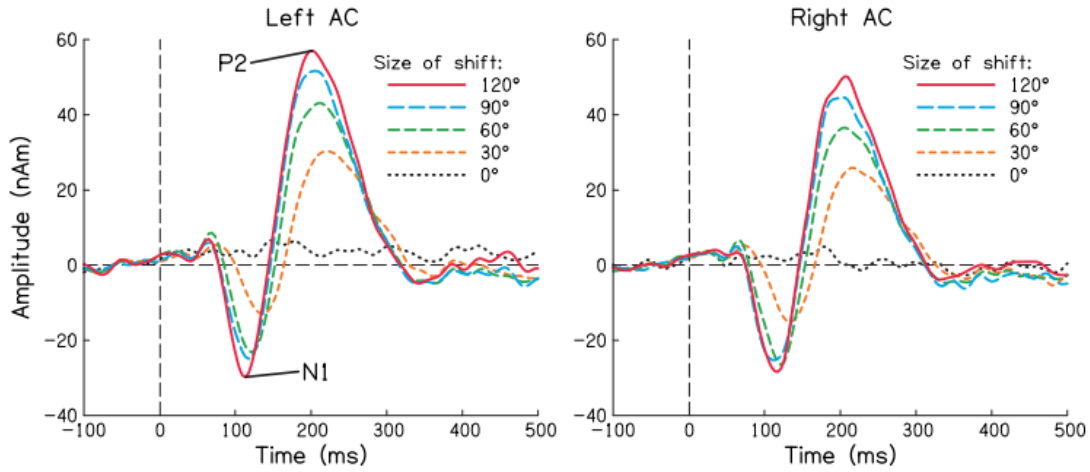



*Figure 2-15. EEG evoked responses of the left and right auditory cortices to sound location shifts. The sound stimuli were presented in sequence and originated in one of the five azimuthal locations -60 (left), -30, 0, +30, +60 (right). The data shows that response magnitude of the N1 and P2 components increase with the size of the location shift. Adapted from Briley et al. (2013)*

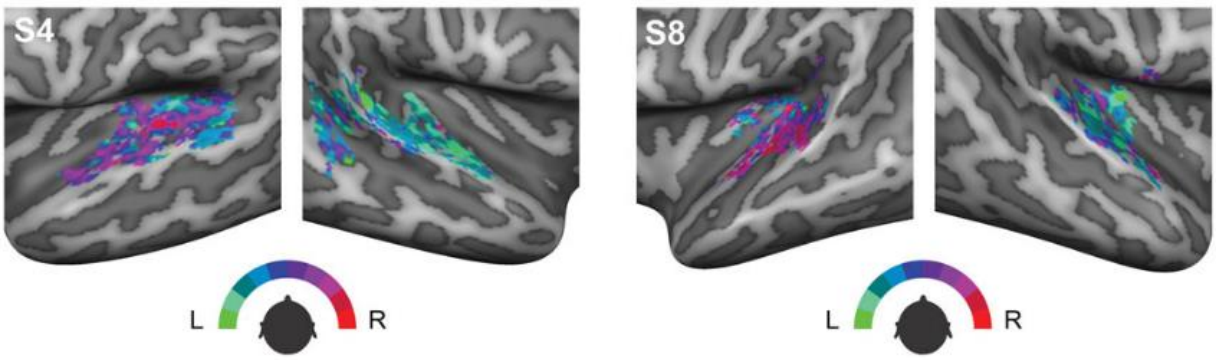



*Figure 2-16. fMRI activation map of the cortical surface with color-coded preference to different sound source positions in the horizontal plane shown for two subjects. The stimuli were presented from different spatial locations and the maps was obtained using phase-*

*encoding paradigm. Adapted from Derey et al. (2016)*

Based on neurophysiological observations, several cortical encoding models of auditory space have been proposed. The currently popular "opponent-channel" model (Stecker *et al.*, 2005) is a population encoding scheme which states that the azimuthal sound position may be computed by calculating the difference between contra- and ipsilaterally tuned neural populations within each hemisphere. This model is supported by a number of neuroimaging studies (Magezi & Krumbholz, 2010; Młynarski, 2015; Derey *et al.*, 2016; Ortiz-Rios *et al.*, 2017). Others hypothesized that the sound position might be encoded using more than two neuronal channels. The existence of a third channel was proposed by (Dingle *et al.*, 2010; Dingle *et al.*, 2012; Briley *et al.*, 2016) and even encoding by a larger number of relatively narrowly tuned channels was proposed (Carlile *et al.*, 2016). Another alternative is "labelled-line" representation, where sound location is represented by activity patterns of individual neurons. As discussed in a ferret study by (Wood *et al.*, 2018), although the A1 neurons within each hemisphere respond generally to the contralateral hemifield, the receptive fields of neurons in a single hemisphere were narrower than predicted by opponent channel theory and had best azimuth distributed across the opposite hemifield. The authors proposed that each hemifield of auditory space is encoded by contralateral A1 using labelled-line system.

### 2.4.3. Cortical Representation of Sound Elevation

Finally, as most studies investigating cortical encoding of sound location focus only on horizontal plane localization (azimuth), not much is known on how the brain encodes elevation. Based on the somewhat limited evidence, it appears that the cortical regions that process sound azimuth and elevation overlap (Pavani *et al.*, 2002; Trapeau & Schönwiesner, 2018). Nevertheless, in our recent study (Bednar *et al.*, 2017) as well as in a study by Fujiki *et al.* (2002), it has been shown that the monaural cues are processed later than binaural cues, which suggests their separate processing. With respect to elevation encoding, a recent fMRI study (Trapeau & Schönwiesner, 2018) showed that location sensitive AC regions decrease in activation with increasing elevation, however, the observed tuning was quite broad. Also, the study showed that the cortex does not encode the acoustical elevation, but rather perceived sound elevation.

### 2.4.4. Encoding Location Multiple Sound Sources

When sound waves from two concurrent sounds arrive simultaneously, the interaction between the sound waves might corrupt the binaural spatial cues, which become less

informative (Blauert, 1997). However, psychophysical experiments have shown that humans can localize more than one sound source, although the localization is worse than in the single-source case (Yost & Brown, 2013). Also, this study proposed that the ability to localize multiple sources might be explained by brief temporal windows "glimpses", where the acoustic power of one of the sources dominates over the other source. This way, assuming there is enough of these glimpses, one could accumulate enough spatial information about several sources. With respect to our ability to resolve a pair of two concurrent spatial sources, it has been shown that for broadband noise stimuli that were only distinguishable based on their location, subjects were able to detect spatial separation of sources when the stimuli were horizontally separated only about 20 degrees. However, larger separation was needed for more lateral locations (Best *et al.*, 2004a).

The spatial sensitivity of the central auditory system to multiple sources has been investigated by animal neurophysiological studies. In an experiment that involved two spatial sounds, Middlebrooks and Bremen (2013) found that neurons in cat's primary AC preferentially synchronized to one of the two stimuli and in effect segregated the spatial streams. A rabbit study that investigated firing patterns of IC neurons in response to spatial auditory stimuli showed that the pattern activity decoder can correctly distinguish between two simultaneously presented stimuli and a single sound source (Day & Delgutte, 2013). In line with that, another study showed that the neural activity measured in monkey IC was alternating between patterns corresponding to each of the spatialized sounds to which monkeys were simultaneously presented (Caruso *et al.*, 2018).

Human spatial studies that employed multiple sources focused mainly on "stream segregation". The process of stream segregation describes the ability to disentangle competing sounds that come from multiple sources (Bregman, 1994). It has been shown that several different sound characteristics play a role in this, namely, temporal envelope, fundamental frequency, phase spectrum as well as spatial acoustic cues contribute to the segregation process, see review by Moore and Gockel (2002). The spatial stream segregation in humans has been shown for both binaural cues ITD and ILD as well as for elevation, although the successful elevation-based segregation depended on the stimulus characteristics (Middlebrooks & Onsan, 2012; Carl & Gutschalk, 2013). Interestingly, it has been suggested that the spatial stream segregation is based on differences between spatial acoustic cues corresponding to concurrent sound sources but not on perceived difference in locations of the target and masker sources (Middlebrooks & Onsan, 2012). Consequently, the authors propose that different neural pathways might be involved in spatial stream segregation and

location discrimination. Somewhat in line with this are the results of a recent fMRI study (Shiell *et al.*, 2018), who showed that human cortex encodes spatial separation between the sources rather than their absolute locations.

### 2.4.5. Spatial Attention and Cocktail Party Scenario

In the previous section, we have discussed the human ability to perceive and localize multiple sound sources. However, in the real-world, we typically pay attention to only one of the objects within the scene at a time. How attentional selection among multiple sources influences the auditory neural processing has long been investigated mostly in the context of "cocktail party", which describes a situation when one is attending a speaker in the presence of other competing speakers (Cherry, 1953). In such situations, it has been shown that the low-frequency neural activity dynamically tracks predominantly the envelope of the attended speech stream and to a lesser extent the unattended stream (Ding & Simon, 2012a; Mesgarani & Chang, 2012; O'Sullivan *et al.*, 2015).

Several studies employed a cocktail-party experiment with multiple competing sources that required listeners to perform a spatial task, and identified that PT and posterior STG (Zündorf *et al.*, 2013; Zündorf *et al.*, 2014; Lewald & Getzmann, 2015) regions are involved in selective spatial attention. An EEG study (Gamble & Luck, 2011) used a paradigm, where subjects indicated whether one of the two simultaneously presented sounds was a target stimulus. The study identified a prominent ERP component reflecting allocation of selective attention named N2ac (N2-anterior-contralateral) with latency around 350 ms post-stimulus, which was calculated as the difference between responses to left- and right- spatialized targets. The N2ac component was as well observed in later studies that involved similar experimental paradigms (Lewald & Getzmann, 2015; Lewald *et al.*, 2016). In addition, the latter study reported a longer latency component, LPCpc, which was also suggested to reflect attention allocation. See figure 2-17 that shows the morphology of the N2ac and LPCpc components.

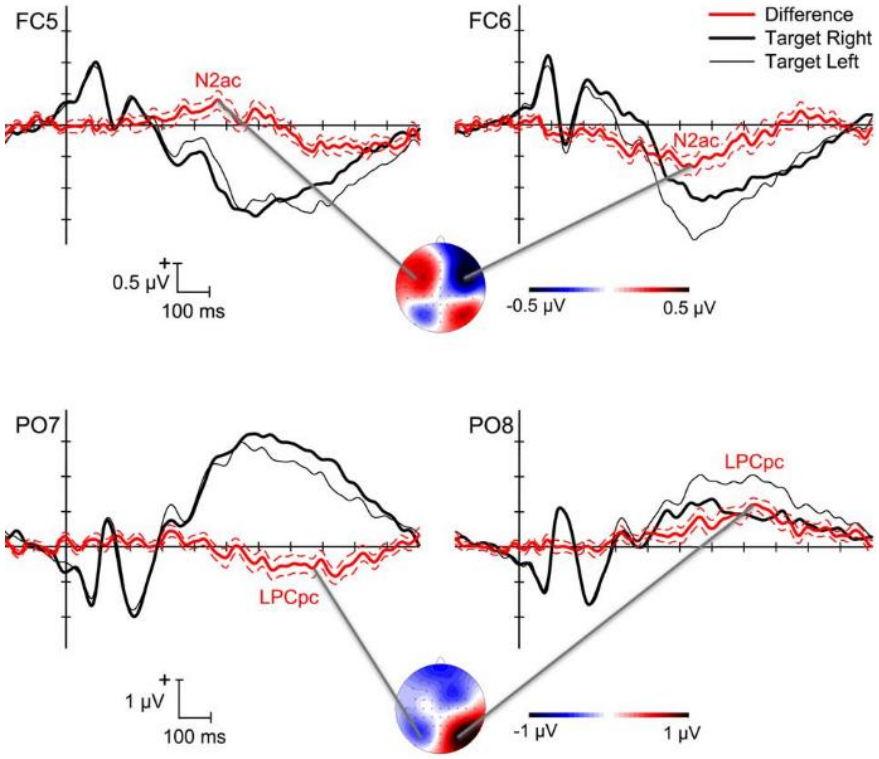



*Figure 2-17. N2ac and LPCpc subcomponents of the difference evoked response that was obtained by subtracting responses to spatial target stimuli (left minus right) in multi-speaker condition. From Lewald et al. (2016)*

Along with the N2ac and LPCpc components, it has been also found that the power of alpha EEG (8-12 Hz) reflects the locus of deployed attention. Specifically, it has been shown that the posterior alpha power increases ipsilaterally and decreases contralaterally to the location of the attended stimuli. This effect has been demonstrated first for visual (e.g. Worden *et al.*, 2000; Kelly *et al.*, 2009) as well as for spatial auditory stimuli (Kerlin *et al.*, 2010; Wöstmann *et al.*, 2016; Wostmann *et al.*, 2018). See an example of auditory alpha lateralization in figure 2-18.

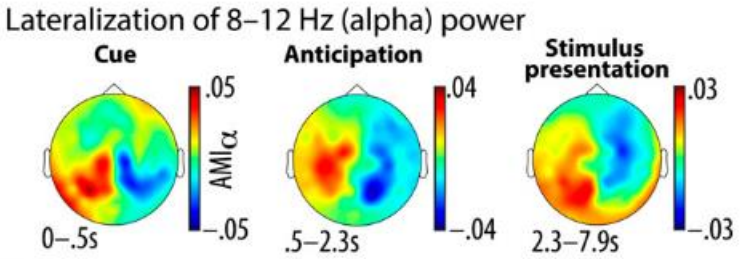



*Figure 2-18. Topographic map of the attentional modulation of the alpha power (8-12Hz) during an auditory selective attention task. The attentional modulation was calculated as the normalized difference in alpha power in responses to lateralized attended stimuli (left minus right). From Wöstmann et al. (2016)*

### 2.4.6. Auditory Motion Sensitivity

In the visual system, there are specialized low-level motion detectors that are direction- and velocity- sensitive (Borst & Egelhaaf, 1989). In the auditory domain, the existence of such low-level detectors has not been confirmed. Instead, it might be that we perform a higher-level motion detection based on already extracted sound source location data. However, how this higher-level motion perception might work is under debate (Carlile & Leung, 2016).

Based on psychophysical experiments, (Grantham, 1986; 1997) proposed a "snapshot hypothesis", which states that the subject can infer the direction or extent of motion by spatial comparison of snapshots taken at the endpoints of moving trajectories. Therefore, the auditory system does not need to have motion detectors per se but can simply rely on the same neural structures that are used for localization of non-moving sources. However, this theory in its original form does not explain results from psychophysical studies that found sensitivity to the source position between the trajectory endpoints (Perrott & Marlborough, 1989), and sensitivity to sound acceleration (Perrott *et al.*, 1993). These latter findings instead support the existence of spatialized motion detectors within the auditory system.

Many neurophysiological studies found that neurons in the auditory system are sensitive to moving stimuli. At the subcortical level, it was found that neurons in IC are sensitive to stimulus dynamics i.e., their activity depends on prior stimulation (Spitzer & Semple, 1991; Spitzer & Semple, 1998). However, McAlpine *et al.* (2000) argued that this effect could be explained by adaptation and does not necessary imply motion sensitivity in the form we observe in the visual system. Human imaging studies investigated neural responses to moving sound stimuli and found that posterior temporal areas, planum temporale and IPL are involved in motion processing (Griffiths *et al.*, 1998; Baumgart *et al.*, 1999; Griffiths & Green, 1999; Griffiths *et al.*, 2000; Pavani *et al.*, 2002; Warren *et al.*, 2002; Krumbholz *et al.*, 2005a; Poirier *et al.*, 2005; Krumbholz *et al.*, 2007; Alink *et al.*, 2012; Poirier *et al.*, 2017). In addition, the majority of studies found the neural responses to moving stimuli to be different (stronger) in comparison to responses to "corresponding" static stimuli. This could be interpreted in such a way that supports the possibility of exclusive motion sensitive structures within the auditory system. However, comparisons between moving and static sounds are questionable from the methodological point of view, since static stimuli are always somewhat impoverished relative to moving stimuli. See (Smith *et al.*, 2004; Smith *et al.*, 2007) for evidence against the existence of spatialized motion detectors.

The existence of the auditory motion aftereffect (Grantham & Wightman, 1979; Grantham,

1989; Dong *et al.,* 2000), which is analogous to the "waterfall effect" in the visual domain, might be considered as evidence in favour of specialized motion sensitivity. This phenomenon occurs after adaptation to repeated moving adaptor stimuli, which cause subsequent static probe stimuli to be perceived as moving in the opposite direction to the adaptor. Interestingly, Magezi *et al.* (2013) showed that the EEG correlates of auditory motion aftereffect are not direction selective i.e. the adaptor motion direction did not matter.

To summarize, humans are sensitive to moving sound and its properties like direction, velocity and acceleration. However, the exact neural mechanisms underlying motion sensitivity are unclear and although several models of motion perception have been proposed, there is no consensus on this matter. More information on auditory motion processing can be found in reviews by (Warren *et al.*, 2002; Neuhoff, 2004; Ahveninen *et al.*, 2014; Middlebrooks, 2015; Carlile & Leung, 2016).

### 2.4.7. Sound Velocity Sensitivity

There are several psychophysical experiments showing that we are able to resolve the velocity of moving sound (Chandler & Grantham, 1992; Carlile & Best, 2002; Kaczmarek, 2005), and that we can also discriminate between accelerating and decelerating sounds (Perrott *et al.*, 1993). However, physiological evidence for motion and velocity sensitivity of neurons is lacking and it is often argued that the velocity is estimated based on location cues (Locke *et al.*, 2016)

A handful of neurophysiological studies investigating evoked responses to motion onset found that latency and amplitude of the motion related component of the response is dependent on the stimulus velocity (Makela & McEvoy, 1996; Xiang *et al.*, 2005; Getzmann, 2009). Specifically, the response had a shorter latency and larger amplitude with increasing velocity. This is also in line with a psychophysical study that employed a reaction time task and found that listeners responded faster to higher velocity sounds (Getzmann, 2008). In addition, Getzmann (2009) found that responses exhibited larger contralaterality and right hemispheric bias for the higher-velocity stimuli and suggested that a different mechanism might be involved in fast- and slow- motion processing.

Compatible with these results, an fMRI study by (Meng *et al.*, 2016) compared neural responses to slow and fast-moving sounds and found that the higher velocity sounds cause stronger activation of PT and increased activation of premotor areas. The authors hypothesised that premotor areas might be engaged in preparation of head movements, which are typically performed for higher velocity sounds. To the best of our knowledge,

neural sensitivity to sound acceleration has not been investigated.

## 2.5. Summary

In this chapter, we have first introduced acoustic cues that are used for sound localization, namely binaural interaural time- and level- differences (ITD and ILD) as well as monaural spectral cues. Next, we provided an introduction on neuroimaging techniques with emphasis on electroencephalography (EEG), which is a technique we employed in this work. We have then described the basic principles of traditional event-related (ERP) analysis of EEG data and discussed details of more advanced model-based EEG prediction (forward mapping) and stimulus reconstruction (backward mapping) techniques that are used throughout this work.

In the following section, we have reviewed the neural representation of auditory space. First, we have briefly introduced spatial processing at the subcortical level where the spatial acoustic cues are extracted and processed in functionally separate pathways. Next, we discussed how auditory space is represented at the cortex. Specifically, we have identified cortical regions of the dorsal auditory stream that are sensitive to sound source location. We then provided an overview of the current literature on cortical encoding of sound location, which indicates that the sound position is represented by some form of population neural coding rather than topographically as is often seen in other sensory systems. However, the exact details on how auditory space is cortically encoded are unknown.

In the next sections, we discussed how we localize sound in more naturalistic scenarios with multiple overlapping sources. We describe how spatial acoustic cues help to separate overlapping sources from the mixture of sounds. And we also discuss how spatial attention help us to select the most relevant source and what are the neural correlates of selective attention.

Finally, we reviewed literature on auditory motion processing, which suggests that we likely do not have low-level motion detectors but rather analyse motion using the same neural structures that are used for sound localization. Nevertheless, it appears that the cortex is modulated by motion as some cortical regions respond differently to static and moving sound and a small number of studies have reported cortical sensitivity to motion velocity.

In the following chapter we present the results from our first study which investigated the cortical tracking of the azimuthal time-series of a sound applying the methods described earlier in this chapter.

# Chapter 3. Decoding the Trajectory of Continuously Moving Sound Source

## 3.1. Introduction

In chapter 2.4, we have discussed that auditory space is encoded by broadly tuned neurons within each hemisphere of auditory cortex, which are activated primarily by sounds coming from the contralateral hemifield. This was shown in mammals (Middlebrooks & Pettigrew, 1981; Stecker *et al.*, 2003; Werner-Reiss & Groh, 2008; Ortiz-Rios *et al.*, 2017), and in humans (Palomaki *et al.*, 2000; Krumbholz *et al.*, 2005b; Palomaki *et al.*, 2005; McLaughlin *et al.*, 2015; Derey *et al.*, 2016).

At a more fundamental level, there is uncertainty as to whether this contralateral neural tuning applies both to ILD and ITD cues (Johnson & Hautus, 2010; McLaughlin *et al.*, 2015), and whether these cues are processed independently at the level of cortex (Ungan *et al.*, 2001; Altmann *et al.*, 2007; Johnson & Hautus, 2010; Edmonds & Krumbholz, 2014; Higgins *et al.*, 2017).

It is unclear how this distributed opponent encoding might reflect the dynamics of sound motion. EEG studies have shown that the brain responds to a sudden stimulus motion onset (MOR) (Krumbholz *et al.*, 2007; Getzmann & Lewald, 2010; Getzmann, 2011); or to an onset of mismatching moving stimuli (MMN) (Altman *et al.*, 2010; Shestopalova *et al.*, 2012). Nevertheless, as these studies employed only transient and discretized stimuli with abrupt sound energy and motion onsets, it remains unknown whether cortex tracks spatial changes in more natural auditory situations, where sound stimulus and its movement are continuous. In order to address this issue, we have designed a study that consist of two experiments:

In Experiment 1, we aimed to use a model-based decoding method, which we introduced in section 2.3.5, to attempt to learn whether the cortex reflects continuous, time-varying location of a moving sound source. This builds up on previous spatial studies that used short and often static discrete sound stimuli to investigate sound location encoding in humans. To test this, we presented subjects with HRTF- spatialized noise stimuli that was perceived to be randomly moving in the horizontal plane while we recorded their EEG. Next, we attempted to build a reconstruction model, that would predict the sound trajectory from the EEG data.

Experiment 2 was aimed to test whether the sound encoding relies on sound amplitude envelope fluctuations and to address a question whether we encode ITD and ILD using the

same cortical structures. In this experiment, we employed the same stimulus-reconstruction approach but used spatially impoverished ITD- or ILD- only pulse stimuli, which were again simulating random sound source motion in the horizontal plane.

Previously, we have shown that we can successfully predict positions of static noise bursts using a linear classifier (Bednar *et al.*, 2017). However, to the best of our knowledge this is the first study to attempt to "decode" EEG signals to determine the time-varying location of a moving auditory input. Such decoding approaches have been used previously for stationary auditory inputs, based on their intensity (Lalor *et al.*, 2009; Mesgarani *et al.*, 2009a) and to assess how attention is deployed in complex environments (Ding & Simon, 2012a; O'Sullivan *et al.*, 2015).

The results of this study were published as a research article: Bednar, A., & Lalor, E. C. (2018). "Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG." *NeuroImage*, 181, 683-691.

## 3.2. Material and Methods

***Participants.*** In total thirty-three participants (median = 21 years; min = 18 years; max = 27 years; 22 females; 26 right handed) participated in this study with informed consent. Sixteen subjects took part in the first experiment and seventeen in the second. All subjects reported no neurological diseases and normal hearing. The experiments were conducted in accordance with the Declaration of Helsinki and were approved by the Research Subjects Review Board of the University of Rochester.

***Experimental procedure.*** Participants listened to auditory stimuli (described below) presented via headphones while performing a simple target detection task. The task required subjects to respond with a button press to infrequent tremolo targets (modulation frequency 4 Hz, 2 s long), which were embedded in the stimuli. The number of targets within each trial ranged from 1 to 6 per trial. During the experiment, the subjects sat in a dark soundproof room. To minimize movement, the participants were asked to look at a fixation cross displayed on a computer screen directly in front of them. In both experiments, subjects undertook a total of 24 trials. Each trial lasted 3 minutes and consisted of one continuous sound stimulus that was perceived as randomly moving within the horizontal plane. The stimuli were different for the first and the second experiment.

***Stimuli.*** In both experiments, the auditory stimulus was perceived as moving within the frontal part of the horizontal plane. The trajectory was pseudo-random, simulating smooth but unpredictable sound movement. The stimulus and method of sound spatialization was different for the first and the second experiment.

In the first experiment, the sound stimuli consisted of continuous pink noise with frequency roll-off 10 dB/decade. The sound source motion was implemented by virtual acoustic space (VAS) using Oculus Audio SDK, which simulates head-related transfer function (HRTF) filtering. The sound was spatialized so that it was perceived to be pseudo randomly moving on a semi-circular trajectory in the horizontal plane between -90˚ (left) and +90˚ (right) relative to the subject (see figure 3-1A). The simulated motion of the source had an average angular velocity of 80˚/second. Each trial (sound stimulus) lasted exactly 180 s.

In the second experiment, we used trains of short pink noise bursts. The noise bursts were presented at a rate of 100 Hz, each burst lasted 1 ms and the silence interval between the bursts was 9 ms (see figure 3-1B). The stimuli were spatialized using either interaural level differences (ILDs) or interaural time differences (ITD). Similar to the first experiment, the acoustic cues were manipulated so that the sound appeared to be moving smoothly in a pseudo-random manner between the left and the right ear. The ITD cues varied between +/-

750 µs and ILD cues within the range of +/- 20 dB. From the total of 24 trials, 12 trials were spatialized using only ILDs and 12 trials contained only ITDs. The ILD and ITD trials were alternating. As in the first experiment, all trials were 180 s long.

***EEG data preprocessing.*** The EEG data were recorded using a 128-channel ActiveTwo acquisition system (BioSemi, The Netherlands) at a sampling rate of 512 Hz. Preprocessing was done in MATLAB using custom written scripts and the EEGLAB toolbox (Delorme & Makeig, 2004). The data were filtered between 0.02 and 30 Hz and downsampled to 64 Hz. Bad channels were interpolated from the surrounding channels using the spline function from EEGLAB. Finally, the data were re-referenced to the average of all electrodes.

***Data analysis.*** In order to assess if and how the location of the moving stimuli was represented in the EEG, we used a multivariate linear reconstruction model $g$ to reconstruct the sound stimulus trajectory $S$ from the neural data $R$ (see section 2.3.5 for more details). This mapping can be described as:

$$\hat{S}(t) = \sum_{n=1}^{N} \sum_{\tau} g(\tau, n) R(t + \tau, n), \qquad \textit{Eq 3-1}$$

where $\hat{S}(t)$ is the reconstructed estimate of the stimulus position at time $t$=1…T, $g(\tau, n)$ is the decoder model which is a function of the time lag $\tau$ and the electrode channel $n$=1…N, and $R(t, n)$ is the neural response at time $t$ and electrode $n$.

The stimulus trajectory, $\hat{S}(t)$, is a measure of the azimuth, where azimuth of +90˚ corresponds to the right and -90˚ corresponds to the left (see fig. 1A). The decoder $g(\tau, n)$ integrates EEG over time lags $\tau$ from 0 ms to 250 ms poststimulus to reconstruct each sample of the stimulus trajectory. This range of lags was selected as it typically encompasses major cortical ERP components to an auditory stimulus. As indicated by the equation, the decoder $g(\tau, n)$ is essentially a multivariate impulse response function calculated from all 128 EEG electrodes and all time-lags simultaneously and the stimulus trajectory is estimated by convolving this impulse response with the EEG data.

Leave-one-out cross-validation was used as a performance measure of our stimulus reconstruction. We fitted the model (decoder) on all but one trial and then evaluated this decoder on the remaining left out trial. Afterwards, we measured the similarity between the reconstructed and the original sound trajectory using Pearson's r. We repeated this for each trial. These correlation coefficients were then averaged across trials for each subject.

To measure the within-subject statistical significance of our decoding, we used a non-parametric permutation approach (Combrisson & Jerbi, 2015). First, we established a null

distribution of the predictions by running the decoding process 1,000 times with randomly permuted stimulus trajectories. After, we used the tail of this empirical distribution to calculate the p-value for the original classification. This was done for trial-averaged data and for each subject separately.

All comparisons on a group level were conducted using two-sided Wilcoxon signed-rank tests.

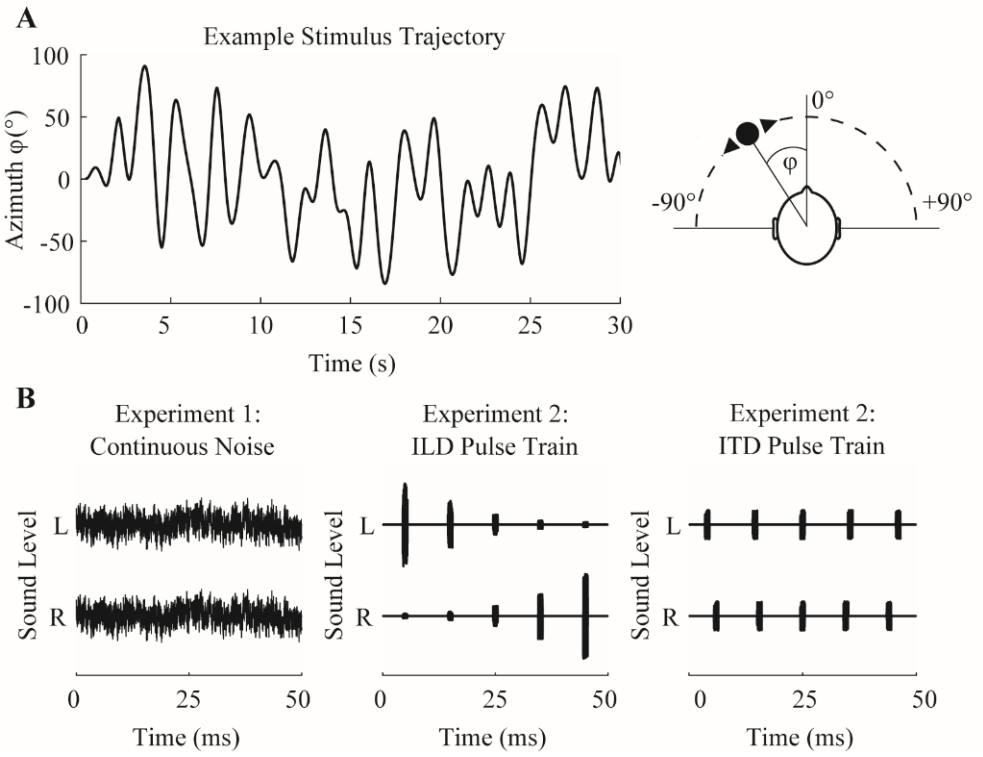



*Figure 3-1. Continuously moving stimuli. In Experiment 1, the pink noise stimulus was HRTF-filtered to be perceived as randomly moving on a semi-circular trajectory around the listener between the left and right. In Experiment 2 the noise stimuli were replaced by ILD- and ITD-only spatialized pulse trains. (A) Example of the sound trajectory. The sound source simulated random and smooth motion between the left (-90 deg) and the right (+90 deg) ear. (B) Sound stimulus waveform examples. Left: HRTF-filtered pink noise was used in Experiment 1. Center: Experiment 2- ILD condition used a pulse train with embedded level differences. Right: Experiment 2- ITD condition used a pulse train with introduced time differences between pulses.*

## 3.3. Results

### 3.3.1. Experiment 1: Reconstructing azimuth of moving sound

We investigated whether the scalp recorded EEG reflects the dynamics of a moving sound source in the horizontal plane. As we have done in the past for other sound features e.g. sound intensity (O'Sullivan *et al.*, 2015), we used the decoding approach to reconstruct the time-varying azimuth of the moving sound. In the first experiment, the subjects listened to

spatialized continuous pink noise. We used VAS to simulate a moving sound source and so listeners had both binaural cues available, as well as spectral cues.

***Predicting stimulus trajectory from broadband EEG.*** Initially, we tested the feasibility of reconstructing the sound trajectory using broadband EEG (0.02-30 Hz). As input to the decoder, we used all 128 EEG channels and lags within the range of 0 and 250 ms.

We used Pearson's r to compare the reconstructed sound trajectories with the original ones ('decoding'). The mean reconstruction correlation averaged across all subjects and trials was *r*=0.076. The mean chance-level correlation ('control') was r=0.008 and was computed by averaging the permutation test correlations across all trials and repetitions (figure 3-2A). On a group-level, the decoding correlation values were significantly larger than the control condition (p=4.4e-4). See examples of reconstructed trajectories in figure 3-2C. Although the reconstruction correlation values are relatively low, it is important to point out that they represent stimulus trajectory predictions from novel unaveraged EEG and should be interpreted in this context given that background EEG can be an order of magnitude larger than stimulus evoked activity.

We also ran the analysis on a single-subject level. For all subjects, the correlations averaged across all trials were higher than the control condition, and for eleven out of sixteen subjects, this difference was significant (see figure 3-2B).

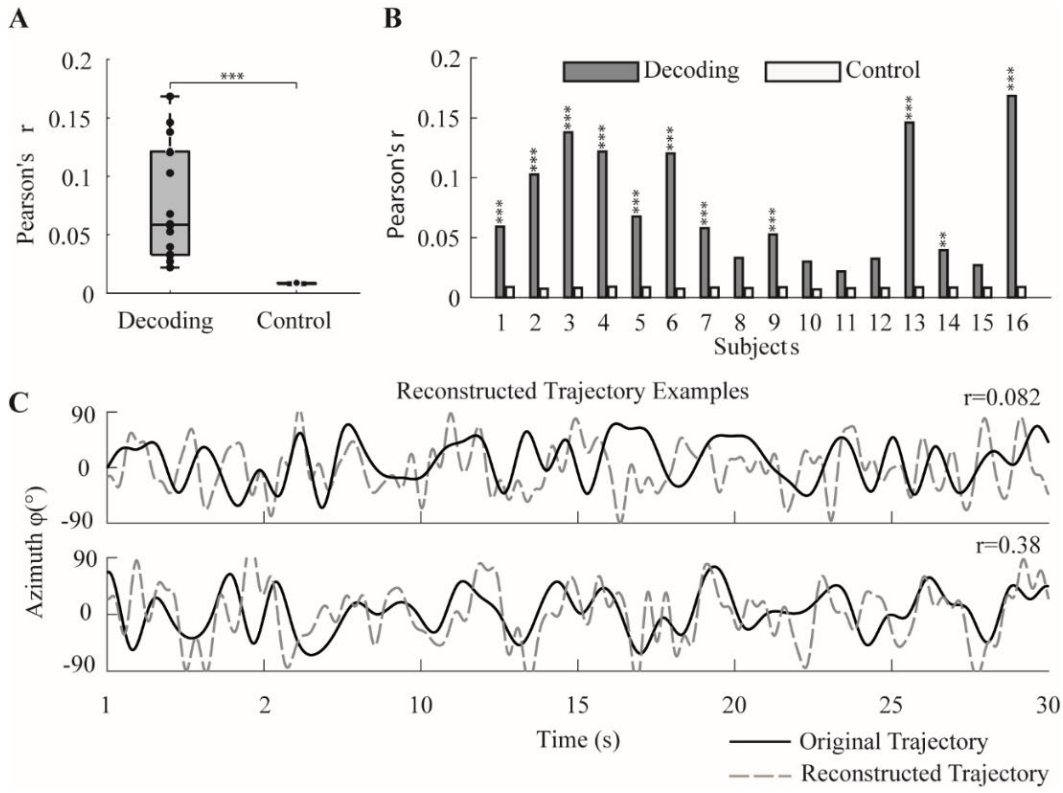



*Figure 3-2. Azimuth reconstruction from broadband EEG. (A) Decoding results shown as*

*an average across subjects and trials. The decoding performance is represented as Pearson's r between the reconstructed and the original sound trajectory. As a control condition, to estimate whether our decoding is above the chance-level, we trained the decoders with randomly permuted trajectories between the trials. We found there is a significant difference between the decoding and the control condition ($p<0.001$). (B) Decoding results shown for each individual subject. The black bars show the reconstruction correlation values averaged across all trials. The white bars show the correlation values corresponding to the permutation test ('control'). (C) Examples of original sound stimulus trajectories along with their reconstructions shown for two different trial segments of one subject. The correlation values between the original and reconstructed trajectories are shown above the plots.*

***Decoding at different EEG bands***. The next stage of our analysis focused on investigating what EEG frequencies are best for stimulus angle reconstruction. We extracted different EEG bands using a sliding bandpass filter from 0.02 to 30 Hz which had a passband window with a width of 2 Hz. We then ran our decoding analysis using each of these EEG bands (figure 3-3A). The results revealed that the decoding is driven by the lowest frequency EEG delta band (0.02-2 Hz). We also ran the analysis after calculating the analytic envelope of the EEG signal using the Hilbert transformation following the bandpass filtering. In this case, the greatest decoding accuracy was achieved within the alpha band of the EEG (i.e., the EEG power in the range 8-12 Hz). In general, the performance decreased with increasing frequency, except for a small improvement around 18 Hz. In the following analyses, we therefore focused on the delta and alpha power EEG decoders, from now on abbreviated as 'Delta' and 'Alpha Pwr'.

The decoding results for Delta and Alpha Pwr decoders are shown in figure 3-3B. The average reconstruction accuracies for both the Delta and Alpha Pwr were significantly above the chance level r=0.08, p=4.3e-04 and r=0.046, p=4.3e-04 respectively.

The Delta decoder performed better than the Alpha Pwr decoder (p=9.7e-3). For the Delta decoder, eleven subjects had decoding accuracy significantly above the chance level. For the Alpha Pwr decoder, eight out of sixteen subjects showed significant decoding on single subject level. In comparison to broadband EEG decoders, both the Delta decoder and the Alpha Pwr decoder had higher reconstruction accuracies than the broadband EEG decoder (p=4.3e-4 and p=5.2e-3 respectively).

To test whether delta and alpha power EEG carry independent information about the source azimuth, we combined the Delta and Alpha Pwr decoders. This was done by performing the

regression using 256 channels (128 channel delta EEG + 128 channels alpha power EEG). Then we compared its performance with each of the decoders separately. On average, this combined 'Delta+Alpha Pwr' decoder performed better than the individual Delta and Alpha Pwr decoders. However, this improvement was only significant when compared with Alpha Pwr decoder (p=4.3e-4), not with the Delta decoder (p=0.063).

***Spatiotemporal decoder characteristics.*** In order to obtain some insight into the possible cortical areas involved in encoding spatial information, we examined which scalp regions were modulated by sound source position and driving our decoding. To do this we plotted the forward transformations of the decoder weights 'activation patterns', which are more interpretable in terms of the underlying physiology (Haufe *et al.*, 2014). As shown in figure 3-3C, the Delta decoder showed strong activations over temporal scalp bilaterally, indicating the likely involvement of auditory cortex. The activation patterns were opposite between the hemispheres, i.e., an increase in activity in one hemisphere was accompanied by a decrease in the other. The Alpha Pwr decoder indicated more posterior activation. The stimulus caused a relative decrease in parietal activity contralateral to the stimulus position, and a relative increase over ipsilateral scalp.

To further compare the Delta and Alpha Pwr markers of spatial audio encoding, we then evaluated the two performances of the two decoders at individual time lags. The temporal window of lags where the decoding was significantly above chance was relatively long. For the Delta decoder, the stimulus angle reconstruction was best when EEG was lagging the stimulus by 78 ms. For the Alpha Pwr decoder the responses were more delayed and decoding peaked at a lag of 187 ms (see figure 3-4A). The single-lag decoder activation patterns are shown in figure 3-4B.

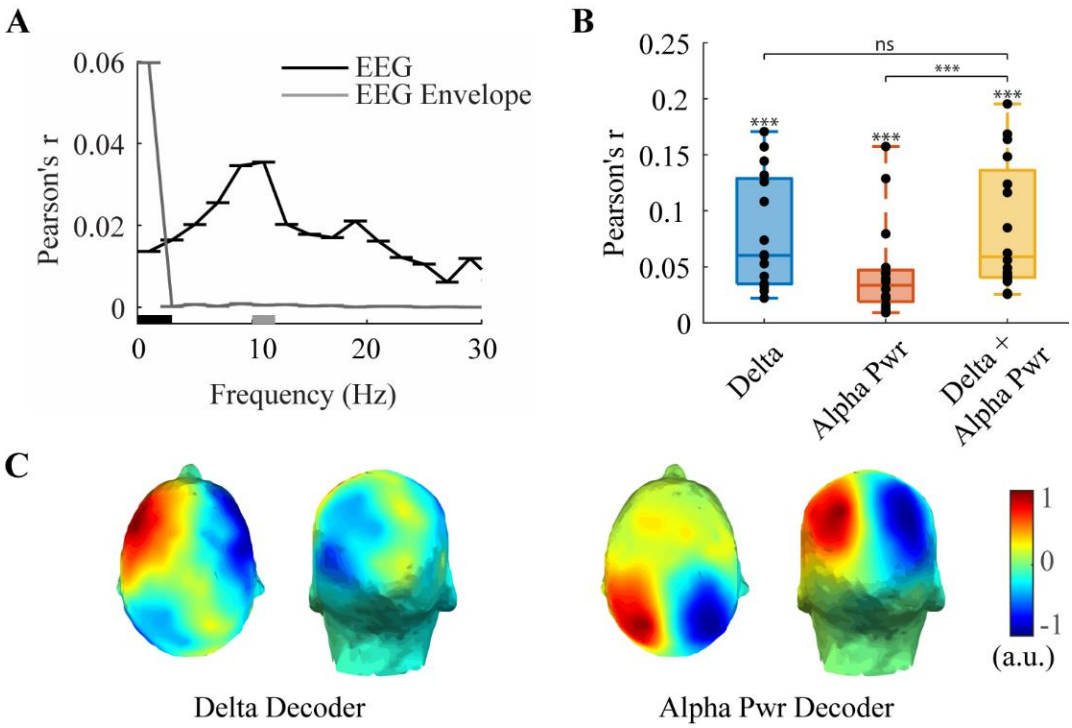



*Figure 3-3. Decoding azimuth for different EEG bands. (A) The decoding performance dependency on frequency bandwidth of raw EEG signal (black) and EEG envelope (grey). Significant frequency windows are marked using thick lines at the bottom of the plot ($p<0.001$). (B) Azimuth reconstruction accuracies shown for Delta (blue), Alpha Pwr (red) and Delta+Alpha Pwr (orange) decoders. All reconstruction accuracies were significantly above the chance level ($p<0.001$). *** indicates reconstruction differences at the level of $p<0.01$. (C) Head plots of decoder activation patterns are shown for Delta and Alpha Pwr decoders.*

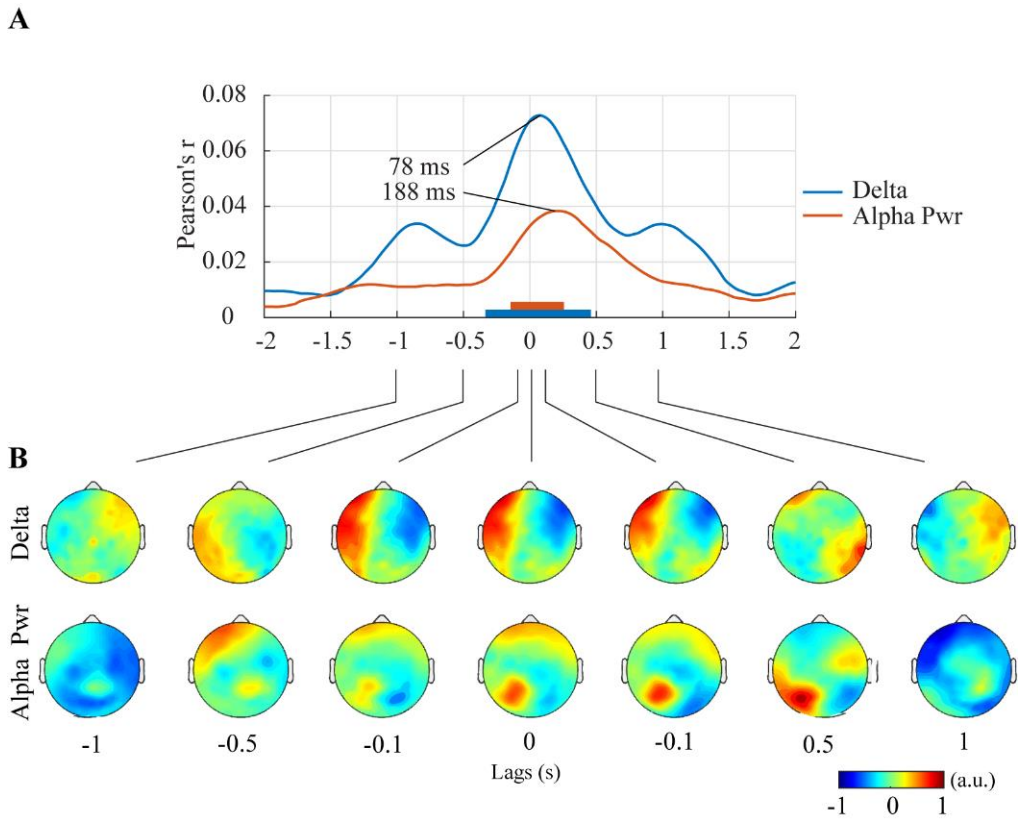



*Figure 3-4. Single lag decoding with decoder activation patterns (A) Single lag decoding shows reconstruction correlations as a function of time lag between stimulus and EEG shown for Delta (blue) and Alpha Pwr (red) decoders. Significant time lags are marked using thick lines at the bottom of the plot (p<0.001). (B) Decoder activation patterns at selected time lags shown for Delta and Alpha Pwr decoders.*

We also investigated if we can reconstruct sound azimuth using EEG channels from a single hemisphere and tested whether this 'simulated lesion' results in a degradation of decoding performance in comparison to using information from the entire scalp. We did this by training our decoders on electrode subsets (n=55) covering either left or right scalp, 'L-hemi' and 'R-hemi', and comparing the results with the decoder that was trained on electrodes covering the entire scalp 'Full Scalp'. To control for the greater number of electrodes in the 'Full Scalp' decoder, this decoder was trained 1000-times on a random subset of 55 electrodes from the entire scalp and the decoding results were averaged together. All midline electrodes (n=18) were omitted.

For both 'L-hemi' and 'R-hemi' decoders using both delta and alpha EEG bands, the azimuth reconstructions were significantly better than chance level (all p<0.005). Also, the results showed that the 'Full Scalp' decoders were significantly better than the decoders trained on a single hemisphere. This was true for 'L-hemi' and 'R-hemi' using delta filtered EEG (p=8e-4 and p=0.016) and for the power of alpha EEG (p=4.5e-3 and p=6.1e-3). See decoding results in figure 3-5.

Also, previous literature has suggested a possible asymmetric representation of auditory spatial information across the hemispheres (Kaiser *et al.*, 2000; Zatorre & Penhune, 2001; Krumbholz *et al.*, 2005b). While our decoder activation patterns (figure 3-3C) didn't show any obvious lateralization, we decided to test this formally by comparing the reconstruction accuracies between the left and the right hemisphere. The accuracies of these different reconstructions were not significantly different for either delta (p=0.26) or the power of alpha EEG (p=0.72). See figure 3-5.

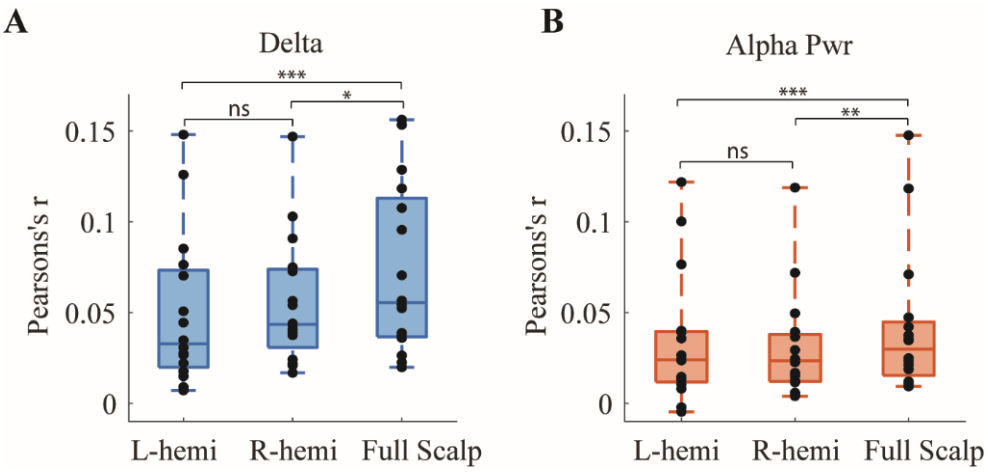



*Figure 3-5. Reconstructing stimulus azimuth using a single hemisphere shown for Delta (A) and Alpha Pwr decoders (B). The decoders were trained using a subset of 55 EEG electrodes corresponding to the left hemisphere 'L-hemi' or the right hemisphere 'R-hemi'. 'Full Scalp' represents a decoder that was trained using 55 randomly selected EEG electrodes from the entire scalp covering both hemispheres (shown as an average of 1000 randomly selected electrode subsets). * ** *** indicates prediction differences at the level of p<0.05, p<0.01 and p<0.005 respectively.*

**Sensitivity to specific position in azimuth.** Finally, we tested whether the EEG tracking of sound position relies on specific position in space or whether it only exhibits sensitivity to stimulus laterality i.e., stimulus on the left or right. We used a forward modelling approach (see Crosse *et al.*, 2016), in which we tried to predict previously unseen EEG data from the stimulus trajectory. Specifically, we used either full trajectory of the moving stimulus 'Full' (continuously varying angle between -90° and +90°, as shown in figure 3-1A) or left/right-only trajectory 'L/R' (rectangular waveform indicating sound position within left or right hemifield).

The EEG prediction accuracies for both Full and impoverished L/R trajectories are shown in figure 3-5A. The comparison of EEG prediction accuracies averaged across all electrode channels show we can predict the EEG signal significantly better using the Full than the L/R trajectory. This was shown for both Delta (p=6.4e-4) and Alpha Pwr EEG (p= 2.3e-3). The scalp distributions of EEG prediction accuracies for Full and L/R trajectories are shown in

figure 3-6B.

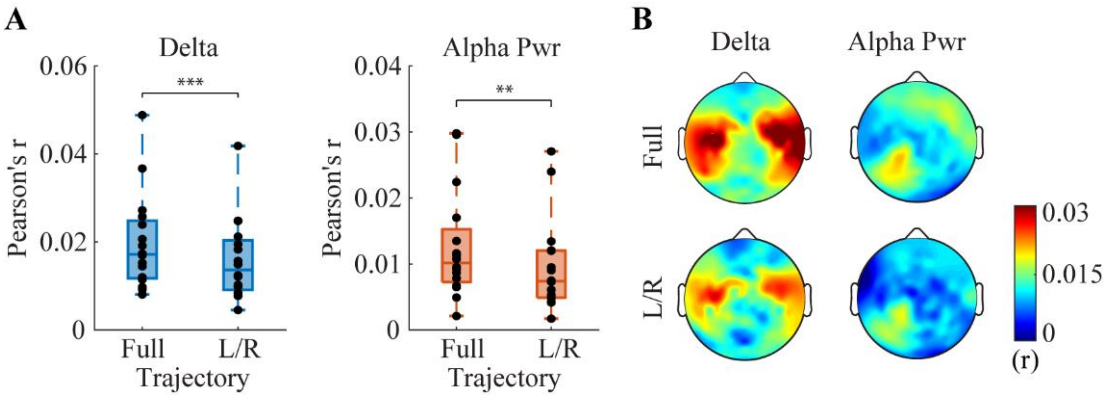



*Figure 3-6. Predicting EEG using full trajectory and left/right-only trajectory. To test whether the EEG tracking of sound position is sensitive to specific position in azimuth rather than just left/right lateralization, we used a forward modelling approach to predict previously unseen EEG from the trajectory data. The sound position was either represented as full trajectory 'Full' (continuously varying angle between -90° and +90°) or left/right-only trajectory 'L/R' (rectangular waveform indicating sound position within left or right hemifield). (A) Comparison of EEG prediction accuracies averaged across all electrodes shown for both trajectories and Delta and Alpha Pwr EEG. ** and *** indicates prediction differences at the level of p<0.01 and p<0.005 respectively. (B) Topographical plots indicating EEG prediction accuracies across scalp shown for both trajectories and Delta and Alpha Pwr EEG.*

***Correlation with behaviour.*** We compared the azimuth reconstruction accuracy with the performance in the tremolo target detection task, which subjects performed during the experiment. On a group level, there was no significant correlation between the trajectory reconstruction accuracy and behaviour performance (r=-0.21, p=0.44) and (r=-0.28, p=0.31) for Delta and Alpha Pwr decoders respectively.

### 3.3.2. Experiment 2: Reconstructing ITD/ILD of moving stimuli

In the second experiment, we examined which features of the acoustic signal were driving our stimulus trajectory reconstruction. We did so by running the decoding on spatially impoverished stimuli that contained either only ITD or only ILD acoustic cues. Similar to the first experiment we simulated smooth movement of the sound in the horizontal plane by manipulating these acoustic cues.

We used the same approach as in the first experiment and we attempted to reconstruct stimulus position (time series of the ILD/ITD stimulus azimuth) from delta band and alpha power of EEG. We found that both ILD and ITD stimuli can be decoded with reconstruction accuracy significantly above the chance level. For Delta decoder, the mean reconstruction correlations of ILD and ITD stimuli were r=0.050, p=9.9e-3 and r=0.045, p=0.017. For

Alpha Pwr decoder, the reconstruction values were r=0.042, p=4.2e-4and r=0.034, p=3.1e-3 for ILD and ITD respectively (see figure 3-7A).

***ILD/ITD decoder activation patterns.*** To characterize the spatial distribution of neural responses to each acoustic cue, we calculated the decoder activation patterns for ILD and ITD decoders (figure 7B). For both ITD and ILD trials, the Delta decoder weights had a lateralised distribution and the activation patterns had an opposite polarity between the hemispheres. Although being noisier, the activation patterns, and especially the patterns corresponding to ILD cue processing, resembled the activation patterns seen in the first experiment. For the Alpha Pwr decoders, the lateralized responses were again more posterior, showing opposite patterns between the left and the right hemispheres. In comparison to the ILD decoder, the ITD weights are less central and located closer to the temporal areas.

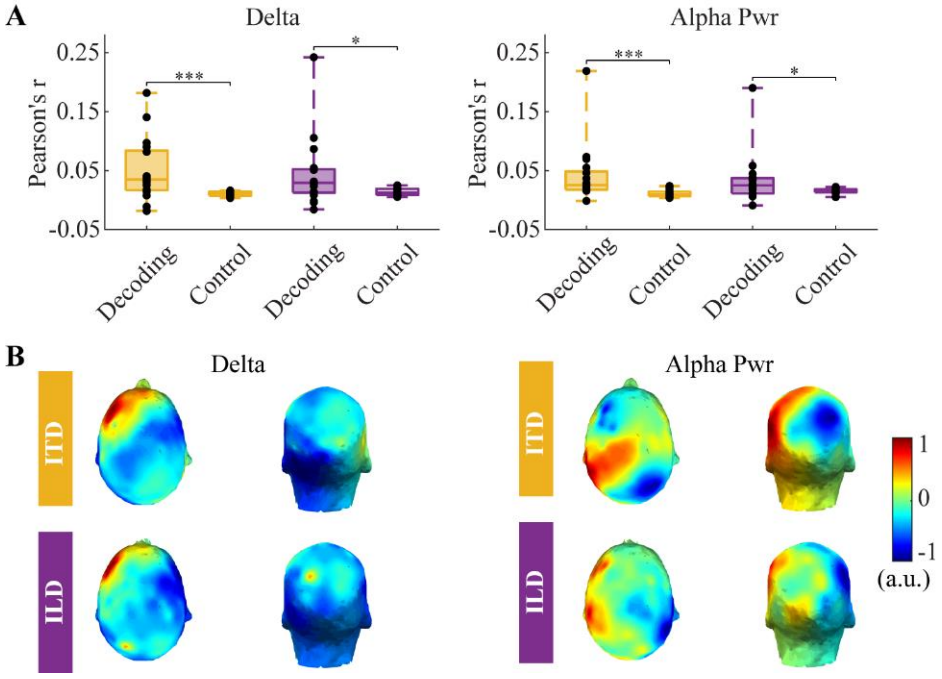



*Figure 3-7. Decoding ITDs and ILDs. (A) Reconstruction accuracy of ILD and ITD trials shown for delta (0.02-2 Hz) and alpha power EEG (8-12 Hz) shown along with chance level correlation values ('control'). * ** *** indicates classification with decoding accuracy significantly above the chance level for p<0.05 p<0.01 and p<0.005 respectively. (B) Head plots of Delta and Alpha decoder activation patterns shown for ILD and ITD trials.*

***Cross-modal decoding.*** We observed that the activation patterns for ILD and ITD decoders were relatively similar. We attempted to quantify this similarity by performing cross-cue classification e.g. evaluating ILD-trained model performance on ITD trials and ITD-trained

model performance on ILD trials.

When evaluating the ILD-trained decoder on ITD data, the reconstruction accuracies were r=0.045, p=0.025 and r=0.032, p=3.6e-3 for Delta and Alpha Pwr. When evaluating ITD-trained decoder on ILD trials, the decoding results were r=0.051, p=1.2e-3 for Delta and r=0.033, p=0.044 for Alpha Pwr. The reconstruction accuracies were significantly above the chance level for all conditions. Pair-wise comparisons between the within-modal (e.g. ILD decoder evaluated on ILD data) and cross-modal reconstruction (e.g. ILD decoder evaluated on ITD trials) are shown in figure 3-8. On a group level, there were no significant differences in reconstruction accuracies between the within-modal and cross-modal decoding.

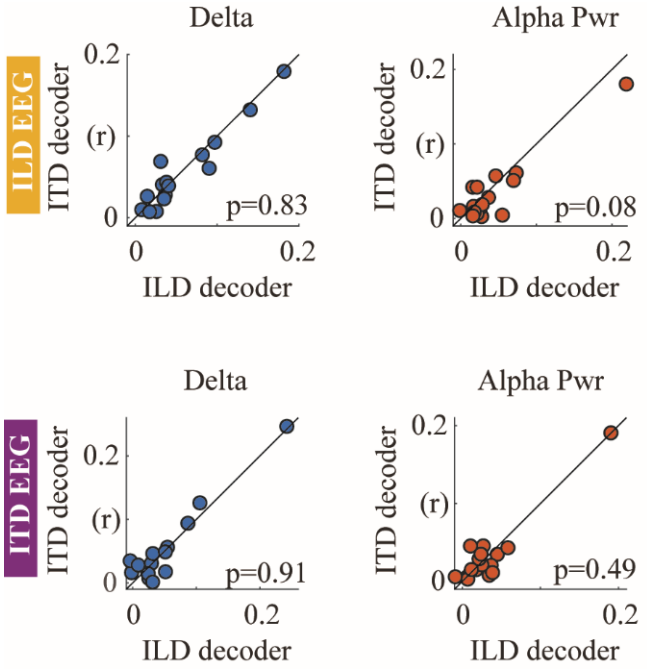



*Figure 3-8. Comparison of within modal and cross-modal classification for Delta and Alpha Pwr decoders. The cross-modal classification was done by evaluating ILD-trained model on ITD trials and ITD-trained model on ILD trials. The p-values correspond to pairwise comparison between the within-modal and cross-modal decoders.*

**Correlation with behaviour.** We compared the azimuth reconstruction accuracy for both ILD and ITD trials with the performance in the tremolo target detection task. For ILD, there was no significant correlation between the trajectory reconstruction accuracy and behaviour performance (r=0.34, p=0.18) and (r=0.21, p=0.41) for Delta and Alpha Pwr decoders respectively. For ITD there was a significant correlation between azimuth reconstruction for Delta (r=0.55, p=0.02) but not for Alpha Pwr (r=0.30, p=0.23) decoders.

## 3.4. Discussion

***EEG decoding reveals cortical tracking of auditory motion.*** Recent studies have shown that cortex is sensitive to sound position, yet it remains unclear to what extent neural activity reflects the on-going dynamics of a moving sound. Here, we demonstrated that it is possible to decode the trajectory of a moving sound source from low frequency delta EEG (0.02-2 Hz) and by the power of alpha EEG (8-12hz). This suggests that cortical activity tracks the time-varying azimuth of a moving sound. We also showed that this tracking does not rely only on the stimulus lateralization i.e., whether sound is on the left or right side of the acoustic space but is rather sensitive to specific sound source location in azimuth.

***Auditory motion is likely encoded in auditory cortex.*** We found that the low-frequency delta EEG (<2 Hz) track the fluctuating stimulus azimuth. This is somewhat in line with previous studies, which demonstrated that low frequency EEG oscillations (<10 Hz) entrain to temporal modulations of sensory input. Specifically, in the auditory domain, this was shown for the speech envelope (Lalor & Foxe, 2010; Luo *et al.*, 2010).

One reason why our decoding performs well in this relatively low and narrow frequency band might be explained by the frequency characteristics of the stimulus. The stimulus consisted of pink noise that was spatialized to be perceived as moving smoothly around the listener. To ensure that the stimulus movement would be comfortable to follow and perceived as continuous, the maximum angular velocity was kept under 280 deg/s. Translating this into the frequency domain, the spectral analysis of the stimulus trajectory time-series showed that the majority of signal power is distributed below 1 Hz (see appendix C). And if, as the results indicate, the cortical activity entrains to the temporally varying sound azimuth in a (quasi-) linear manner, then we would expect that the frequency band of EEG that encodes the stimulus would correspond to the frequency bandwidth of the stimulus trajectory.

The delta EEG decoder weights indicated a contralateral bias. The topographical distribution was relatively symmetrical between the hemispheres but the polarity of weights was opposite between the left and the right hemisphere. Similar to this, predominately contralateral tuning was also shown for lateralized sounds in animals (Stecker *et al.*, 2003; Stecker *et al.*, 2005; Werner-Reiss & Groh, 2008; Ortiz-Rios *et al.*, 2017; Poirier *et al.*, 2017), as well as in humans (Palomaki *et al.*, 2000; Fujiki *et al.*, 2002; Pavani *et al.*, 2002; Palomaki *et al.*, 2005; Getzmann & Lewald, 2010; Lewald & Getzmann, 2011; Derey *et al.*, 2016).

Using single time-lags to decode the trajectory showed that the sound position is best reconstructed when delta filtered EEG is lagging the stimulus trajectory by approximately

100ms. The cortical sensitivity to sound azimuth at this latency was also reported by previous studies (Palomaki *et al.*, 2000; Fujiki *et al.*, 2002; Palomaki *et al.*, 2005; Lewald & Getzmann, 2011; Bednar *et al.*, 2017). This would correspond to the latency of the N100 waveform in classical ERP analysis, which is generated from Heschel's gyrus and planum temporale (Godey *et al.*, 2001).

The second EEG measure that was found to track the sound azimuth was the parieto-occipital alpha power, which is often related to deployment of selective attention (Kerlin *et al.*, 2010; Ahveninen *et al.*, 2013; Wöstmann *et al.*, 2016). Unlike these studies, our experiment did not explicitly employ selective attention as the subjects listened only to a single moving sound stimulus and no competing audio stream was present. However, our results are compatible with recent studies that reported lateralization of occipital alpha oscillations in response to salient but spatially unpredictable sounds (Störmer *et al.*, 2016; Feng *et al.*, 2017).

Different spatio-temporal characteristics of delta and the power of alpha EEG encoding suggests they potentially reflect different aspects of auditory motion processing. We tested this, by combining the Delta and Alpha Pwr decoders and comparing that with the results of the Delta and Alpha Pwr decoders individually. More specifically, if we could show that combining the two decoders led to improved decoding performance, we could argue that the two neural signatures carried complementary information. On average, the combined decoder performed better than the Delta and Alpha decoders separately. However, this result was only significant in comparison to the Alpha Pwr decoder. Therefore, we were unable to confirm if delta and alpha EEG each carry different information about the sound source position.

We then investigated the involvement of each hemisphere in sound location encoding. As shown in a human lesion study (Zatorre & Penhune, 2001), a unilateral lesion caused impairment in sound localization ability. In line with that, we found that the azimuth reconstruction accuracies decreased when we trained the decoders on a subset of EEG channels covering either left or right scalp only. However, the performances of decoders that were trained on a single hemisphere were still significantly higher than the chance level. This suggest that the sound source position is at least partially encoded in a single hemisphere, however, the inter-hemispheric interactions are necessary to fully describe the sound source position. This is also compatible with the opponent-channel model (Stecker *et al.*, 2005), which predicts that sound location is calculated based on a difference of contra- and ipsilaterally tuned neural channels and assumes that both contralateral and ipsilateral

channels are present within each hemisphere.

Interestingly, the comparison of decoding performance between hemispheres did not reveal any significant lateralization, which is in contrast with a number of studies that suggested a larger involvement of right hemisphere in auditory localization (Burke *et al.*, 1994; Kaiser *et al.*, 2000; Zatorre & Penhune, 2001; Palomaki *et al.*, 2005). However, our lack of any inter hemispheric differences may be due to the limited spatial sensitivity of our approach.

***Are we decoding sound position or just fluctuations in the envelope of the sound in each ear?***

It is possible that the decoding in our first experiment was not driven by sound position per se but by the fluctuating envelope of the signal that was caused by ILDs. In the second experiment, we sought to test this by investigating if we can decode sound position independently based on the type of the acoustic localization cue. Analogous to the first experiment, we simulated a smoothly moving sound source in the horizontal plane. However, the stimuli were spatially impoverished and contained either ILD or ITD cues.

The results showed that we can indeed reconstruct the trajectory of a sound source containing only ILD or ITD cues from the EEG with accuracy significantly above the chance level. This indicates that we decoded genuine sound motion trajectory and not only sound envelope changes that correspond to varying ILDs. Interestingly, the performances of ILD and ITD decoders were relatively similar. The correlation values were generally lower than in the first experiment, which could be attributed to several factors. First, the spatially impoverished character of the stimuli possibly caused a reduction in neural responses compared with the stimuli that contained all forms of sound localization cues. This was shown in previous MEG studies (Palomaki *et al.*, 2000; Palomaki *et al.*, 2005). Another possible reason for the reduction in accuracy may be that there were two conditions (ILD and ITD) in the latter experiment and so the decoders were trained on only half the number trials. Indeed, when sub-sampling half the number of trials of Experiment 1 and comparing the decoding performance results with Experiment 2, we found only significant differences for delta EEG and ITD trials ($p=0.048$) but not for ILD trials and there were no differences for alpha filtered EEG (all $p>0.05$).

***ITD and ILD cues may not be represented independently at the level of cortex.*** To determine whether or not ILD and ITD are represented differently in the EEG activity, we performed cross-cue decoding. A similar approach to this was used in an fMRI study by (Higgins *et al.*, 2017). The results presented here show that we can successfully exchange ILD- and ITD- trained models without any significant change in reconstruction accuracy,

despite only approximate ILD-ITD mapping. This suggests that activity at the cortical level represents sound location in an acoustic cue-independent manner. Or, considering the possibility that ILD and ITD are processed by separate cortical networks, the ILD and ITD may activate similar brain regions, whose separation is not detectable by EEG. Whether and where ITD and ILD cues integrate in human cortex is still under debate. And one could interpret the relatively inconsistent electrophysiological evidence seen to date as meaning that although cortical responses to perceptually matched ITD and ILD stimuli show some differences, a cortical region sensitive to sound location independent of the type of acoustic cue possibly exists. See (Ungan *et al.*, 2001; Johnson & Hautus, 2010; Edmonds & Krumbholz, 2014; Salminen *et al.*, 2015; Altmann *et al.*, 2017; Higgins *et al.*, 2017).

***Limitations.*** In this study, we showed that one could decode EEG signals to investigate the neural underpinnings of auditory motion processing. Nevertheless, there are several limitations in the current study.

First, we discuss some limitations arising from the stimulus used. The stimuli were presented over headphones, and the motion was simulated by manipulating the acoustic cues that were embedded in the signal. In the first experiment, VAS was used which is based on non-individualized HRTFs. In the second experiment, subjects were presented with artificially impoverished sound that contained only ILD or ITD binaural cues. Therefore, in both experiments, the fidelity of sound spatialization was imperfect and the neural responses we obtained might not fully represent processing of 'true' auditory motion as it would be when using free-field sound presentation.

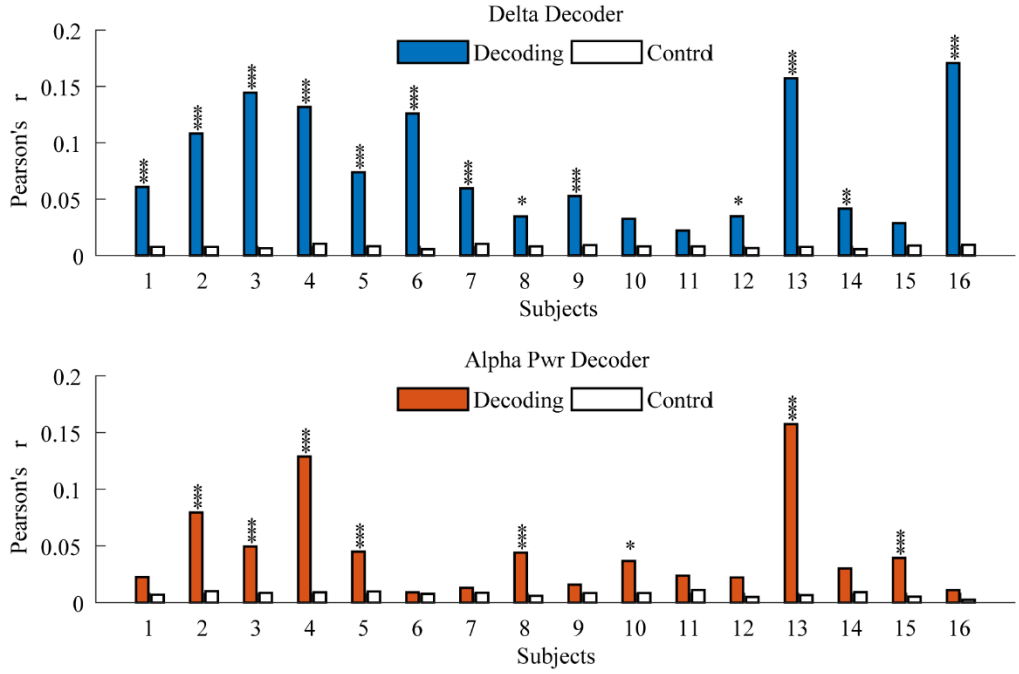
Second, the results of the cross-modal decoding in the second experiment should be interpreted with caution since the ITD and ILD were not perceptually matched. We used only approximate linear mapping between the acoustic cues where the maximum ITDs of +/- 750 µs of corresponded to ILD +/- 20 dB. The absence of evidence for a change in performance when exchanging decoders cannot necessarily be construed as evidence for absence of any difference in how ITD and ILD are represented in cortex. Especially given the inherent spatial limitations and signal-to-noise ratio of EEG.

Finally, the sound trajectory time-series used here could also be improved to better elucidate the underlying neural dynamics of tracking sound motion. In particular, the stimuli trajectories happened to be relatively strongly autocorrelated (see appendix D). This caused 'smearing' of the temporal dimension of our decoding and prevented us from providing more detailed information on the neural dynamics of sound motion processing.

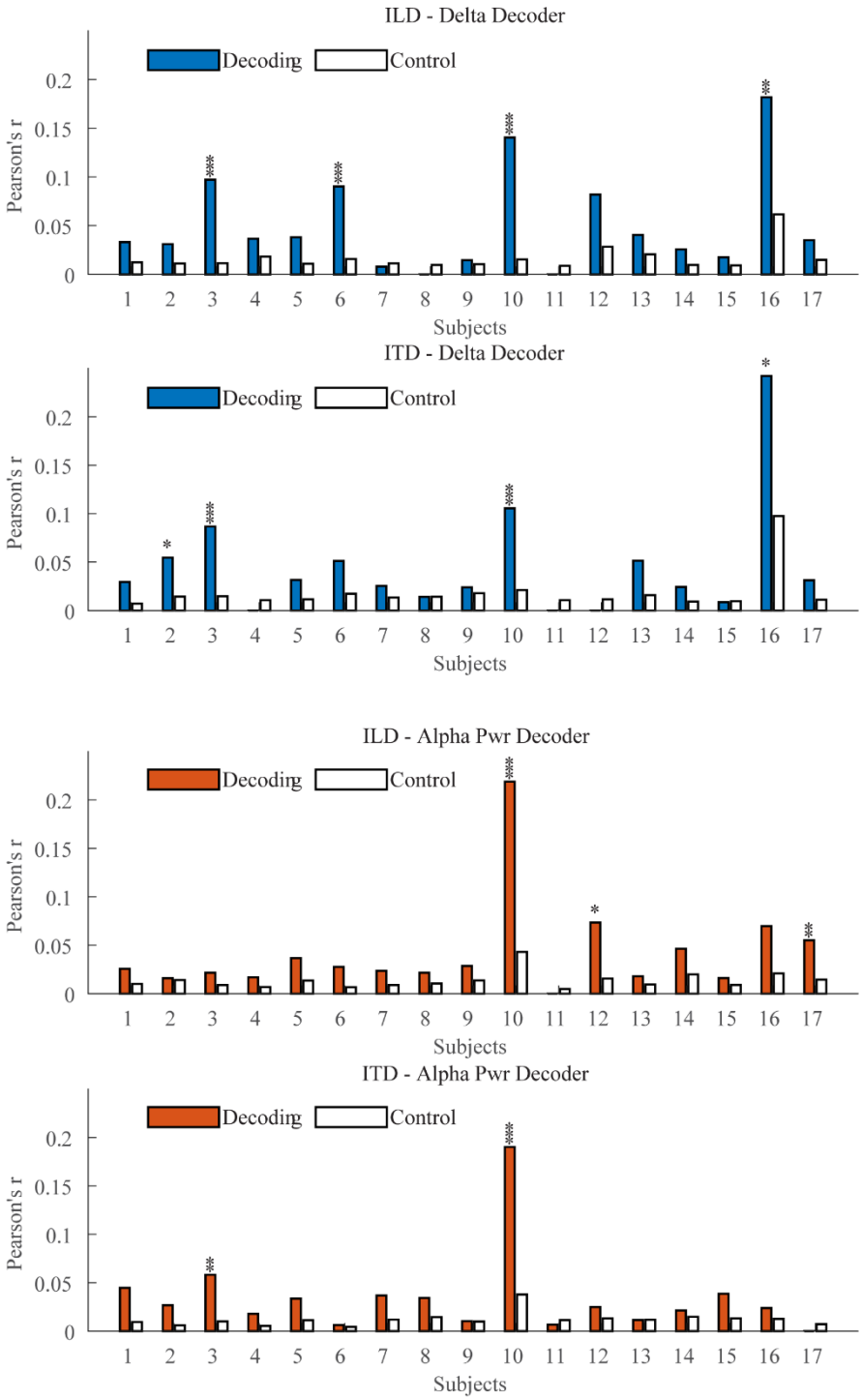
## 3.5. Conclusion

We demonstrated that cortical activity fluctuates in a phase-locked manner with the dynamics of a moving sound source. Specifically, the delta phase and alpha power of EEG were both found to contribute to the decoding of sound azimuth. Moreover, using spatially impoverished stimuli, we found that this cortical tracking is equally present for both binaural acoustic cues ILD and ITD.

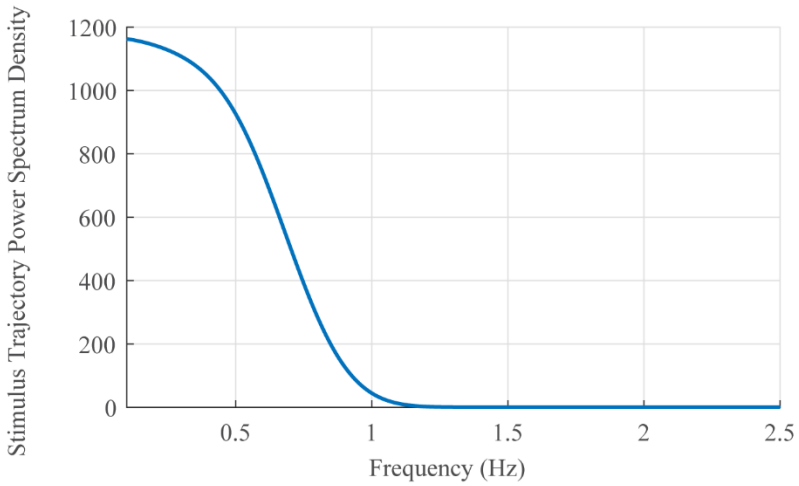
## 3.6. Appendix



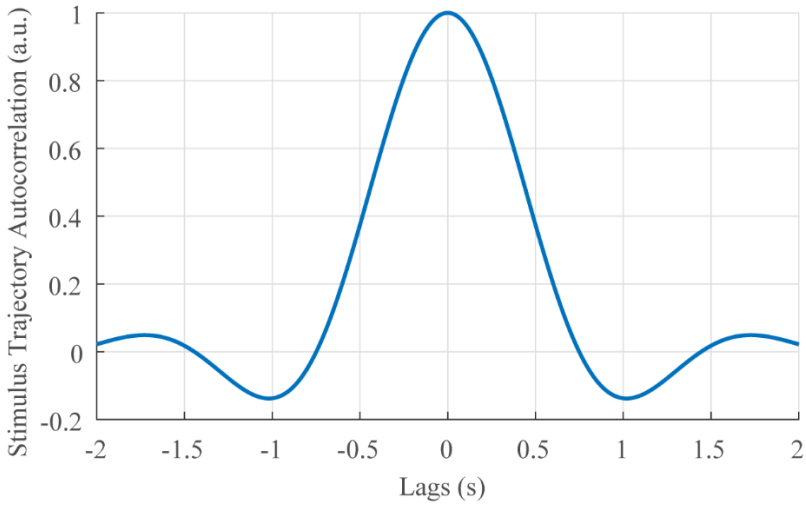

*Appendix A. Individual azimuth reconstruction accuracies shown for Delta (blue), Alpha Pwr decoders (red). The decoding performance is represented as Pearson's r between the reconstructed and the original sound trajectory. As control condition, to estimate whether our decoding is above the chance-level, we trained the decoders with randomly permuted trajectories between the trials. * ** *** indicates within-subject classification with decoding accuracy significantly higher than the chance level of $p<0.05$ $p<0.01$ and $p<0.005$ respectively.*




*Appendix B. Individual azimuth reconstruction accuracies shown for Delta (blue), Alpha Pwr decoders (red) for ILD and ITD trials. The decoding performance is represented as Pearson's r between the reconstructed and the original sound trajectory. As control condition, to estimate whether our decoding is above the chance-level, we trained the decoders with randomly permuted trajectories between the trials. * ** *** indicates within-subject classification with decoding accuracy significantly higher than the chance level of p<0.05 p<0.01 and p<0.005 respectively*




*Appendix C. Frequency spectrum of moving stimulus trajectory (azimuth time-series). Majority of stimulus trajectory power is distributed under 1 Hz.*




*Appendix D. Autocorrelation plot of stimulus trajectory (azimuth time-series).*

# Chapter 4. Decoding Locations of Attended and Unattended Moving Sound Sources using EEG

## 4.1. Introduction

From section 2.4.6, we know that our auditory system allows us to identify the location of multiple sound sources in our environment and to follow those sources as they change position. This is behaviourally important in and of itself, but also helps us to attend to and understand speech in complex listening scenarios (Cherry, 1953; Shinn-Cunningham *et al.*, 2001).

As we also discussed in chapter 2, while a substantial body of literature exists on how single-sound source location is encoded in subcortical and cortical neurons, much work remains to be done to fully characterize how cortex represents of the location of multiple auditory objects in complex scenes. An animal study by Middlebrooks and Bremen (2013) showed that primary AC neurons preferentially synchronized to one of the two stimuli and, in effect, segregated the streams. From human studies on auditory scene analysis, we know that spatial cues alone can facilitate the stream segregation even when other cues are unavailable (Best *et al.*, 2004b; Middlebrooks & Onsan, 2012; Carl & Gutschalk, 2013). This suggests we can process overlapping spatial cues corresponding to several sound sources and one can speculate that the spatial representations of multiple sound sources might exist within the ascending auditory pathway. However, it has been proposed that distinct neural pathways are involved in spatial stream segregation vs sound location processing (Middlebrooks & Onsan, 2012) and a recent fMRI study demonstrated that spatial separation between sources but not their locations can be decoded from the auditory cortex (Shiell *et al.*, 2018).

When investigating hearing in acoustic scenarios with multiple sources, it is also important to consider the effects of top-down attentional mechanisms on cortical responses. One of the most common examples of top-down selection in complex acoustic environments is the so called “cocktail party problem”, which describes a situation when one is attending to one speaker in the presence of other competing speakers (Cherry, 1953). In such situations, it has been shown that the auditory cortex primarily tracks the dynamic changes of the attended stream. Nevertheless, the unattended stream tracking is also reflected in the cortex although the tracking is weaker (Ding & Simon, 2012a; Mesgarani & Chang, 2012; O'Sullivan *et al.*, 2015).

In our previous study (chapter 3), we showed that the cortex tracks the time-varying location of a continuously moving sound source and we demonstrated that it is possible to reconstruct

the sound trajectory from EEG using a linear reconstruction model. Here, we used the same stimulus-reconstruction framework to investigate how cortex encodes spatial information in a multi-source acoustic scenario and we wanted to assess the effect of attention on our ability to decode the locations of simultaneous, competing stimuli. We tested the trajectory reconstruction in two different experiments that differed in their stimulus and behavioural task:

Experiment 1 employed continuous band-passed filtered noise stimuli with an active sound localization task. This builds on our previous study from chapter 3, where we successfully reconstructed the trajectory of noise stimuli in a single source scenario and we were interested in applying this approach in a multi-source environment using already tested stimuli. The active localization task was used as it was previously shown that spatial attention augments the location sensitive responses of the cortex (Zatorre *et al.*, 2002; Ahveninen *et al.*, 2006; Altmann *et al.*, 2008).

In experiment 2, we used competing speech stimuli. Unlike the noise stimuli, the speech stimuli had fluctuating envelope over time. This allowed us to test whether the EEG activity reflects sound azimuth of more complex and less artificial audio signals. The second reason to choose speech stimuli was to investigate if we can use location reconstruction approach to decode selective attention in a cocktail party scenario. For this experiment, we chose to use speech-comprehension task rather than a sound localization task as it more naturalistic in scenarios where subjects listen to speech and we also wanted to confirm that the location tracking does not rely on active localization task. In our previous study from chapter 3, we found that we can reconstruct a single sound source trajectory from the phase of delta (<2 Hz) and power of alpha (8–12 Hz) EEG. Here, we predicted that in the multi-source scenario the delta and alpha components of EEG might be differentially sensitive to bottom-up source location and top-down attention: (1) We predicted that the delta EEG component would track the trajectory of both the attended and the unattended sources, however, the unattended would be tracked to a lesser extent. This was based on envelope-reconstruction studies, which showed that the low-frequency cortical oscillations predominantly synchronize with the temporal structure of the attended sound stimulus, and less with the unattended source (Ding & Simon, 2012a; O'Sullivan *et al.*, 2015). (2) We predicted that the alpha power tracking would be restricted to the attended sound as it has a long-standing association with allocation of spatial attention (Kerlin *et al.*, 2010; Wöstmann *et al.*, 2016).

The findings described in this chapter have been submitted as a research article: Bednar, A., & Lalor, E. C. “Where is the cocktail party? Decoding locations of attended and unattended

moving sound sources using EEG." *NeuroImage*, in review.

## 4.2. Material and Methods

***Participants.*** In total 29 participants (median = 20 years; min = 18 years; max = 30 years; 12 females; 23 right handed) participated in this study with informed consent. Nineteen subjects took part in the first experiment and ten in the second. Two subjects were excluded from the first experiment due to inability to perform the task. All subjects reported no neurological diseases and normal hearing. The experiments were approved by the Research Subjects Review Board of the University of Rochester.

***Experimental Procedures.*** Participants were presented with two simultaneous continuously moving auditory stimuli via headphones and were asked to pay attention to one of the presented audio stimuli, 'attended' stimulus, while ignoring the other 'unattended' stimulus.

There were two experimental conditions, which differed in the type of auditory stimulus and the task performed by the subjects. In experiment 1, we used filtered noise stimuli and subjects performed a spatial task. In experiment 2, subjects listened to continuous speech stimuli and performed a speech comprehension task.

In both experiments, the attended and unattended stimuli were spatialized to be perceived as randomly moving within the frontal part of horizontal plane using Oculus Audio SDK, which simulates head-related transfer function (HRTF) filtering. The pseudo-random trajectories for both stimuli were generated independently, simulating smooth but unpredictable sound movement. The stimuli were set to move on a semi-circular trajectory in the horizontal plane between -90˚ (left) and +90˚ (right) relative to the subject (see figure 1A and B). The simulated motion of the source had an average angular velocity of 50˚ / second. During the experiment, the subjects sat in a dark soundproof room. To minimize movement, the participants were asked to look at a fixation cross displayed on a computer screen directly in front of them.

***Experiment 1- Moving Noise Stimuli.*** Subjects were presented with two moving concurrent band-passed noise stimuli, each having different spectral content, 'Low' and 'High'. Both stimuli were continuous band-pass filtered noise with a bandwidth of 1 kHz, with the center frequency being 750 Hz for the Low stimulus and 4500 Hz for the High stimulus (see figure 4-1C and D). Before each trial, the subjects were cued to attend to either the Low or High stimulus by text on the computer screen.

The subjects performed a target localization task. The task required subjects to respond with a button press to infrequent tremolo targets (modulation frequency 4 Hz, 2 s long) in the

attended stream and indicate whether the target originated from the left or right side of the auditory space using left or right mouse buttons. After each trial, the subjects were shown a target detection score on the screen. The subjects undertook a total of 44 trials, each trial lasting 2 minutes. The number of targets within each trial ranged from 1 to 4 per trial. The experiment was divided into 4 blocks, each having 11 trials. In experimental blocks 1 and 3 subjects attended to the Low stimulus, while in blocks 2 and 4 attended to the High stimulus.

***Experiment 2- Moving Speech Stimuli.*** Subjects were presented with two different speech streams (audiobooks), each read by different male speaker. Silent gaps in the audio streams exceeding 0.5 s were truncated to 0.5 s in duration. See O'Sullivan *et al.* (2015) for details. Again, the stimuli were spatialized, so the speakers performed random continuous motion in the horizontal plane. Before each trial, subjects were cued to attend to one of the speakers. See figure 4-1D for example stimuli.

The subjects performed a speech comprehension task and were required to answer between 4 and 6 multiple-choice questions on both stories after each trial. Each question had 4 possible answers. After answering the question, the subjects were shown correct answers. The subjects undertook a total of 40 trials, each trial lasting 1 minutes. The experiment was divided into 2 blocks, each having 20 trials. For the first 20 trials the subjects paid attention to one speaker and for the next 20 trials they paid attention to the other speaker.

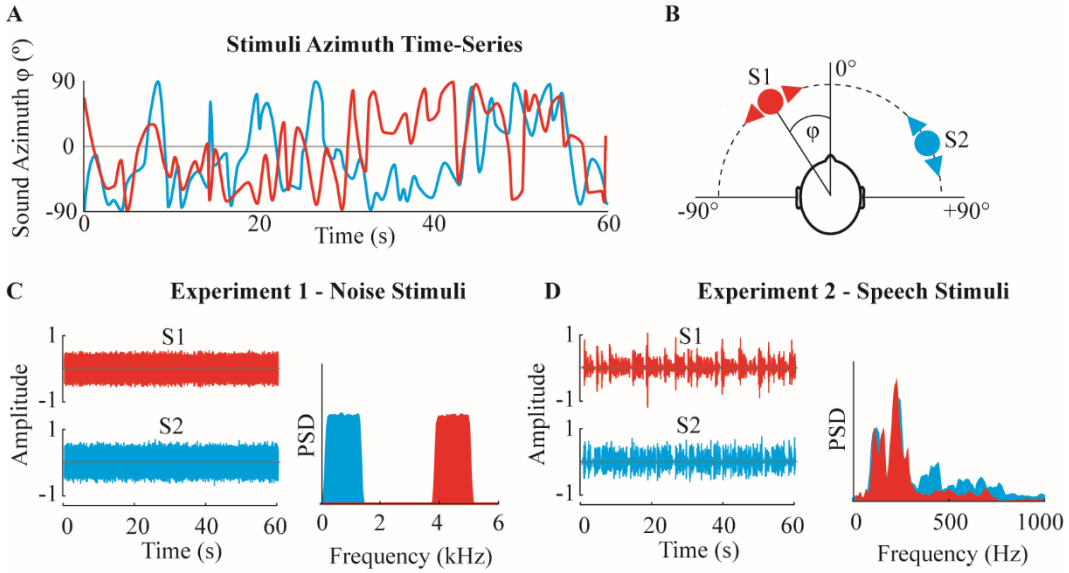



*Figure 4-1. Continuously moving auditory stimuli. In experiment 1 (noise stimuli) and experiment 2 (speech stimuli), the subjects were presented with two concurrent stimuli, which were HRTF-filtered to be perceived as randomly moving in the frontal part of the horizontal plane between -90 (left) and +90 degrees (right). (A) Example of stimulus trajectories shown as azimuth time-series. (B) Top view of the auditory scene. (C) Left:*

*Example waveforms of band-passed filtered noise stimuli that were used in Experiment 1. Right: Power spectrum density (PSD) shown for Low and High noise stimuli having bandwidth of 1kHz and center frequencies of 750 Hz and 4500 Hz respectively. (D) Left: Example waveforms of speech stimuli that were used in Experiment 2. Right: PSD shown for both speech waveforms.*

***EEG data preprocessing.*** The EEG data were recorded using a 128-channel ActiveTwo acquisition system (BioSemi, The Netherlands) at a sampling rate of 512 Hz. Preprocessing was done in MATLAB using custom written scripts and the EEGLAB toolbox (Delorme & Makeig, 2004). The data were filtered between 0.02 and 30 Hz and downsampled to 64 Hz. Bad channels were interpolated from the surrounding channels using the spline function from EEGLAB. Independent component analysis (ICA) was used to remove eye movement artifacts. The components to be removed were preselected by the ADJUST algorithm (Mognon *et al.*, 2011) and then marked for rejection by visual inspection. Finally, the data were re-referenced to the average of all electrodes.

***Multivariate backward modelling.*** We used the same multivariate linear reconstruction approach as in the previous study (chapter 3) to reconstruct the trajectory of the attended as well unattended sound trajectory from the EEG data. We also repeated this analysis where we focused on reconstructing the spatial trajectory of one stimulus, while controlling for (i.e., regressing out) the effect of the other stimulus on the EEG data. To do this we used a linear predictive (forward) model, which we described in detail in chapter 2.3.4: We first trained a forward model to predict EEG from a given stimulus trajectory. Next, we subtracted that predicted stimulus trajectory from the original EEG signal. Then we fed the residual EEG data into our stimulus trajectory reconstruction analysis described above.

***Decoding selective attention.*** In the second experiment, we attempted to use our trajectory reconstruction method to decode selective attention i.e. to determine which of the two speakers was attended from the EEG data. The traditional approach for determining the attended speech stream has been based on stimulus envelope reconstruction. Briefly, the stimulus envelope is reconstructed from the EEG and compared with actual attended and unattended speech envelopes using Pearson's r, then the envelope which has the largest correlation with the reconstructed envelope is deemed the attended envelope (e.g. O'Sullivan *et al.*, 2015).

Here, we used a similar approach as O'Sullivan, with a main difference that we were reconstructing sound trajectory instead of sound envelope: First, we used a leave-one-out cross-validation approach to train an "attended model", which reconstructed the attended stimulus trajectory from the EEG, and used this model to reconstruct the trajectories of the

left-out trials. Afterwards, for each subject and each trial, we compared the reconstructed trajectory with the actual attended and unattended stimulus trajectories by calculating Pearson's r, which we will refer to as $r_{traj\text{-}ttended}$ and $r_{traj\text{-}unattended}$. We then used a linear discriminant analysis (LDA), which used $r_{traj\text{-}ttended}$ and $r_{traj\text{-}unattended}$ as input features, to determine whether the given trial was attended or unattended. Again, this was done using a leave-one-out cross-validation approach. Afterwards, the single-subject decoding accuracy was calculated as the proportion of correctly classified trials. This was done for Delta and Alpha Pwr models separately.

In addition, to get a better idea how the performance of trajectory-based attention decoders compares to envelope-based methods, we use the above mentioned approach to train an envelope decoder on the current dataset. That is, we repeated the analysis using the stimulus envelope instead of the stimulus trajectory and ran the LDA using $r_{env\text{-}ttended}$ and $r_{env\text{-}unattended,}$ values corresponding to the correlation values between the reconstructed envelope and the actual attended and unattended envelopes, as input features. As done by O'Sullivan *et al.* (2015), this decoding was performed on EEG that was filtered between 2 and 8Hz.

Finally, we also tried to combine both the trajectory and envelope based attention decoders. To do so, we ran the LDA decoder using an input feature vector containing 6 correlation values: two pairs of $r_{traj\text{-}attended}$ and $r_{traj\text{-}unattended}$ corresponding to the Delta and Alpha trajectory decoders and $r_{env\text{-}attended}$ and $r_{env\text{-}unattended}$ from the envelope decoder.

To statistically assess whether classification performance is above theoretical chance on a single-subject level, we used a nonparametric permutation test (Combrisson & Jerbi, 2015). First, to establish a null distribution of the classification accuracies, we repeated the LDA 1000 times with randomly permuted classification labels. Afterwards, we used the tail of this empirical distribution to calculate the p-value for the original classification. The group level statistical comparison was conducted using one-sided Wilcoxon signed-rank test by testing the actual classification accuracy of each subject against the chance level accuracy that was obtained by averaging the null-distribution.

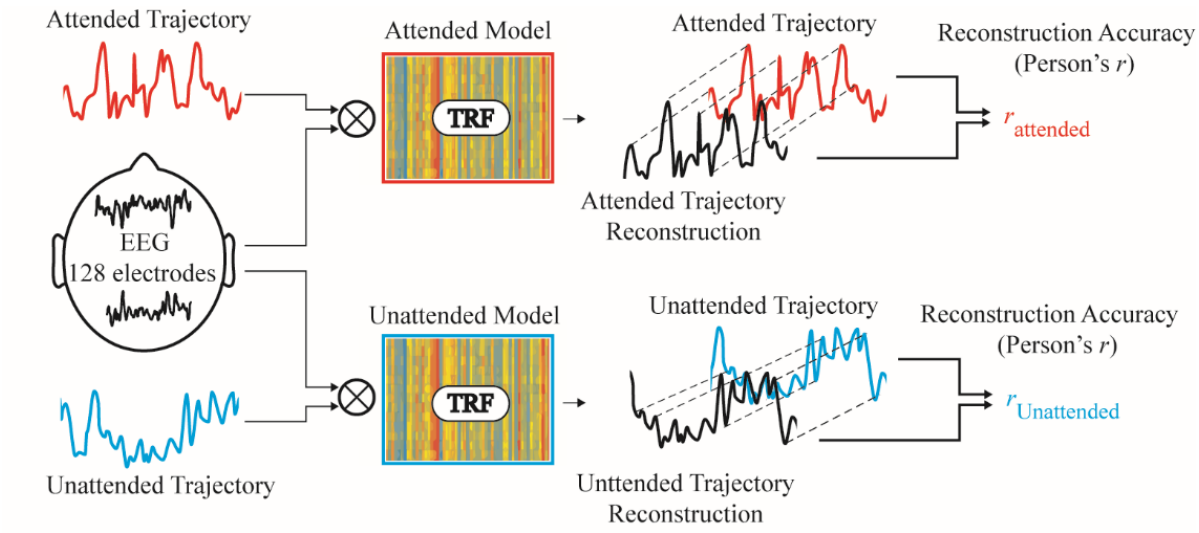



*Figure 4-2. Decoding EEG analysis. For both attended and unattended sound stimuli, we trained regression models to separately reconstruct the attended and unattended trajectory from EEG. The reconstruction models were using simultaneously information from all EEG channels and time lags corresponding to time window 0-250ms. The models were trained using leave-one-out cross-validation approach and the reconstruction accuracy of both trajectories was assed using Pearson's r by correlating the actual and reconstructed trajectories.*

## 4.3. Results

### 4.3.1. Experiment 1 – Noise Stimuli

We investigated whether the spatial dynamics of multiple moving sound sources are separately reflected in scalp recorded EEG and were interested in how attention affects the neural tracking of those spatial dynamics. Specifically, we attempted to decode the trajectories of two simultaneously presented noise stimuli from EEG in a situation when one source was attended and the other unattended. Our hypothesis was that the delta phase and alpha power components of EEG might be differentially sensitive to bottom-up effects induced by spatial properties of the stimuli and top-down attentional selection effects.

***Behavioural results.*** During the experiment, subjects were asked to indicate the spatial location of tremolo targets within the attended audio stream. This was done to confirm that subjects attended the correct stimuli and to quantify how well subjects perceived the spatial aspects of the auditory scene. Subjects responded quite accurately to this challenging task with 72.8% of attended targets correctly detected. And they were successfully able to ignore the unattended with only 4.7% of targets in that stream being responded to as false alarms. The group-average sensitivity d-prime index was d=2.08. We also calculated d-prime sensitivity indices separately for low (d= 1.67) and high stimuli (d= 1.51) and found no

significant differences.

***Trajectory reconstruction results.*** First, we investigated what EEG signal frequency bandwidth would be the most informative for sound trajectory reconstruction. To test this, we filtered the EEG into different frequency bands and for each band we separately ran the trajectory reconstruction models. This was performed on EEG amplitude signal as well as on EEG power. Consistent with our previous single-source study (Bednar & Lalor, 2018), for both attended and unattended EEG the delta (0.05-2Hz) and the alpha power (8-12Hz) of EEG were the most informative for sound trajectory reconstruction (see figure 4-3). In further analysis, we therefore focused on these two EEG subcomponents. Specifically, we individually trained the trajectory reconstruction models on the delta phase and the alpha power of EEG, naming the models "Delta" and "Alpha Pwr" respectively.

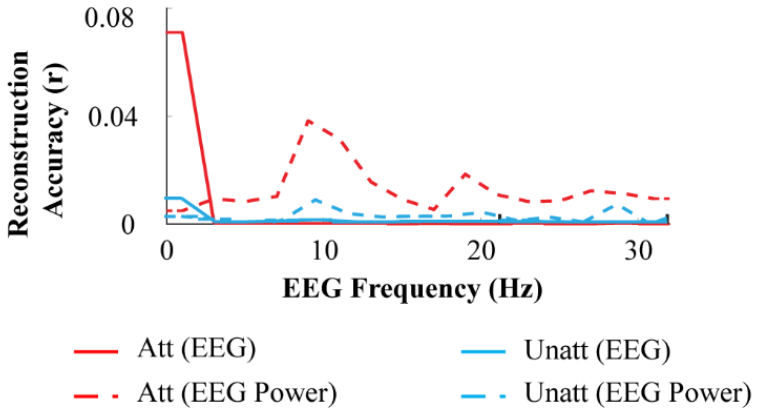



*Figure 4-3. Experiment 1- Noise Stimuli: Trajectory decoding performance dependency the EEG bandwidth. The trajectory reconstruction accuracies shown for different frequencies of raw EEG signal (solid) and EEG power (dashed) for the attended (red) or unattended (blue) stimuli. The EEG power was calculated as an absolute value of a Hilbert transform of a raw EEG signal. Different frequencies of the signals were obtained using a sliding band-pass filter with pass-band width of 2 Hz.*

Using both Delta and Alpha Pwr decoders, we found that we could reconstruct the sound trajectory of the attended noise source with accuracies above the chance level. For the attended sound source, the average reconstruction correlations were r=0.079, p=1.6e-4 and r=0.055, p=1.9e-4 for Delta and Alpha Pwr respectively. For the unattended sound source, the reconstruction accuracies were below the chance level, r=0.01, p=0.56 for the Delta decoder and r=0.013, p=0.13 for the Alpha Pwr decoder. The trajectory reconstruction accuracies showed strong effects of attention. Specifically, the reconstruction correlation values were larger for the attended sound source than the unattended for Delta (p=3e-4) as well as the Alpha Pwr of the EEG (p=6e-4). (See left panel in figure 4-4.)

In our experiment, the attended and unattended source trajectories shared a common location

at times. In order to minimize any confounding effects that location overlap between the sources might have on our assessment of the reconstructed trajectories, we ran the decoders on the EEG that had the unattended trajectory partialled-out, i.e., that contained only spatial information about the attended source. Similarly, we evaluated the decoders on EEG that had regressed-out the attended trajectory (see methods section for details). When using EEG data with partialled-out unattended source, the trajectory reconstruction accuracy of the unattended source decreased to zero (trivially) and the reconstruction accuracy of the attended source did not significantly change. This pattern was the same for the Delta and the Alpha Pwr decoders. (See middle panel in figure 4-4).

When we removed the attended trajectory information, the reconstruction accuracy of the attended trajectory decreased below the chance level ($p<0.05$). The accuracy of the unattended source trajectory reconstruction increased on average, however it remained below the significance threshold of $p<0.05$. Again, this effect was the same in both Delta and Alpha decoders. (See right panel in figure 4-4).

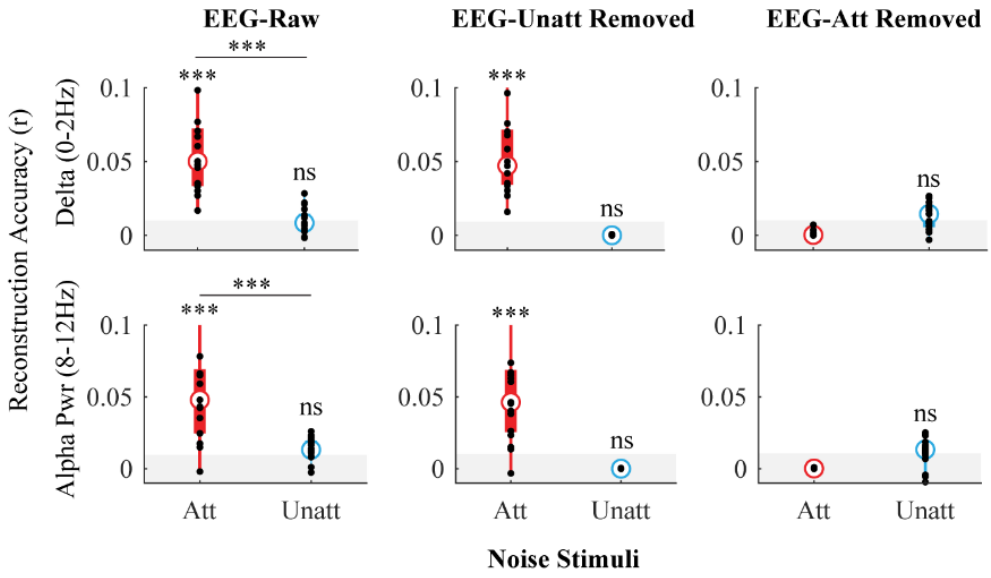



*Figure 4-4. Experiment 1- Noise Stimuli: Reconstructing sound trajectories of two continuously moving noise sources from EEG. Stimulus trajectory reconstruction accuracies are shown for the attended ("Att", red) and unattended stimuli ("Unatt", blue) for Delta and Alpha Pwr decoders. The reconstruction was performed on: (1) filtered-only EEG ("Raw"), (2) filtered EEG with partialled-out unattended stimulus trajectory ("Unatt Removed") and (3) filtered EEG signal with partialled-out attended stimulus trajectory ("Att Removed"). Shaded gray area at the bottom of each plot indicates significance threshold ($p<0.05$). * ** *** indicates reconstruction accuracies at the level of $p < 0.05$, $p < 0.01$ and $p < 0.005$ respectively.*

We then ran the decoders at different time-lags between the EEG and the stimulus trajectory using a sliding window of 50 ms. We found that the attended source decoding peaked at around 200 and 125 ms for Delta and Alpha Pwr decoders respectively. For the unattended source, the single-lag accuracy was below the significance level threshold. As discussed in Bednar and Lalor (2018), due to the relatively low velocity of the sound stimuli, the reconstruction accuracy remains high over a relatively large time-lag interval.

For selected time-lags, we plotted the decoder activation patterns (see figure 4-5 below). For the attended source, the Delta decoder showed strong activations over temporal scalp bilaterally. The activation patterns were opposite between the hemispheres, i.e., an increase in activity in one hemisphere was accompanied by a decrease in the other. The Alpha Pwr decoder indicated more posterior activation in parieto-occipital areas. The stimulus caused a relative decrease in activity contralateral to the stimulus position, and a relative increase over ipsilateral scalp. These activation patterns indicating activity over fronto-temporal areas strongly resemble the patterns we observed in our previous study that involved listening to single moving sound source (Bednar & Lalor, 2018). For the unattended decoder, although the activation patterns were also opposite between the hemispheres, the spatial distribution of activation patterns differed from the attended decoders and were generally noisier. The Delta decoder patterns were again more frontal and indicated a relative decrease in activity contralateral to the stimulus position. The Alpha Pwr unattended decoder patterns were localized more posteriorly and were more central in comparison to the attended Alpha Pwr decoder. However, one needs to be careful to interpret these patterns as the reconstruction accuracies of these models were below the chance level. Nevertheless, there was still some structure in the decoder weights, which might suggest some weak cortical tracking of the unattended source.

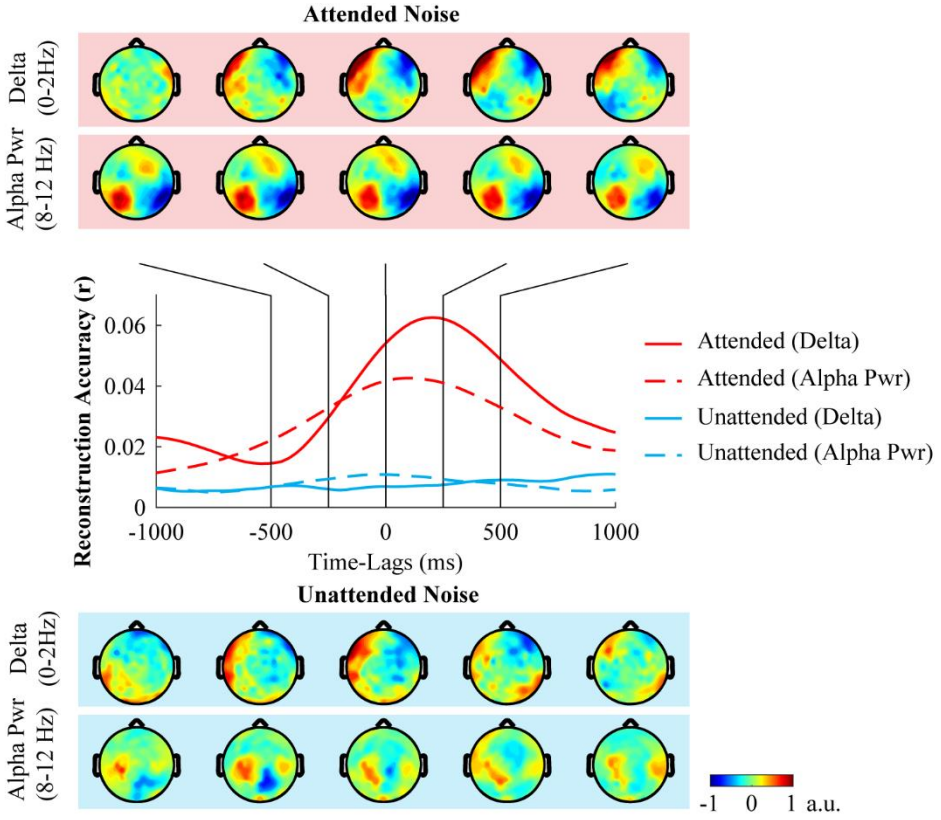



*Figure 4-5. Noise Stimuli: Trajectory decoder activation patterns and single-lag decoding. The decoders were independently evaluated at single time-lags between the EEG and the trajectory signal. Middle: Reconstruction accuracies shown for each individual time-lag. Top and Bottom: Decoder activation patterns corresponding to different time-lags are shown for the attended (red background) and unattended speech stimuli (blue background). The decoder patterns for each time-lag were individually normalized.*

***Correlation with behavioural results.*** We compared the behaviour and trajectory reconstruction data across subjects and found there was no significant correlation between the target detection sensitivity and trajectory reconstruction of the attended or unattended stimuli from the EEG data (all $p>0.05$).

### 4.3.2. Experiment 2- Speech Stimuli

In the second experiment, we were interested in whether we could employ a more naturalistic attention paradigm and attempted to reconstruct the trajectories of two concurrent non-stationary speech sources (speakers). Also, in contrast to the first experiment, the subjects performed a non-spatial speech comprehension task. This was done to see if our spatial decoding measures were still affected by attention even when the task wasn't specifically spatial. Finally, we tried using trajectory reconstruction to decode selective attention in multi-speaker environment and compare it with 'traditional' envelope-based cocktail party decoders. In addition, we tried to see if markers of spatial attention could be added to traditional cocktail party measures to improve the selective attention decoding performance.

***Behavioural results.*** In the second experiment, subjects were asked to answer multiple-choice questions on the attended and unattended story after each trial. On average, the subjects answer correctly 69.5% ($p<0.005$) of the questions from the attended story and 25.1% from the unattended story, which was not greater than chance (25%, $p>0.05$). The number of correctly answered question was significantly higher for the attended than the unattended stream ($p<0.005$).

***Trajectory reconstruction results.*** Using the Delta decoder, we found we could successfully reconstruct the trajectory of the attended and unattended speech sources with reconstruction correlation values of r=0.057, p=4.8e-3 and r=0.024, p=0.032 respectively. For the Alpha Pwr decoder, the correlation values for the attended and unattended stimuli were r=0.042, p=4.2e-3 and r=0.019, p=0.5. The reconstruction values were significantly larger for the attended than the unattended speech stimuli only for the Delta EEG p= 0.019 but not for the Alpha Pwr decoder p=0.065. See figure 4-6 for reconstruction results.

As in the first experiment, we attempted to reconstruct the attended and unattended sources in isolation by regressing-out the other source. Again, the reconstruction accuracies of the attended sound source were unaffected when we removed the trajectory information of the unattended source (see figure 4-6, middle panel). Importantly, when we removed the positional information of the attended sound source from the EEG, it was still possible to reconstruct the unattended source trajectory using the Delta decoder (r=0.023, p=0.032). This indicates that the above chance-level reconstruction accuracy of the unattended sound source trajectory was not simply caused by the time intervals where both sound sources were co-located.

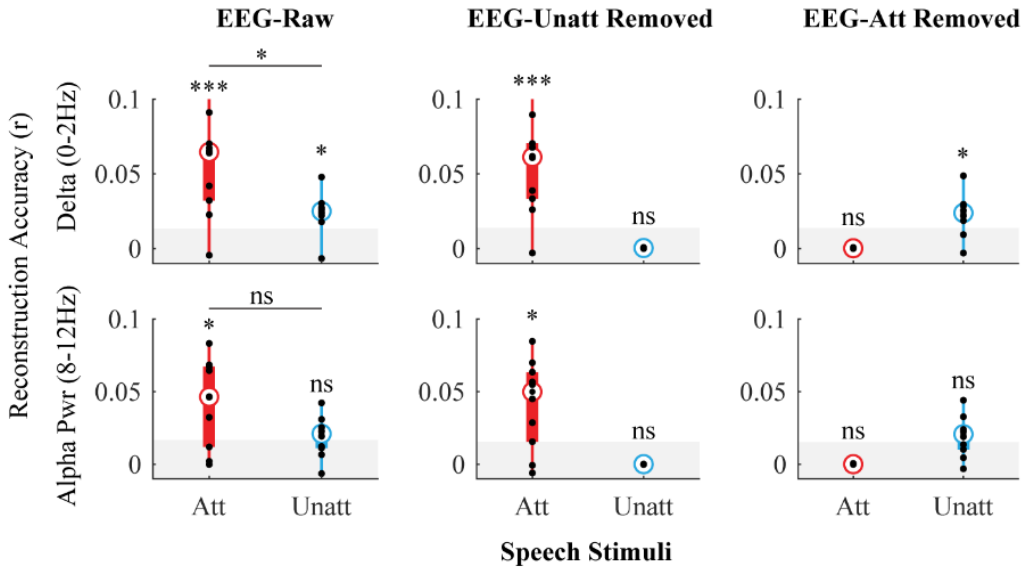



*Figure 4-6. Experiment 2- Speech Stimuli: Reconstructing sound trajectories of two moving speech sources from EEG. Stimulus trajectory reconstruction accuracies shown for the attended ("Att") and unattended stimuli ("Unatt") for Delta and Alpha Pwr decoders. The reconstruction was performed on: (1) filtered-only EEG ("Raw"), (2) filtered EEG with partialled-out unattended stimulus trajectory ("Unatt Removed") and (3) filtered EEG signal with partialled-out attended stimulus trajectory ("Att Removed"). Shaded grey area at the bottom of each plot indicates significance threshold ($p<0.05$). * ** *** indicates reconstruction accuracies at the level of $p < 0.05$, $p < 0.01$ and $p < 0.005$ respectively.*

Figure 4-7 shows the activation patterns of the speech trajectory decoders and the single-lag decoding results. The attended decoder patterns indicate strong contralateral tuning and opposite activations between hemispheres and their morphology resembles the activation patterns from the first experiment.

The activation patterns corresponding to the unattended decoder were relatively noisy and interestingly less lateralized than the attended decoder. Again, this should be interpreted with caution since the reconstruction accuracy of the model was weak. The single-lag decoding showed that the trajectory of the attended sound is best decoded at 195 and 250 ms for Delta and Alpha decoders respectively. In comparison to the first experiment, the Alpha decoding peaked around 125 ms later, which might be possibly attributed to the different nature if the attention task.

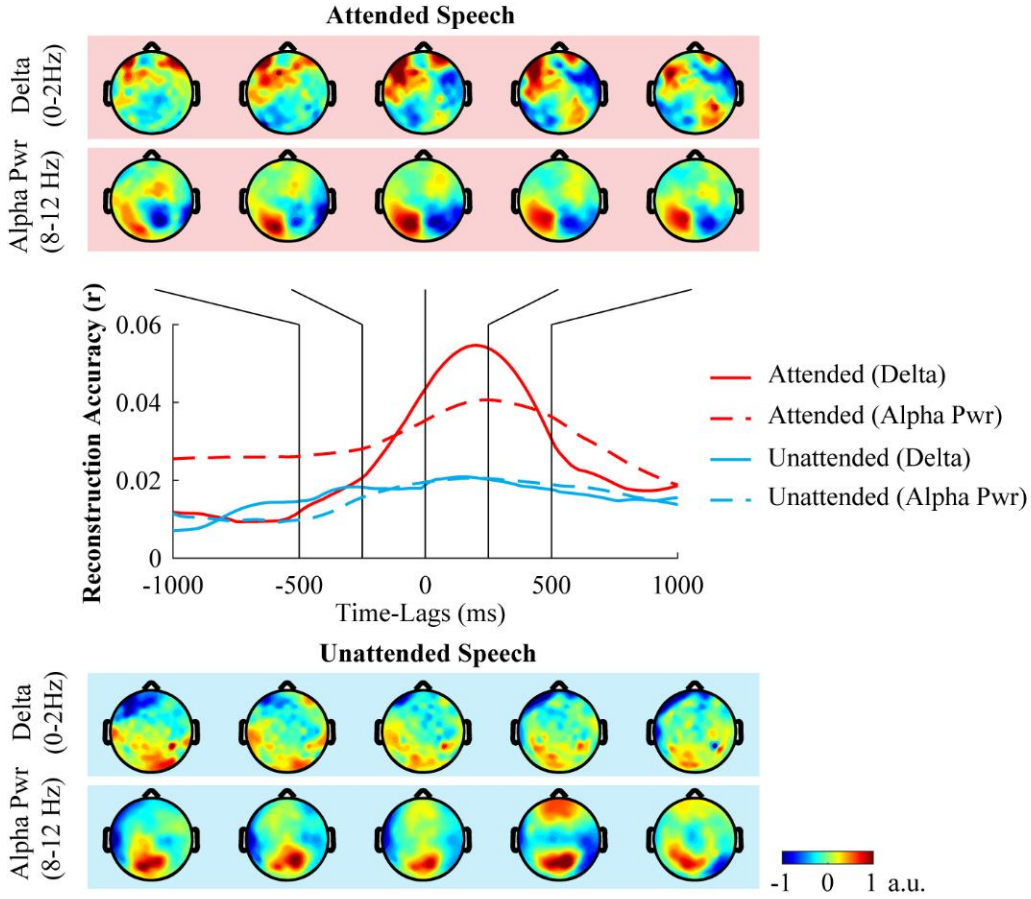



*Figure 4-7. Experiment 2- Speech Stimuli: Trajectory decoder activation patterns and single-lag decoding. The decoders were independently evaluated at single time-lags between the EEG and the trajectory signal. Middle: Reconstruction accuracies shown for each individual time-lag. Top and Bottom: Decoder activation patterns corresponding to different time-lags are shown for the attended (red background) and unattended speech stimuli (blue background). The decoder patterns for each time-lag were individually normalized.*

***Decoding selective attention at single-trial level using sound trajectory and envelope reconstruction.*** As the trajectory reconstruction appeared to be strongly modulated by selective attention, we tested the possibility of using the spatial decoders to decode selective attention in a cocktail party, i.e., to decode which of the two speakers was attended. We used an approach similar to (O'Sullivan *et al.*, 2015), who used a method based on sound amplitude (envelope) reconstruction to decode selective attention from EEG at a single trial level.

Briefly, our method was to train a trajectory reconstruction model on the attended speech stimulus. Next, using a previously unseen EEG trial, we reconstructed the stimulus trajectory and assessed the correlation between the predicted trajectory and each of the two original speech stimulus trajectories. Finally, we used LDA to classify the correlation values as attended or unattended speech stream. As well as this trajectory-based decoding, we also

used the same dataset to perform envelope-based decoding of selective attention, which is using the same regression approach, and compared the results. Finally, we also tried to combine the trajectory- and envelope- based decoder together. See methods section for details.

On a group-level, the trajectory-based approach decoded selective attention with an accuracy significantly above the chance level for the Delta decoder with an accuracy of 66.75% (p=2e-3), but not for the Alpha Pwr decoder, which reached 53.75% (p=0.28). The average decoding accuracy of the envelope decoder was 92% (p=9.8e-4) and was better than both trajectory decoders (p<0.005). Finally, the decoder that combined both trajectory decoders and envelope decoder reached average decoding accuracy of 93% (p=9.8e-4). This was on average higher decoding accuracy than for envelope decoder alone, however, the difference was not significant (p=0.53).Within-subject statistical analysis showed that for the Delta and Alpha Pwr trajectory decoders, the decoding accuracy was above the chance level for 6/10 and 3/10 subjects respectively. For the envelope and combined decoder, the single subject data showed significant decoding accuracy for 10/10 subjects. See results in figure 4-8.

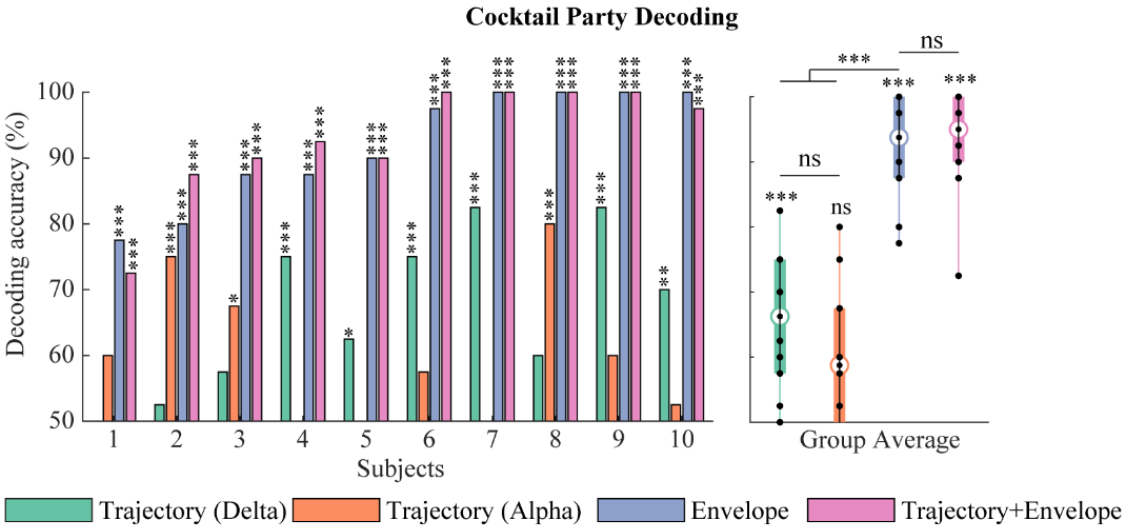



*Figure 4-8. Experiment 2- Speech Stimuli: Decoding selective attention using trajectory and envelope reconstruction models. The decoding was performed using Delta and Alpha Pwr trajectory decoders, envelope decoder and "Trajectory+Envelope" decoder that integrated both trajectory decoders and envelope decoder together. The decoding accuracy is shown for individual subjects (left) as well as group average (right). Subjects are sorted according to the performance of their envelope decoder.* ** *** indicates prediction differences at the level of p < 0.05, p < 0.01 and p < 0.005 respectively.*

## 4.4. Discussion

Here in this study, we were interested in whether we can decode sound source locations from EEG when listeners were presented with two concurrent moving stimuli and attended to one

of them. Specifically, we tried to reconstruct the trajectory of attended as well as unattended stimuli. This was tested in two experiments which differed in their stimulus type and behavioural task: Experiment 1 employed filtered noise stimuli that were discriminable by their frequency content and subjects performed a spatial task. Experiment 2 involved listening to two different speakers telling a story and a speech comprehension task.

**The trajectory of an attended sound can be reconstructed from EEG even in the presence of another stimulus.** Previously, it has been shown that a single-source location can been decoded from recorded neurophysiological data (Zhang *et al.*, 2015; Derey *et al.*, 2016; Bednar *et al.*, 2017). Moreover, in our previous study (Bednar & Lalor, 2018), we showed that in an acoustic environment with one continuously moving sound source, the cortex dynamically tracks the source's location in time and demonstrated that the trajectory of the source can be reconstructed using a regression model. Importantly, we also demonstrated that the trajectory can be successfully applied to the ITD-spatialized pulse-train, which shows that our trajectory reconstruction is not purely driven by envelope fluctuations.

Here in this paper, we showed that in a more complex acoustic environment with two continuously moving sources, the sound trajectory of the attended sound source can be successfully decoded from EEG. As in our previous study, we identified two distinct EEG components that tracked the location of the attended sound: (1) low frequency delta (0-2Hz) component, which showed strong bilateral activation indicating involvement of auditory cortex and (2) the alpha power (8-12Hz) of EEG that was over the parieto-occipital region. The cortical activation within the delta band indicated the hemispheric preference for contralateral stimuli as shown previously (e.g. Palomaki *et al.*, 2000; Palomaki *et al.*, 2005). The alpha EEG activation pattern looked similar to patterns observed in studies that involved deploying spatial attention to auditory stimuli (Kerlin *et al.*, 2010; Wöstmann *et al.*, 2016).

**Active sound localization task is not required to successfully reconstruct the attended sound trajectory.** It has been shown that paying attention to the location of a sound enhances neural tuning to this stimulus feature (Zatorre *et al.*, 2002; Ahveninen *et al.*, 2006; Altmann *et al.*, 2008). In the first experiment, we successfully decoded the sound trajectory when subjects performed a spatial task. However, it was unclear whether it is possible to apply our decoding framework in situations where one does not perform an active sound localization task, as in this case one might expect reduced spatial sensitivity of the cortex. However, as we showed in our second experiment, which involved a non-spatial speech comprehension task, it is possible to reconstruct the attended sound trajectory even when subjects were not

tasked with localizing targets. This is in line with observations from an fMRI study by Deouell *et al.* (2007) that demonstrated that the cortex is spatially sensitive even when subjects do not perform a localization task.

**Do we track the position of the unattended sound?** Previous studies showed that we are sensitive to spatial separation between sound sources within the auditory scene (Zatorre *et al.*, 2002; Shiell *et al.*, 2018). However, it has been unclear whether the exact location of the unattended source is represented within the auditory cortex. We tested this by attempting to reconstruct the unattended sound source trajectory from the EEG signal.

The data from our first experiment, which involved noise stimuli, showed that the unattended sound source position cannot be successfully reconstructed from the EEG signal. However, in the second experiment that involved speech, the results indicated that the unattended source position can be successfully decoded using the delta EEG component. As one might expect, the unattended reconstruction accuracies were lower than for the attended stimulus, which is in line with studies that showed that the selective attention enhances the neural representation of the attended sound (Woldorff *et al.*, 1993; Petkov *et al.*, 2004).

However, in both of our experiments, the trajectories of attended and unattended sources were partially correlated and at times shared the same location. Therefore, we were concerned that the decoding of the unattended sound source was possibly driven by the moments when the locations of both sound sources coincided. In order to test this possibility, we first regressed-out the attended trajectory from the EEG signal and then attempted to reconstruct the unattended sound source location. The results showed that it is still possible to reconstruct the unattended speech stimulus trajectory with a decoding accuracy above the chance level, which further supports the idea that cortex actually tracks the unattended speech source trajectory.

**Location tracking by delta and alpha power EEG possibly represent different mechanisms.** We predicted that the Delta decoder might track attended as well as unattended trajectories and we expected that the unattended tracking would be weaker. This assumed that the delta EEG trajectory tracking would directly encode the spatial location of the stimulus and that that would just be modulated by attention, similar to amplitude envelope tracking by sub-alpha band (<8hz) EEG, which can be seen in cocktail party studies (e.g. O'Sullivan *et al.*, 2015) that showed this pattern. In contrast, for the alpha power EEG component, which is often associated with spatial attention per se (Kerlin *et al.*, 2010; Wöstmann *et al.*, 2016), we hypothesized that it would track only the attended sound source trajectory. Indeed, the results from our second experiment, which employed speech stimuli,

support this dissociation of delta and alpha power components. However, for the noise stimuli (experiment 1), the decoding accuracy of the unattended sound source was below the detection level for both frequency components of EEG.

A question remains then as to what the delta and alpha power EEG components actually represent in the context of sound location tracking. The morphology of the delta EEG reconstruction model indicated enhanced cortical responses contralaterally to the stimulus location. This suggests that the delta EEG response reflects intrinsic contralateral tuning of the auditory system that has been described in other studies (e.g. Palomaki *et al.*, 2000; Lewald & Getzmann, 2011; Briley *et al.*, 2013; Derey *et al.*, 2016). With respect to the alpha EEG component, as we mention above, the lateralized posterior scalp alpha power is often related to auditory spatial attention. In line with this, we found that the alpha tracking is restricted to the attended sound source. This might suggest that our alpha power-based reconstruction reflects allocation of spatial attention rather than the spatial location of the sound source per se. Indeed the parietal distribution of our alpha power decoder weights supports the notion that this might reflect auditory spatial attention (Banerjee *et al.*, 2011). However, we cannot completely rule out the possibility of a cross-modal effect with our auditory stimulus eliciting covert visual attention effects. In line with this idea, a recent study by Feng *et al.* (2017) showed that a spatial auditory cue caused posterior alpha lateralization and further demonstrated that the changes in alpha activity predicted the accuracy in subsequent visual task. The authors also speculated that the posterior alpha desynchronization might reflect a general priming mechanism that facilitates visual processing.

**Is the trajectory tracking motion specific?** It has been discussed whether the moving sounds engage different cortical networks than static sounds. Neuroimaging studies have indicated that non-primary auditory cortex and particularly planum temporale (PT) is involved motion processing (Baumgart *et al.*, 1999; Pavani *et al.*, 2002; Warren *et al.*, 2002; Krumbholz *et al.*, 2005a). However, these structures often respond to spatial static sounds and so it is debatable as to whether they represent true motion detectors. Nevertheless, a recent study provided strong evidence for cortical motion sensitivity and showed that activity in posterior belt and parabelt cortical regions cannot be explained by static spatial and spectro-temporal processes (Poirier *et al.*, 2017). Although the cortical motion sensitivity might have contributed to our trajectory reconstruction, we speculate that the spatial tracking as we observed is not motion specific. As we discussed above, the contralateral tuning of the cortex (Palomaki *et al.*, 2000; Palomaki *et al.*, 2005; Magezi & Krumbholz, 2010; Briley *et*

*al.*, 2013) as well as alpha lateralization (Kerlin *et al.*, 2010; Wöstmann *et al.*, 2016; Feng *et al.*, 2017) have been mainly demonstrated for static sound stimuli.

**Different spatial sensitivity to moving noise and speech stimuli.** It has been proposed, that in a multi-source environment, listeners are able to collect spatial information from both sources during "glimpses" where sound envelopes do not overlap and this allows listeners to track the position of multiple sources (Yost & Brown, 2013). This could explain why we were unable to reconstruct the unattended sound trajectory in the first experiment, which used noise stimuli, and why we were successful in the second experiment that employed speech. Specifically, in the first experiment, the noise stimuli had relatively constant intensity and were ongoing during the whole trial length, which possibly did not allow to listeners to "collect" enough information about the unattended sound source. In contrast, in the second experiment that employed competing speech, the stimuli had varying intensity included periods of silence, e.g., breaks between words and sentences. Therefore, subjects could possibly utilize these glimpses to tracks both sources. Another possibility as to why we were able to decode the location of the unattended speech but not the unattended noise stimulus is that it is easier to localize and follow speech than filtered noise due to having access to more spatial acoustic cues, as it has been previously shown that it is easier to localize amplitude modulated signals (Blauert, 1997).

**Potential for using stimulus trajectory reconstruction to decode selective attention.** With more practical applications in mind, we tested whether one can use the trajectory reconstruction approach to build a decoder of selective attention, which can be used in cognitively steered hearing aids (Lunner *et al.*, 2009; Mirkovic *et al.*, 2015). We used the same approach as the envelope-based reconstruction decoder, which is capable of successfully solving the "cocktail party problem" on a single-trial level (O'Sullivan *et al.*, 2015). We demonstrated that it is possible to use the trajectory reconstruction approach to decode to which speaker the listener was attending with above-chance accuracy. However, note that this method assumes that the locations of speakers are known, which might not be easily achievable in practice.

We also compared the results with the envelope-based decoder of selective attention. By a clear margin, the envelope-based model outperformed the trajectory reconstruction model, showing that the neural tracking of the attended stimulus envelope is more robust than that of the stimulus location. We also tried to combine the trajectory and envelope decoders together, however, although the combined decoder led to slightly better average accuracy, this gain was not significant. Nevertheless, unlike the envelope decoder, the trajectory

reconstruction model directly predicts the sound source angle, which can be theoretically used as control signal for cognitively steering hearing aids. Therefore, it might be of interest in future studies to investigate the possibility of integrating both types of decoders

## 4.5. Conclusion

We showed the trajectory of an attended sound source can be reliably reconstructed from EEG data and that the reconstruction results are robust to the presence of distracting auditory stimuli and the specific behavioural task. With respect to reconstructing the trajectory of the unattended sound, our data suggested that the cortical representation of the unattended source position is below detection level for the noise stimuli but we observed weak tracking of the unattended source location for the speech stimuli. Our data indicated that the ability to track the unattended sound source position in time was likely driven by temporal glimpses in which the sound intensities of the attended and unattended sources did not overlap. We also demonstrated that the trajectory reconstruction method can be in principle used to decode selective attention on a single-trial basis, however, its performance is inferior to envelope-based decoders.

# Chapter 5. Higher Orders of Motion: Decoding Velocity, Speed and Acceleration

## 5.1. Introduction

It has been shown previously that the auditory cortex is sensitive to sound stimulus azimuth (see chapter 2.4.2) and we have shown in study 3 that the time-varying sound stimuli azimuth of a moving sound can be reconstructed from the EEG data. However, as we discussed in section 2.4.7, less is known about cortical representation of higher-order measures describing sound source motion such as velocity and acceleration and it is unclear whether these measures have representation in the cortex that is separate from azimuth encoding.

With respect to our ability to perceive sound velocity, there are several psychophysical studies showing that we are sensitive to the velocity of moving sound (Chandler & Grantham, 1992; Carlile & Best, 2002; Kaczmarek, 2005). Furthermore, an fMRI study demonstrated that posterior superior temporal regions and premotor areas are possibly involved in velocity processing and showed increased activation in these structures for when contrasting fast- and slow-moving sources. In line with that, electrophysiological studies found that latency and amplitude of the neural responses to motion onset are modulated by stimulus velocity (Makela & McEvoy, 1996; Xiang *et al.*, 2005; Getzmann, 2009). However, above-mentioned studies were limited as they used short discrete sound stimuli and so it is unclear whether the cortex respond to dynamic changes in velocity of a sound stimulus that is continuously moving.

Although we can clearly perceive sound acceleration, i.e., we can also distinguish between sound sources that are speeding up or slowing down, to our best knowledge, the cortical encoding of sound acceleration has not been investigated in the past and the neural correlates of sound acceleration are unknown.

In this study, we propose that our stimulus reconstruction framework may be suitable to investigate the neural sensitivity to velocity and acceleration. Specifically, we want to test whether the cortex reflects time-varying velocity and acceleration by reconstructing these measures from the EEG signal. Firstly, this method allows us to use continuous stimuli, which goes further than previous studies investigating neural responses to velocity using motion-onset evoked responses and short discrete stimuli. Secondly, as our reconstruction approach is more sensitive than traditional approaches of analysing EEG data, we consider it particularly suitable for exploratory investigation of sound acceleration encoding. Lastly, using the regression modelling approach, one can control for the potential correlation

between the variables and investigate each of the motion measures in separation.

## 5.2. Material and Methods

***Participants***. We used the dataset from our second study that employed simultaneous presentation of two noise stimuli, attended and unattended (see section 4.2). Here, we chose to run our analysis only on the attended stimulus trajectory, which was found to have stronger cortical representation (see chapter 4 for more details).

Note that originally, we attempted to answer these questions using the data from study 1 (chapter 3). However, we found this dataset to be unsuitable for this analysis as the first and second derivatives of azimuth (velocity and acceleration) exhibited strong periodicity and were strongly autocorrelated. This has been fixed in the second study.

***Stimuli.*** For each stimulus trial, we extracted first and second differentials of the stimulus trajectory (azimuth time-series) i.e. angular velocity and angular acceleration, to which we will refer as 'velocity' and 'acceleration'. We also calculated stimulus speed as the absolute value of velocity. See figure 5-1 below for comparison of these motion measures.

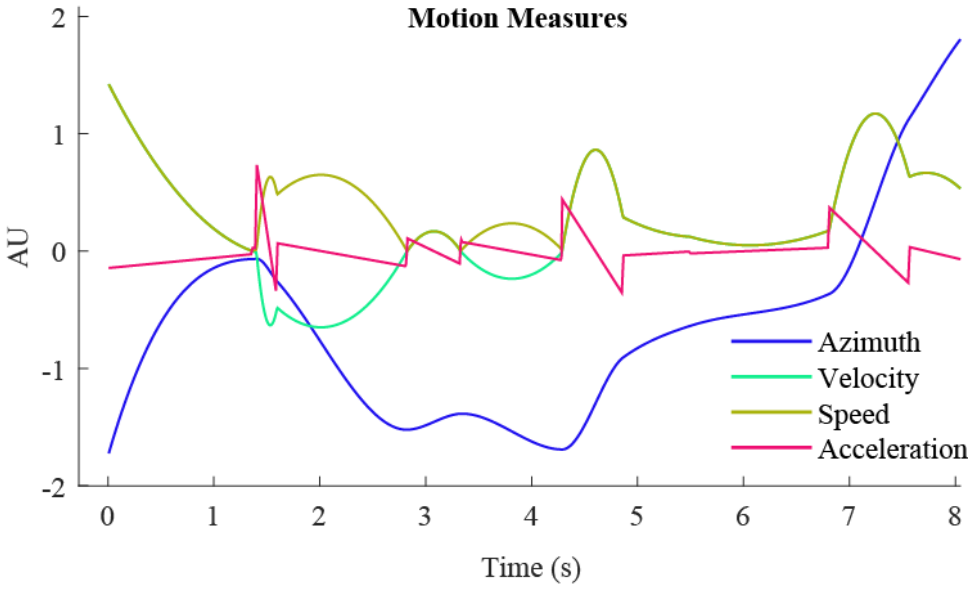



*Figure 5-1. Relationship between sound stimulus trajectory azimuth, velocity, speed and acceleration shown for ten seconds of an example stimulus. All measures were normalized to allow for comparison.*

***EEG data preprocessing.*** The EEG data were filtered between 0.5 and 30 Hz and downsampled to 64 Hz. We also ran the analysis after calculating the analytic envelope of the EEG signal using the Hilbert transformation.

***Backward modelling.*** First, we attempted to use the backward modelling approach to

reconstruct different stimulus motion representations from EEG. The methodology was similar to the stimulus trajectory reconstruction framework that was used in our previous studies (e.g. see chapter 4.2). However, instead of azimuth, here we attempted to reconstruct other representations of motion, namely, velocity, speed and acceleration.

The azimuth and motion measures, velocity, speed and acceleration were relatively highly correlated between each other (see figure 5-2 for cross-correlation plots). Therefore, it was important to control for this in our analysis and investigate these measures in isolation. We did so by repeating the reconstruction analysis for each of the measures while regressing out the contribution from the other motion measures on the EEG data. Specifically, we used a linear predictive (forward) models (see Crosse et al., 2016): We first trained a forward model to predict EEG from the stimulus motion measure we wanted to partial-out. Next, we subtracted the EEG predicted by that stimulus measure from the original EEG signal and then we fed the residual EEG data into our stimulus reconstruction analysis.

***Forward modelling.*** In addition to the reconstruction analysis above, we also used EEG prediction approach (forward modelling) to validate our finding that speed and acceleration are independently represented in the EEG data. To do this, we built three different models that predicted EEG from sound stimuli based on: (1) azimuth, (2) combination of azimuth and speed, and (3) combination of azimuth, speed and acceleration. We reasoned that if speed and acceleration are represented independently from sound azimuth, we should achieve better prediction accuracies of EEG when combining these measures in comparison to using azimuth only.

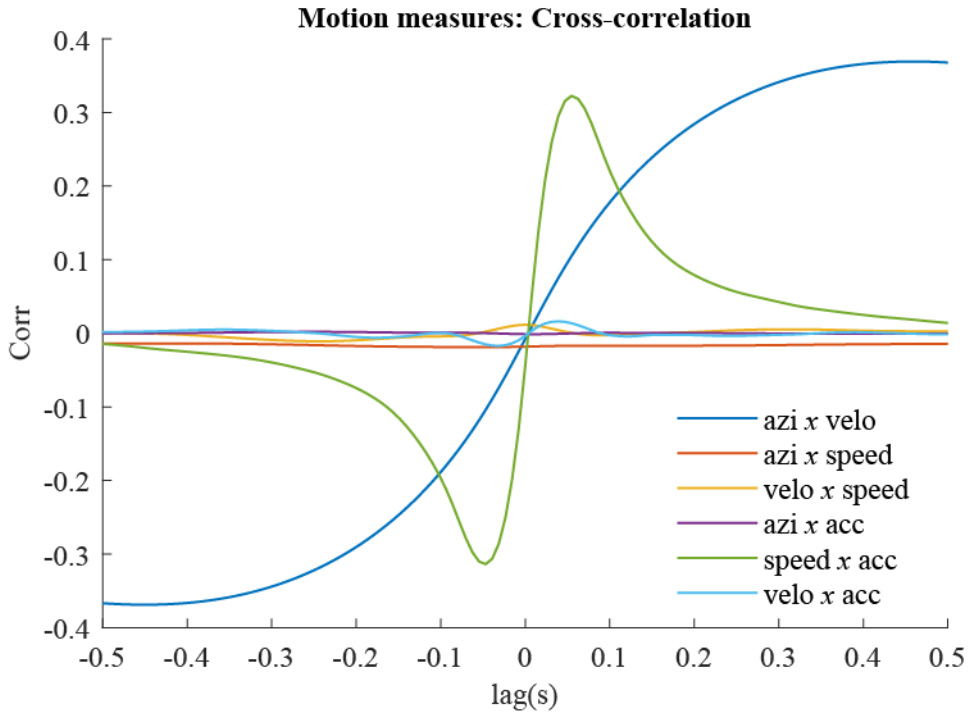



*Figure 5-2. Cross-correlation plots between azimuth, velocity, speed and acceleration. The plot shows the average cross-correlation between motion measures across all trials.*

## 5.3. Results

***Backward modelling.*** First, we attempted to reconstruct the sound azimuth and motion measures, namely, velocity, speed and acceleration, from the EEG signal when no variable was partialled-out. The reconstruction correlation values are summarized in table 5-1 and plotted in figure 5-3 along with the corresponding model activation patterns. All motion measures were reconstructed with accuracy better than the chance level from EEG as well as EEG power except for acceleration, which was successfully reconstructed only from EEG but not from EEG power.

*Table 5-1. Reconstruction accuracies for different motion measures when all measures were present (not-partialled out) in the EEG signal. Upper number in each cell represents Pearson's r correlation between the reconstructed and the actual stimulus and lower number in brackets is the corresponding p-value. Correlation values that are significantly above the chance level (p<0.05) are shaded grey.*

| | Azimuth | Velocity | Speed | Acceleration |
|---|---|---|---|---|
| EEG | 0.0243 (<0.001) | 0.0391 (<0.001) | 0.0309 (<0.001) | 0.0276 (<0.001) |
| EEG Power | 0.0328 (0.001) | 0.0131 (0.003) | 0.0271 (0.011) | 0.0106 (0.103) |

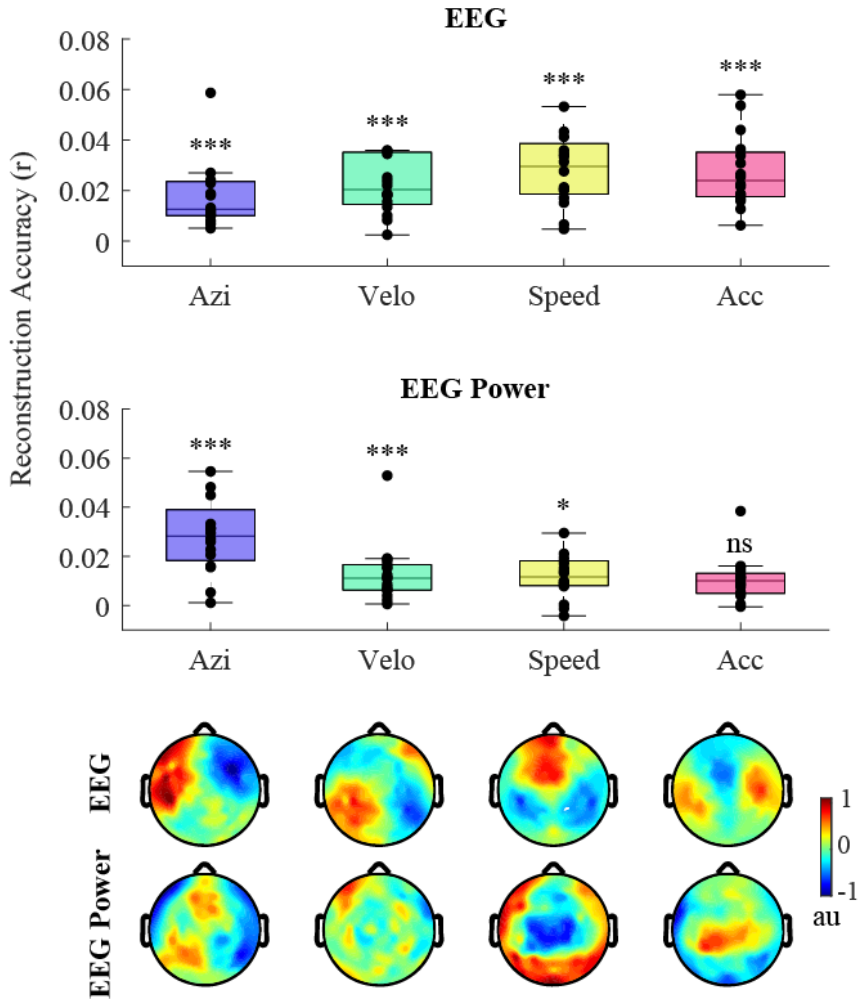



*Figure 5-3. Reconstructing azimuth, velocity, speed and acceleration from the EEG data. Upper panel shows boxplot of average reconstruction accuracies for each motion measure. Lower panel shows corresponding model activation patterns, which were averaged across time-lags 0-250 ms. * ** *** indicates prediction differences at the level of p < 0.05, p < 0.01 and p < 0.005 respectively.*

Then we repeated the analysis after controlling for the correlation between the motion measures (see cross-correlation plot in figure 5-2 above). When we partialled-out the

stimulus azimuth from the EEG data, the reconstruction accuracies for azimuth and velocity decreased below the chance level. However, speed and acceleration were successfully reconstructed from the data. See table 5-2 and left panel in figure 5-4.

*Table 5-2. Reconstruction accuracies for different motion measures when we partialled-out sound azimuth from the EEG data. Upper number in each cell represents Pearson's r correlation between the reconstructed and the actual stimulus and lower number in brackets is a corresponding p-value. Correlation values that are significantly above the chance level (p<0.05) are shaded grey. Red text indicates the partialled-out variable.*

| | Azimuth | Velocity | Speed | Acceleration |
|---|---|---|---|---|
| EEG | 0.0001 (1.000) | 0.0046 (0.982) | 0.0255 (<0.001) | 0.0193 (<0.001) |
| EEG Power | 0.0003 (1.000) | 0.0038 (0.997) | 0.0274 (0.007) | 0.0065 (0.009) |

Finally, we also regressed-out all measures except acceleration i.e. azimuth, velocity and speed from the EEG data. As expected, the backward regression modelling revealed that azimuth, velocity and speed could not be reconstructed from the EEG data. However, we were able to reconstruct sound source acceleration with decoding accuracy above the chance level. See table 5-3.

*Table 5-3. Reconstruction accuracies for different motion measures when we partialled-out sound azimuth, speed and velocity from the EEG data. Upper number in each cell represents Pearson's r correlation between the reconstructed and the actual stimulus and lower number in brackets is a corresponding p-value. Correlation values that are significantly above the chance level (p<0.05) are shaded grey. Red text indicates partialled-out variables.*

| | Azimuth | Velocity | Speed | Acceleration |
|---|---|---|---|---|
| EEG | 0.0000 (1.000) | 0.0022 (0.997) | 0.0030 (0.992) | 0.0074 (<0.001) |
| EEG Power | 0.0000 (1.000) | 0.0005 (1.000) | 0.0046 (0.997) | 0.0086 (0.001) |

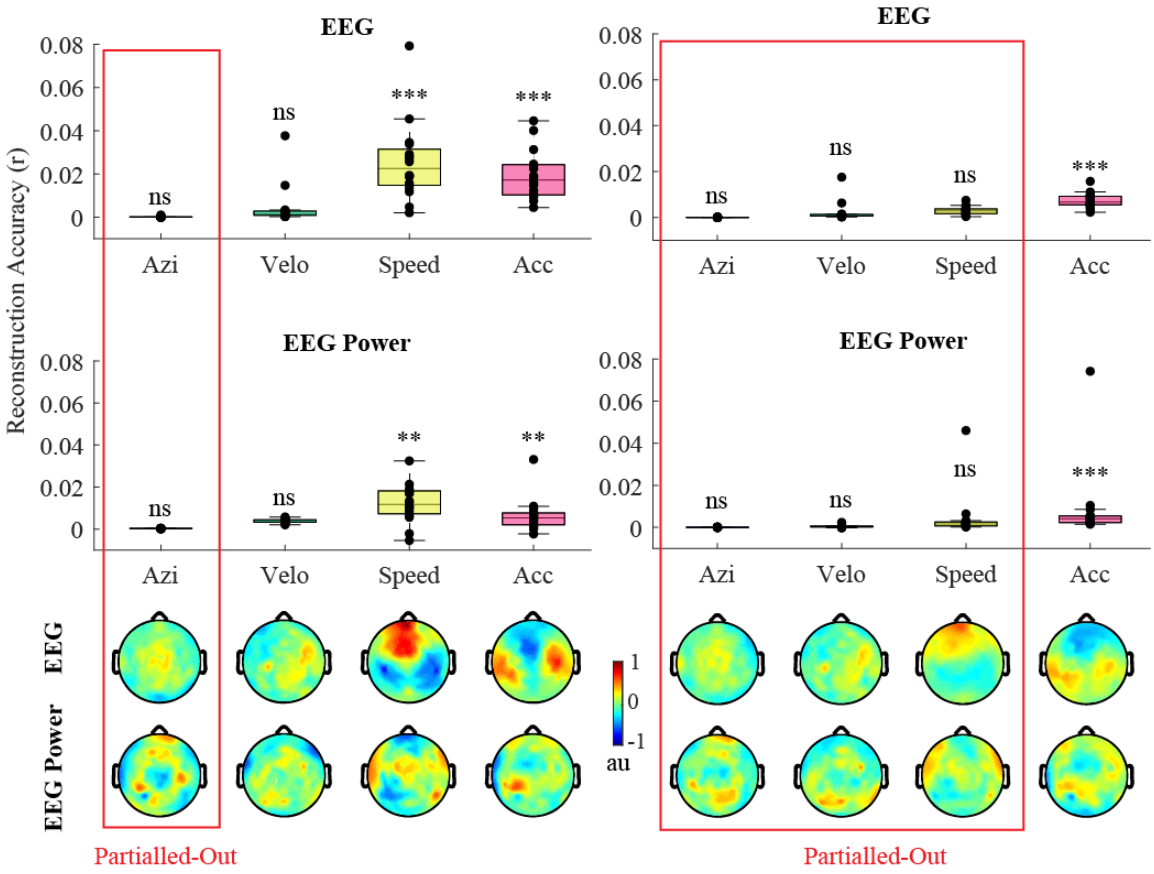



*Figure 5-4. Reconstructing azimuth, velocity, speed and acceleration from the EEG data when we controlled for azimuth as well as azimuth, velocity and speed respectively. Upper panel shows boxplot of average reconstruction accuracies for each motion measure. Lower panel shows corresponding model activation patterns, which were averaged across time-lags. * ** *** indicates prediction differences at the level of $p < 0.05$, $p < 0.01$ and $p < 0.005$ respectively.*

Since we were successful in reconstructing the sound speed and acceleration independently even when controlling for other variables, we decided to run an additional analysis in order to obtain more information about the spatio-temporal properties of our speed and acceleration reconstruction models.

We did that by running our decoders at individual time-lags between the stimulus and the EEG signal. See figure 5-5 below. For EEG, the plot of correlation values against time-lags showed that the acceleration, peaking at 160 ms, is represented in EEG at a shorter latency than the speed with peak at around 250 ms. For EEG power, this pattern was similar, however both speed and acceleration peaked later than for raw EEG. Specifically, the acceleration was best reconstructed at a time-lag of 190 ms and speed at 375 ms.

With respect to the activation patterns of the speed reconstruction model, for raw EEG, the topography at around 200ms indicated negative activations around the temporal areas and positive weights centrally. Later, at around 600 ms the polarity of this pattern reversed. For

the EEG power models, we observed prominent negative central activation and positive weight that were distributed mostly in the posterior regions. This pattern was stable activation across the range of time lags.

For acceleration reconstruction models, the activation pattern corresponding to the peak at 200 ms showed negative weights in the central area and positive activation in the temporal and occipital regions. At 400 ms, the polarity of this activation pattern reversed. The EEG power model showed positive central and negative posterior activation. This pattern was similar across wide range of time-lags, however, it was most prominent at latency of 200 ms, which corresponded to the time-lag of the highest acceleration reconstruction accuracy.

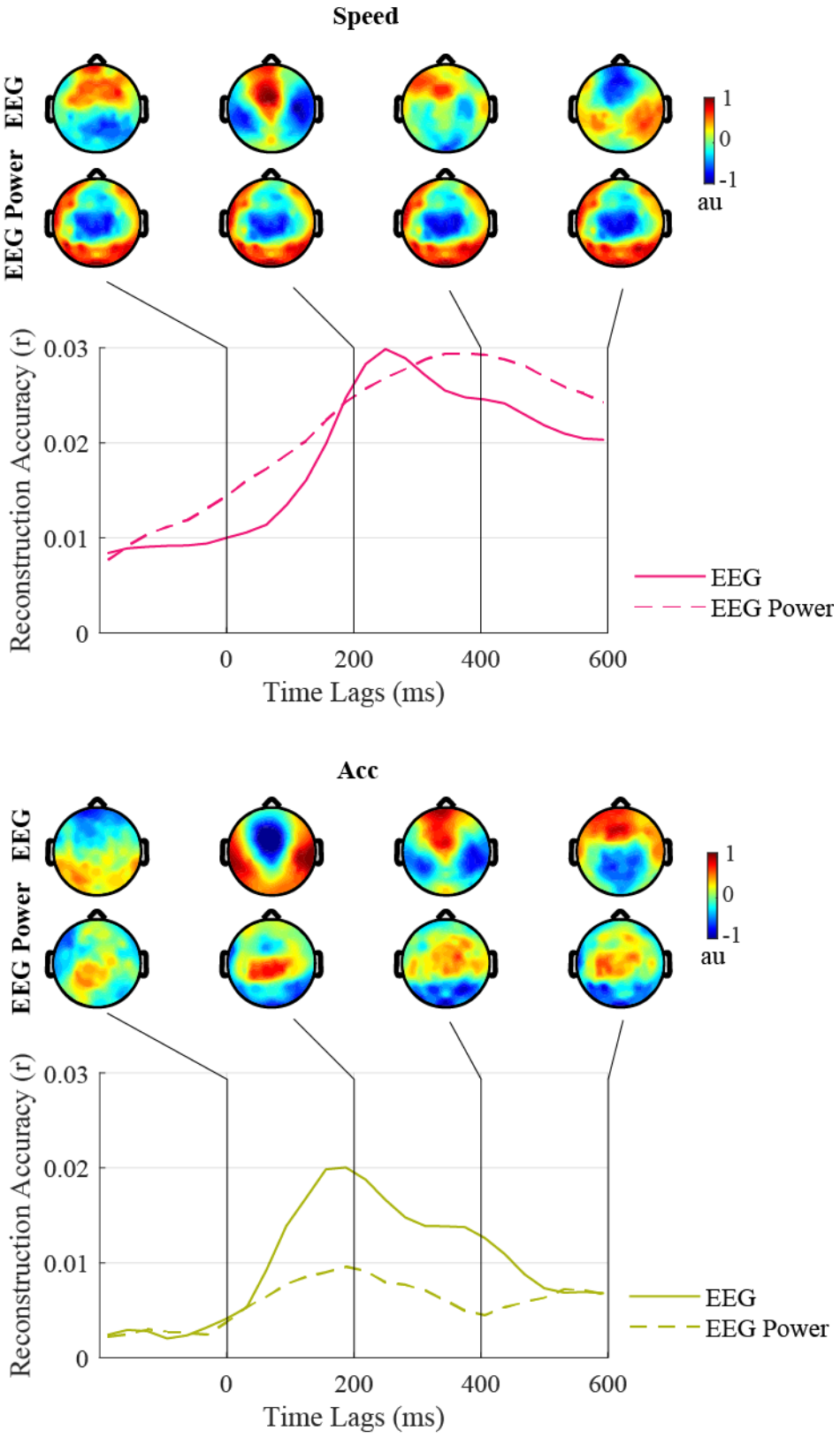



*Figure 5-5. Reconstructing sound speed and acceleration at single time-lags. The decoders performances were independently evaluated at single time-lags between the EEG and the stimulus. This was done for both velocity and acceleration using EEG as well as EEG power.*

***Forward modelling.*** In addition to our stimulus reconstruction analysis (backward modelling), we used an EEG prediction (forward modelling) approach to confirm our

previous results i.e. to test whether the velocity and acceleration are represented at the level of cortex. Specifically, we attempted to predict the EEG using the following representations of sound stimulus motion: (1) azimuth, (2) azimuth and speed "azimuth+speed" (3) azimuth, speed and acceleration "azimuth+speed+acceleration". We reasoned that if speed is represented independently from sound azimuth, we should be able to reach better prediction accuracies of EEG when combining these measures in comparison to using azimuth only. Similarly, we would expect superior prediction accuracies when predicting the EEG using all three motion measures. Again, this was done for raw EEG as well as EEG signal power.

See figure 5-6 below for the forward modelling results. The results showed that we can predict EEG from azimuth with average prediction accuracy of r=0.0206, p=1.9e-4. When using the combination of azimuth+speed, the prediction accuracy was r=0.0231, p=2.3e-4. Finally, combining all three measures, velocity+speed+acceleration, led to a prediction accuracy of r=0.0248, p=2.3e-4. As expected, for raw EEG, the combination of azimuth+speed was better than azimuth alone (p=0.004). And the combination of azimuth+velocity+speed gave superior EEG prediction accuracy to azimuth alone as well as to the combination of azimuth+speed (both p=0.001).

We also predicted EEG signal power from the motion measures. The prediction values were r=0.0075, p=0.16 for azimuth, r=0.0074, p=0.001 for azimuth+speed and r=0.0085, p=0.0023 for azimuth+speed+acceleration respectively. The comparison between the models showed there are significant differences between the azi+speed and azi+speed+acceleration models (p=0.004). However, the azimuth+speed and azimuth+speed+acceleration were not significantly better than the azimuth alone. See figure 5-6.

Finally, we considered the possibility that the prediction accuracies of the combined models (azimuth+speed and azimuth+speed+acceleration) were better than the azimuth-only model due to the larger dimensionality of the combined models. To test this, we ran the EEG prediction using the combined models while permuting the speed and acceleration features across trials. Thus, we built models with the same dimensionality but the features (speed and acceleration) were not informative. We found that the accuracies of these permuted models were not significantly different from the performance of the azimuth-only model (all p>0.05), which indicates that the superior performances of the combined models were not due to their larger dimensionality.

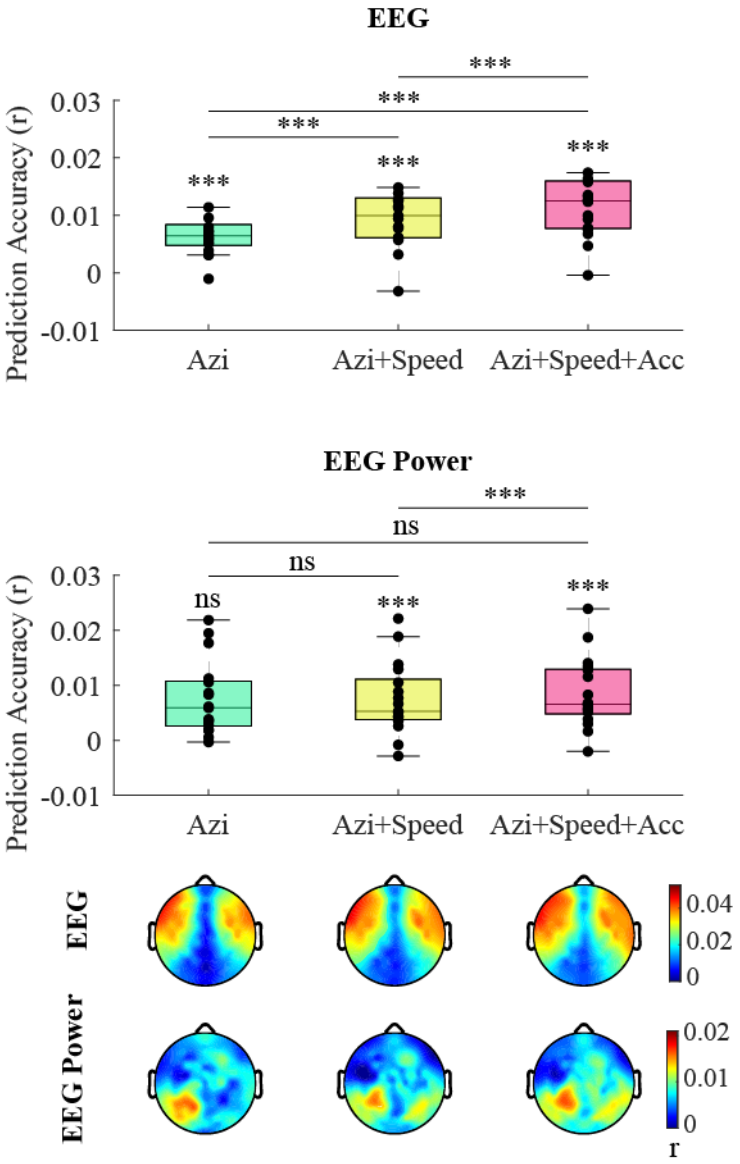



*Figure 5-6. Predicting EEG from azimuth, azimuth+speed and azimuth+speed+acceleration. The upper two panels show boxplots EEG and EEG power prediction correlation values for different motion measures. * ** *** indicates prediction differences at the level of $p < 0.05$, $p < 0.01$ and $p < 0.005$ respectively. At the bottom, the topographic plots show the prediction correlation values on the scalp surface.*

## 5.4. Discussion

Previously, we have shown that the sound stimuli azimuth is represented in the EEG. Here, in this study, we aimed to test whether the cortex also represents higher order motion measures, namely velocity, speed and acceleration.

**Sound source speed is encoded in the cortex independently on sound source azimuth.** We found that stimulus speed can be successfully reconstructed from the raw broadband EEG as well as from the EEG power. This was true even when we the sound azimuth was partialled-out from the EEG signal, which indicates that sound stimulus speed is cortically encoded independently from sound azimuth. Similarly, using a forward encoding model, we

found that we are able to better estimate EEG from the combination of azimuth and speed than from azimuth only time-series. This further demonstrates that the cortex reflects not only the moving sound source trajectory but is also sensitive to the speed of the sound.

These results are in line with other neurophysiological studies that have shown cortical sensitivity to velocity using short discrete stimuli (Makela & McEvoy, 1996; Xiang *et al.*, 2005; Getzmann, 2009; Meng *et al.*, 2016). In particular, our findings are similar to a study by (Getzmann, 2009), who compared motion onset responses to moving stimuli of different speeds. Specifically, the topography of the cP2 component of the motion onset response looked similar to our speed reconstruction model activation patterns. This component showed bilateral negative activation in parieto-temporal regions and positive activation over the central area. Also, the cP2 waveform peaked on average around 235 ms after the motion onset. Somewhat similar to this, we found that the single-lag reconstruction performance of our model was best at a time-lag corresponding to 250 ms. Our findings are also compatible with an fMRI study by Meng *et al.* (2016), who compared slow- and fast-moving sources, and found activation in posterior superior temporal regions and premotor ventral-rostral areas.

Interestingly, we were able to reconstruct the sound source angular velocity only before we partialled-out the sound azimuth from the EEG data. In our experiment, the angular velocity was different from the speed measure only in its directionality. Therefore, we would expect that if we can reconstruct speed, it should also be possible to reconstruct velocity from EEG. We speculate that this somewhat unexpected outcome might be caused by the fact that the azimuth was highly correlated with velocity but not with speed (see figure 5-2).

Finally, although we were successful in reconstructing the sound source speed from the EEG signal power with accuracy above the chance level, the activation patterns were noisier and difficult to interpret as they did not show any consistent pattern. To our best knowledge there is no literature on sensitivity of EEG power to velocity or speed and so it is difficult to put this observation in context.

**Sound acceleration can be reconstructed from EEG independently from azimuth and speed.** Our backward mapping results showed that it is possible to reconstruct sound acceleration from EEG with above-chance accuracy even when we control for all other motion measures, namely azimuth, velocity and speed. Again, this was also confirmed by our forward modelling, which showed that EEG prediction accuracy is the best when we include sound acceleration as a predictor (along with azimuth and speed).

As discussed above, psychophysics have shown that we are able to discriminate between

accelerating and decelerating sounds (Perrott *et al.*, 1993). Nevertheless, acceleration sensitivity of auditory neurons has not been demonstrated and to our best knowledge no other neurophysiological study has shown neural sensitivity to sound acceleration and it is likely that sound acceleration estimation is a higher-level process based on already extracted sound source location data (Carlile & Leung, 2016).

## 5.5. Conclusion

Here, using continuous moving stimuli, we showed that the cortex is sensitive to the speed of sound stimuli. Also, using both forward and backward mapping, we showed that the cortical representation of sound source speed is at least partially segregated from the representation of sound source trajectory. Similarly, we also showed that the sound acceleration can be decoded from the EEG independently on other measures. However, given the lack of other evidence about cortical sensitivity to acceleration makes it difficult to put too much weighting on this finding.

# Chapter 6. Decoding Sound Trajectory during Free Head Movement with VR

## 6.1. Introduction

The vast majority of studies have investigated spatial hearing using a fixed-head listening scenario, where listeners have been unable to move within the environment. However, this situation is highly unnatural and the inability to employ head movements causes sound localization deficits such as front-back confusion. See section 2.1. (Wallach, 1940; Wightman & Kistler, 1999).

In humans, the ears have a fixed position on the head and so the spatial acoustic cues i.e. interaural level and time differences (ILD and ITD) as well as monaural cues are represented relative to the cranio-centric (head-centric) coordinate frame. When we move our head or move within the environment, the sound position relative to our head changes in the opposite direction to our motion. Therefore, in order to localize a sound source with respect to our environment i.e. using a world-centric (allo-centric) coordinate frame and to keep a stable perception of the acoustic space, we need to integrate the information from our proprioceptive system to compensate for our movement. However, exactly how this transformation process is performed by the auditory pathway and whether the sound location is cortically represented using both cranio- and allo-centric coordinate frames is unclear.

One of the main reasons for this is methodological. In order to investigate this, the cranio-centric and allo-centric reference frames needs to be experimentally dissociated, which requires listeners to move or rotate within the environment. Unfortunately, fMRI and MEG imaging techniques require subjects to be stationary and although it is theoretically possible to use EEG, complications arise because of the electromyographic (EMG) motion artefacts e.g. see figure 1 in Altmann *et al.* (2009). As result there is only limited research on this topic. As we discussed in chapter 2, two EEG studies that have investigated auditory processing in experiments involving head rotation found that only cranio-centric and not allo-centric representations of location exist within human auditory cortex (Altmann *et al.*, 2009; Altmann *et al.*, 2012). In contrast, another EEG study demonstrated that both cranio- and allo-centric reference frames are represented within the auditory cortex (Schechtman *et al.*, 2012). As well as providing conflicting results, these studies were not based on direct EEG measures of sound source location per se, but rather reflected neural responses to unexpected location changes.

Some work has been done on directly measuring cortical representation of both spatial

coordinate frames in a recent ferret study (Town *et al.*, 2017), which tested spatial neural sensitivity in animals that were free to move during the recordings. This study found that the majority of neurons in auditory cortex encode sound location cranio-centrically, and a small number of neurons were sensitive to sound location in allocentric terms.

Here, we aimed to investigate whether we encode sound location using cranio- and/or allo-centric coordinates. This has been tested by reconstructing the sound source position with respect to both cranio- and allo-centric frames of reference in a situation where listeners move their head. If we could successfully reconstruct the sound position in given coordinate frame, it would indicate that the cortex encodes sound location using this frame of reference. As we mention above, the head motion was implemented as it is required to dissociate the coordinate frames.

For this task we decided to employ the virtual reality (VR) headset OculusVR CV1 for stimulus presentation. This was used as it has a number of advantages over headphone or free-field (loudspeaker array) presentation: Firstly, the Oculus VR facilitates head position tracking, as it has low latency and high precision rotational and head tracking system so no external head-tracking system was needed. Second benefit is that the system automatically adjusts (compensates) the audio for head motion i.e. the perceived sound source location is unchanged with respect to the allocentric coordinates despite the user's head motion. This technical feature was required in order to keep our experiment as naturalistic as possible. Note that the audio compensation for head movement is also possible to implement using standard headphones or loudspeaker arrays, however, it is technically very challenging. In addition, we were interested to test the combination of VR stimulus presentation and EEG recording in practice as consumer grade VR use for neuroscientific experiments is not very common and we had no previous hands on experience with it.

## 6.2. Material and Methods

***Participants.*** In total eleven participants (median = 26 years; min = 20 years; max = 34 years; 4 females; 8 right handed) took part in this study with informed consent. All subjects reported no neurological diseases and normal hearing. The experiments were approved by the Research Subjects Review Board of University of Rochester.

***Experimental procedure.*** Participants wore an OculusVR virtual reality headset (version CV1, https://www.oculus.com/rift/) and listened to spatial auditory stimuli presented via the embedded headphones while we recorded their EEG (see figure 6-1A). During the

experiment, the subject wearing the headset sat on a fixed desk chair in a dark soundproof room. The room was equipped with head tracking sensors of the OculusVR set and the subject's position was recorded during the experiment.

The experimental virtual environment was custom-developed using a Unity 3D development platform (https://unity.com). The virtual environment consisted of a rectangular floor with green tiles and a clear blue skybox. The subject's avatar was placed in the middle of the floor with a camera (avatar's head) at a distance from the floor that corresponded to a sitting position (see figure 6-1B). The position of the virtual camera was linked to the position of the subject's head in the real-world i.e. when the subject moved his or her head, the virtual camera performed the corresponding motion in the virtual world.

There were two experimental conditions: (1) static head+moving sound and (2) moving head+ moving sound. There were 20 trials in each experimental condition, each trial lasting 60 s.

***Condition 1.*** In the first condition, the subjects listened to moving stimuli while they were asked not to move their heads and look at the fixation cross. The fixation cross was shown in the middle of their field of view.

The sound stimuli consisted of continuous pink noise, which was spatialized so that the sound was perceived to be pseudo randomly moving on a semi-circular trajectory in a horizontal plane between -90˚ (left) and +90˚ (right) relative to the subject. The sound trajectory was pseudo-random, simulating smooth but unpredictable sound movement. The sound source motion was controlled by the Unity Engine and Oculus Audio SDK. This condition was similar to our previous experiment, which was described in chapter 3- section 3.2.

***Condition 2.*** In the second condition, the subjects listened to the same spatial sound stimuli as in condition 1, however, they were asked to move their heads at the same time. The head-motion was indicated by a left- or right-pointing arrow which was shown in the middle of the field of view instead of the fixation cross (see 6-1B). For each time-frame, an arrow indicated the subject to move his or her head towards the desired position according to the predefined head trajectory. The head-trajectory was predefined in a similar way as the sound trajectory i.e. smooth random motion within the horizontal plane between -90 and +90 degrees. An example of a predefined and measured head trajectory is shown in figure 6-1C. As expected, the subjects responded to the arrows that indicated head motion with a delay. Figure 6-1D shows a cross-correlation plot between the predefined and measured head trajectory. As indicated by the peak of the cross-correlation plot, the measured trajectory was

lagging the predefined trajectory by 810ms. Also, as result of the head-movement, the sound position with respect to allo-centric and cranio-centric coordinate frames was dissociated. See figure 6-2 for an example of sound trajectory represented using both coordinate frames.

***Behaviour task.*** In both experimental conditions, the subjects performed a simple target detection task. The task required subjects to respond with a remote control button press to infrequent tremolo targets (modulation frequency 4 Hz, 2 s long), which were embedded in the auditory stimuli. The number of targets within each trial ranged from 1 to 4 per trial.

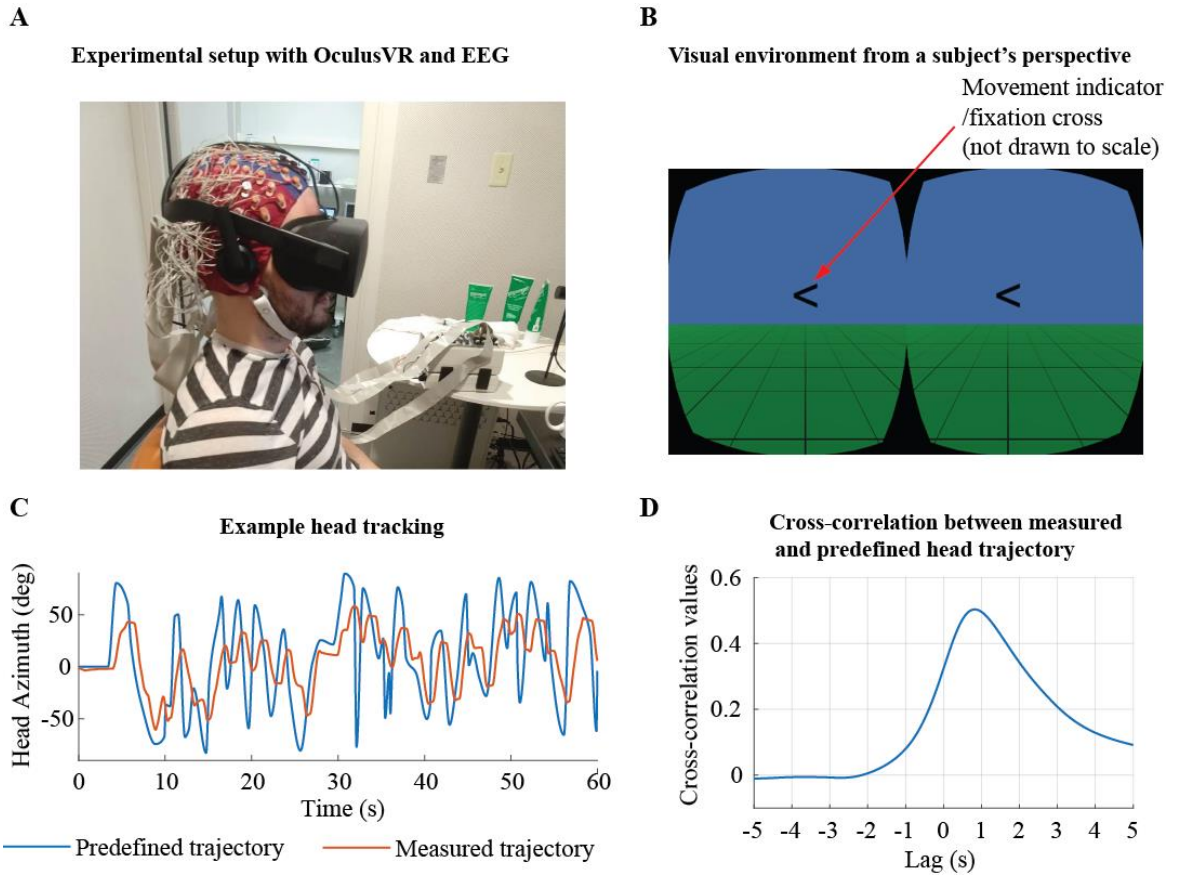



*Figure 6-1. Experimental setup with OculusVR virtual reality headset. (A)The subjects listened to spatialized stimuli over headset's embedded headphones while we recorded their EEG using a 128-channel ActiView Biosemi acquisition system. (B) Subjects were asked to move their head as indicated by an arrow which was presented in the centre of the field of view. (C) Example of predefined (as indicated by the changing arrow) and measured head trajectory shown for a selected subject. (D) Cross-correlation plot between the predefined trajectory and measured trajectory shown as grand-average across all subjects. On average, the measured trajectory was lagging the predefined trajectory by 810 ms.*

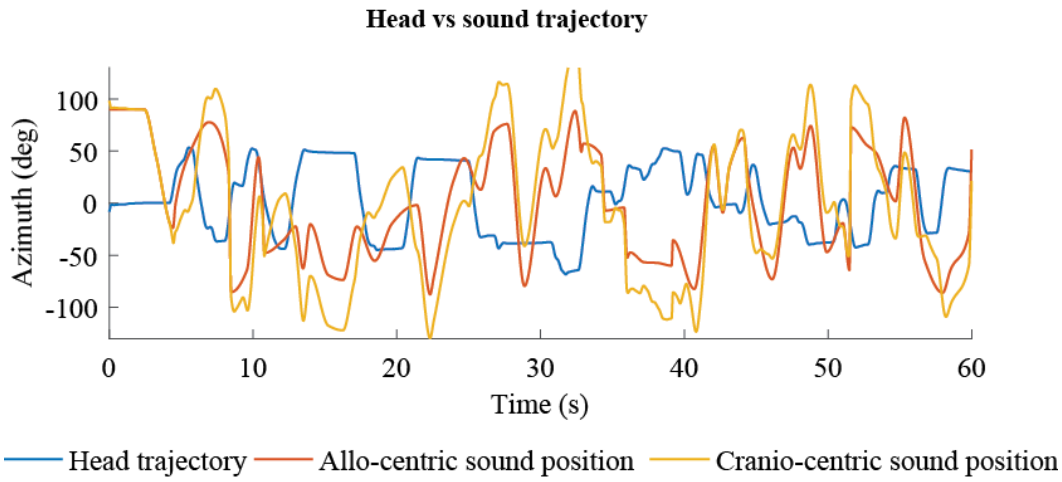



*Figure 6-2. Example of head trajectory and sound trajectory represented in allocentric and egocentric coordinate frames. In experimental condition 2, the subjects were asked to move their head as indicated while they listened to the moving sound stimuli in the same time. The allo-centric coordinates represent location of the sound with respect to the environment and the cranio-centric coordinates with respect to the person's head.*

***EEG data preprocessing and analysis.*** The EEG data were pre-processed using the same pipeline as our previous experiment- see section 3.2.

For the first experimental condition, which required subjects to keep their head still as they listened to a single moving sound source, we used the stimulus trajectory reconstruction approach as described in chapter 3. Briefly, we trained a backward model to reconstruct the sound trajectory from the EEG data.

In the second experimental condition subjects listened to the spatial stimuli while they moved their heads. In order to control for the muscle artefacts caused by the head motion in the EEG signal, we partialled-out the recorded head trajectory from the EEG data. To do so, we trained a forward model (see section 2.3.4 for details) to predict the EEG signal from the recorded head trajectory. As the head trajectory and sound trajectory were independent, this predicted EEG signal contained only information that was related to head motion itself i.e. mainly EMG activity caused by the neck muscles. Next, we obtained the residual EEG by subtracting this predicted EEG signal from the original signal. Afterwards we ran our sound trajectory reconstruction on this "clean" EEG as in the first experimental condition.

## 6.3. Results

***Behaviour task.*** In both experimental conditions, the subjects performed a tremolo target detection task. For the first experimental condition, where subjects kept their heads static the mean average detection score was 91.82±1.54% (SE). For the second condition that required head movement, the targets were successfully detected with accuracy of 88.64±2.10% (SE).

A pairwise comparison using Wilcoxon signed rank test between the experimental conditions revealed no significant differences (p=0.353).

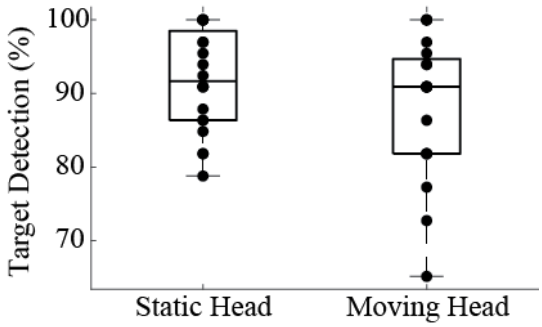



*Figure 6-3. Behaviour results. During both experimental condition, the subjects responded with a button press to a tremolo targets, which were embedded in the audio. No significant differences were found between the experimental conditions.*

***Trajectory reconstruction analysis.*** For the first experimental condition (static head), the Delta decoder mean reconstruction accuracy was r=0.037, p=0.007. For the Alpha Pwr decoder, the reconstruction accuracy was r=0.026 and was not significantly above the chance level (p=0.18). See left panel in figure 6-4.

For the second condition (moving head), we first ran the sound trajectory reconstruction analysis without controlling for EMG artefacts caused by the head rotations. This was done for sound source location represented using cranio- as well as allo-centric coordinate frames.

The results showed that only the cranio-centric representation can be reconstructed significantly above the chance level from EEG for both Delta and Alpha Pwr decoders. Specifically, for the Delta decoder, the allocentric and cranio-centric sound trajectories were reconstructed with average accuracies of r=0.037, p=0.38 and r=0.097, p=0.002 respectively. For the Alpha Pwr decoder, the average reconstruction accuracies were r=0.019, p=0.68 for allocentric and r=0.12, p=0.001 for cranio-centric representation.

In addition, to get an idea of how strongly represented the motion-related activity in the EEG signal was, we tried to reconstruct the head trajectory (rotation angle time-series) from the EEG data. For the Delta decoder, the average head trajectory reconstruction was r=0.28, p=5e-4. For the Alpha Pwr decoder, the reconstruction accuracy was r=0.27, p=1e-3. See right panel in figure 6-4.

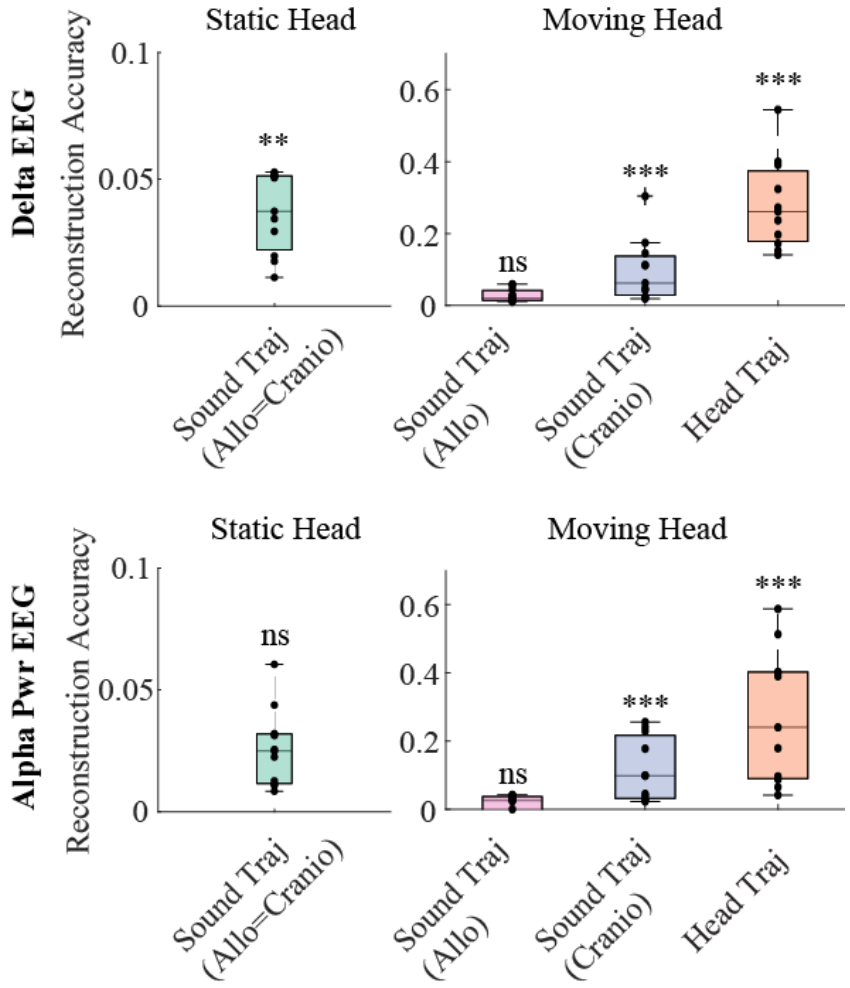



*Figure 6-4. Reconstructing sound and head trajectory from the EEG data. First boxplot (green) shows sound trajectory reconstruction accuracy for experimental condition 1 (static head). In this case the allo-centric and ego-centric coordinate frames were aligned. Second and third box (pink and blue) show sound trajectory reconstruction values from experimental condition 2 (moving head) with respect to allo-centric and ego-centric frames. The rightmost box (red) shows the head trajectory reconstruction accuracy for the second experimental condition.*

Importantly, the very large reconstruction values of the head trajectory from EEG show that the EMG artefacts caused by head turning were strongly represented within the EEG signal. And, as the cranio-centric sound-location is anti-correlated with head motion i.e. when we turn our head to the right, the sound source moves to the left with respect to our head, it was important to separate out the effects of sound source position and head motion on the EEG signal. Therefore, we partialled-out the head trajectory from the EEG signal and tried to reconstruct the sound source trajectory again (see methods section for details).

After we partialled-out the head trajectory from the EEG data, the sound trajectory reconstruction accuracies dropped below the chance level. Specifically, the Delta average sound reconstruction values decreased to r=0.027, p=0.71 for allo-centric and r=0.027,

p=0.86 for cranio-centric representation. For the Alpha Pwr decoder, the reconstruction accuracies were r=0.018, p=0.79 and r=0.017, p=0.86 for allo-centric and cranio-centric representations respectively. Also, as expected, the head trajectory reconstruction accuracies decreased below the chance level, showing that our partialling-out procedure was successful. See figure 6-5 for reconstruction accuracies.

Figure 6-6 shows the model activation patterns for all decoders. For the static head condition, we expected the activation patterns to be similar to those from our first study (see chapter 3). However, the Delta activation pattern was relatively noisy and did not show any consistent pattern. For the Alpha decoder, the topography looked relatively similar to our previous study, nevertheless, the reconstruction of this model was not significantly above the chance level and so one needs to be careful to interpret this. Similarly, it is difficult to comment on the decoder activation patterns corresponding to the moving head condition as the reconstruction accuracies in this condition were all not significantly above chance.

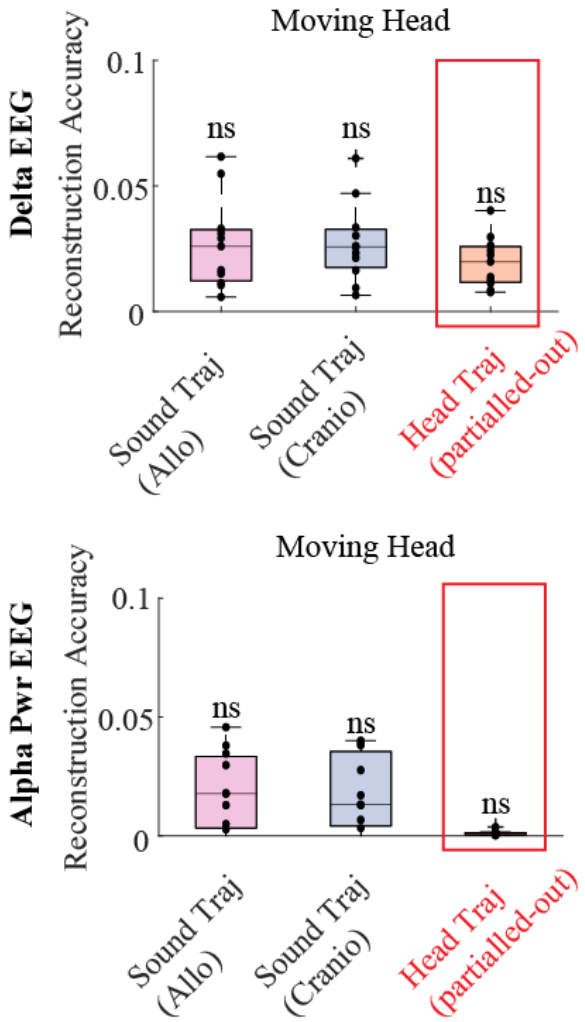


*Figure 6-5. Reconstructing sound and head trajectory after we partialled-out the head trajectory from the EEG signal. First and second box (pink and blue) show sound trajectory reconstruction values from the experimental condition 2 (moving head) with respect to allo-centric and ego-centric frames. The rightmost box (red) shows the head trajectory reconstruction accuracy for the second experimental condition after it was partialled-out.*

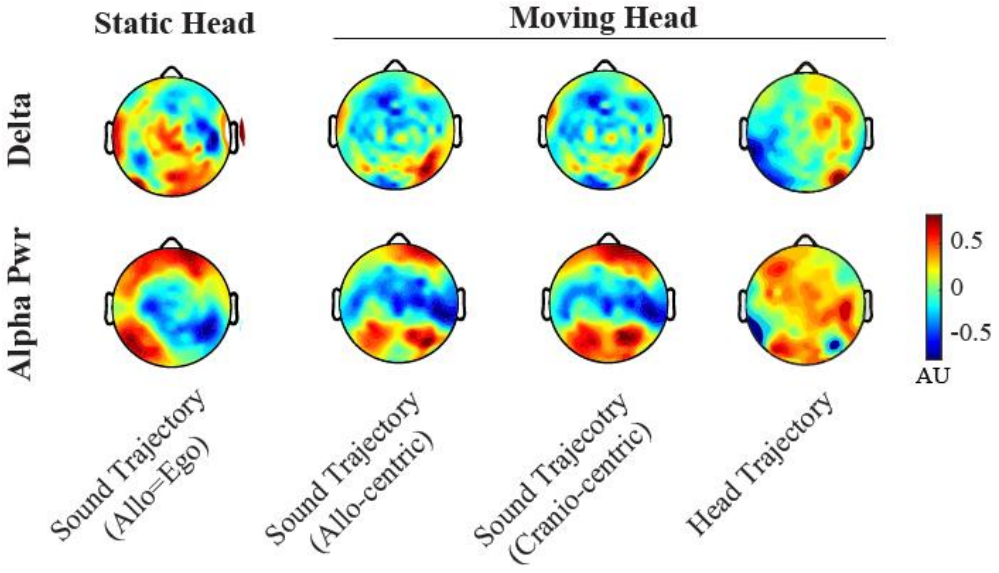



*Figure 6-6. Activation patterns corresponding to sound and head trajectory decoders shown for Delta and Alpha Pwr after we partialled-out the head trajectory. The activation patterns were averaged across subjects and individually normalized.*

## 6.4. Discussion

There were two main aims of this study. First, we were interested to see whether we can apply our trajectory reconstruction models in more naturalistic scenarios where subjects are allowed to move their heads and to test whether we can perform the reconstruction using cranio- and allo-centric representations of the sound. Second, we were interested to test the feasibility of using the Oculus Rift CV1 headset for stimulus presentation in an EEG experiment.

**The trajectory reconstruction in the static head condition was successfully replicated using delta but not alpha power EEG.** In the first experimental condition, the subjects listened to a single moving sound moving sound and were instructed to not move their heads. By doing this, we aimed to replicate our previous study from chapter 3 and validate that we can use the OculusVR hardware for stimulus presentation in our experimental scenario.

Interestingly, we found that we can reconstruct the sound trajectory significantly above the chance level using the Delta but not the Alpha Pwr decoder. This is in contrast with our

results from chapter 3, which showed that both of these EEG components can be used to reconstruct sound trajectory.

In the current and the previous study, the sound stimuli were spatialized using the same method (Oculus Audio SDK). However, there were several differences in other aspects of the experimental design, which could explain the overall lower reconstruction accuracies and below-chance level performance of the Alpha decoder in the current study. The most obvious was the difference in the received visual input. Specifically, in our previous study, subjects were sitting in a dimly lit room and were asked to look at a white fixation cross in the centre of a black computer screen. In contrast, the subjects in the current study were in the VR environment that was composed of a simple visual scene with a green rectangular floor and a blue skybox (see 6-1B), which was done to make the experiment more naturalistic and to make subjects feel immersed within the VR environment. This could result in the cortical responses in the current experiment being dominated by the visual input or more trivially, it is possible that that the subjects paid less attention to the auditory modality.

A second factor that could contribute to this is that the current VR experiment had a shorter duration and so we collected less data than in our previous study. This could potentially cause a suboptimal fitting of the reconstruction models and consequently lower reconstruction accuracies.

Finally, it is possible that the decreased reconstruction accuracy of the Alpha decoder was caused by the problematic head attachment of the OculusVR. Specifically, the attachment strap of the headset applies relatively large pressure on EEG electrodes around the parietal-occipital region of the scalp, which is the area that contributes most to the Alpha decoder. Therefore, it is possible that this led to a decreased signal-to-noise ratio of EEG within this area and subsequently lower reconstruction accuracies of the decoder.

**Head movement substantially impairs sound location decoding with EEG.** In the second condition, we attempted to reconstruct the sound source trajectory when subjects moved their heads using both head- and allo-centric representations of the sound trajectory.

First, we ran the reconstructions on EEG without any compensation for motion artefacts and found that we can successfully decode the sound source trajectory in cranio-centric but not in allo-centric representation. This finding would be in line with the fact that the spatial acoustic cues are defined with respect to our head and previous EEG research that suggested that cortical encoding of sound location has cranio-centric rather than allocentric representation (Altmann *et al.*, 2009; Altmann *et al.*, 2012).

However, as the head rotation proportionally changes cranio-centric location of a sound source, it was important to remove the motion artefacts from the EEG data. Without controlling for this, our above-chance reconstruction accuracy could be possibly driven by head-movement artefacts within the EEG data. Note that we tried to reconstruct head rotation time-series from EEG and obtained correlation values that were approximately ten times larger than for sound source trajectory reconstructions (see figure 6-4). Therefore, in our next step, we partialled-out the head-movement trajectory from the EEG data, which effectively removed movement-related artefacts from the EEG signal. We then found that the reconstruction accuracies for both cranio- and allo-centric models were below the chance level. This shows that we are unable to reconstruct sound azimuth from the EEG data independently of the head movement and consequently makes it difficult to draw a conclusion on whether it is possible to reconstruct sound trajectory from the neural activity when subjects move their heads.

**Is it feasible to use OculusVR version CV1 in EEG experiments?** As this was our first practical experience with OculusVR in an EEG experiment and there are not many other neurophysiological studies employing this hardware we encountered several difficulties:

First, as we already mentioned, even when fully elongated the attachment strap that keeps the relatively heavy headset on the head, presses against the EEG electrodes at the back of the scalp. This can cause potential issues with the EEG signal quality and it can cause discomfort to users that wear the headset over long periods of time, which is a limiting factor for EEG studies. One solution to this issue would be to use a different EEG acquisition system with more compact electrodes or to use a reduced electrode montage i.e. leave out the electrodes that would be in contact with the VR headset. The other option would be to make a custom attachment strap for the headset, however, this would require a relatively complicated hardware modification as the strap is used to deliver the sound signal to the headphones.

Next, there is the issue of synchronizing the audio stimulus presentation with the EEG signal. As the VR environment needs to compensate for user's movement and behaviour, it has to render the audio and visual components of the scene in real-time. Consequently, there is some inherent latency and latency jitter in stimulus presentation. This is in contrast with traditional "offline" experimental stimulus delivery systems, which can preload the stimuli prior to their presentation and then present them at a desired latency with relatively large accuracy and precision. Here, we overcame this issue by designing custom electronic circuitry that measured the electrical signal feeding the headphones and used this signal to

send a trigger to the EEG acquisition system with the latency of actual sound stimulus onset. Although a similar approach could be theoretically adopted for video-EEG synchronization, we have not implemented this in our study.

## 6.5. Conclusion

Here, we tried to reconstruct moving sound trajectory using both cranio- and allo-centric coordinates from EEG while subjects listened to a moving sound and rotated their heads at the same time independently on the sound trajectory.

We found that we were unable to reconstruct the allo-centrically represented sound trajectory from the EEG. With respect to cranio-centric representation, we were only successful in reconstructing the sound source trajectory when not controlling for motion artefacts. The subsequent analysis showed that the reconstruction accuracy drops below the chance level when we partialled-out the head trajectory. Therefore we cannot confirm whether the neural encoding of sound location is represented using a cranio-centric representation.

Our secondary goal was to test the feasibility of using Oculus Rift VR headset for stimulus presentation together with high-density EEG recording. Although we found it is theoretically possible to do this, we encountered several practical issues, for example, subject discomfort when the headset was worn over extended time-periods, which need to be addressed in future studies.

# Chapter 7. General Discussion

## 7.1. A Novel Approach to Study Cortical Representations of Auditory Space

In this thesis we used an analysis framework based on multi-variate linear regression to study neural correlates of spatial sound processing in humans. This analysis method of relating a continuous stimulus representation to un-epoched, un-averaged neural data has been only recently adopted by neurophysiological studies and has for the most part been used for studying speech and language processing. These studies have shown that the stimulus envelope is tracked by low-frequency cortical activity in an approximately linear fashion and showed that the stimulus envelope can be reconstructed from EEG (Mesgarani *et al.*, 2009b; Ding & Simon, 2012b; Zion Golumbic *et al.*, 2013; O'Sullivan *et al.*, 2015)

Here, we demonstrated that the stimulus reconstruction framework can be successfully applied to study human spatial hearing for moving sound stimuli with a stochastic sound trajectory. Specifically, we demonstrated that the moving sound stimulus trajectory, represented as an azimuthal time-series, can be reconstructed from EEG using multivariate linear regression (see chapter 3).

In the case of envelope-tracking by the cortex, where low frequency neural activity is phase-locked to the stimulus envelope, the applicability of linear models is relatively intuitive. In contrast, the use of a linear model to map EEG to sound stimulus azimuth is less obvious. The stimulus spatial position is not explicitly available in the acoustic signal that arrives to an ear. Instead, the sound position is calculated subcortically mainly from the intensity and timing differences between the ears, ILD and ITD, and the higher-level location-sensitive neurons in auditory cortex were found to respond to a broad range of spatial angles. Most of these neurons have contralateral tuning i.e. respond mostly to stimuli from the opposite hemifield (e.g. see Fig 2-13B). This contralateral neural tuning can be reasonably approximated using a linear function and therefore can be mapped using a linear model. However, note that some spatially sensitive cortical neurons respond to the ipsilateral field and other non-linear response patterns such as tuning to several spatial locations have also been observed (Wang *et al.*, 2019). Therefore, our linear regression models reflect only part of the neural activity associated with spatial sound tracking and do not reflect spatially sensitive neurons that have firing patterns which are non-linearly dependent on sound azimuth.

Importantly, unlike the ERP analysis, the stimulus reconstruction method only reflects the

response of the neural system to a specific stimulus parameter (regressor). Therefore, using the same neural data, it is possible to independently reconstruct several different attributes of the same stimulus. This has been used in chapter 5, where we reconstructed velocity, speed and acceleration of the sound. Similarly, it allows the presentation of several sound stimuli simultaneously and analysis of the neural responses corresponding to each stimulus separately. This has been showed in chapter 4, where we presented multiple competing sources and tested whether we can decode the trajectory of both attended and unattended sound sources.

Another advantage of this methodological framework is that it allows the use of continuous stimuli, which is particularly suitable for studying dynamic aspects of auditory motion. And finally, due to its multi-variate nature (the reconstruction model integrates over a number of time-lags between EEG and stimulus as well as over EEG electrodes), it has the potential to be more sensitive than univariate measures.

## 7.2. Future work

Here we used headphone presented stimuli that were spatialized using non-individualized HRTFs. As a consequence, the stimuli contained only approximate spectral cues, which affects mainly localization in the vertical plane and causes front-back confusion. Therefore, we investigated only neural responses to horizontal sound motion, and we restricted the source motion to the frontal part of the auditory space. By using individualized HRTFs or free-field (loudspeaker array) presentation to deliver higher-fidelity spatial audio, one could test aspects of spatial hearing which we have purposely omitted. For example, it would be interesting to test what happens to the spatial coding when a source moves on a full circle (360 degrees) around the subject and to test whether in this situation we can successfully reconstruct sound azimuth from EEG. Or it would be intriguing to use our regression framework to reconstruct elevational trajectory of a sound source that moves within the vertical plane. As we described in section 2.4.3, the number of studies that investigated sound elevation processing is very small and not much is known about cortical encoding of sound elevation. Therefore, any additional research in this area would be certainly beneficial.

Another way to build on this work would be to improve the reconstruction model itself. As mentioned in the previous section, our stimulus reconstruction model can capture only neural activity that is (approximately) linearly related to the stimulus parameter of interest. Although the linearity makes the interpretation of the reconstruction models relatively easy, it is certainly suboptimal in the context of stimulus reconstruction performance. By using a more complex non-linear model, one could better fit the neural data, which would result in

higher reconstruction accuracies and potentially, a more accurate image of how auditory space is encoded within the cortex. One could get inspiration from recent speech research, where non-linear models for multi-variate neural data analysis is beginning to appear in the form of deep neural networks (Yang *et al.*, 2015; Akbari *et al.*, 2019).

Finally, it would be interesting to move towards more naturalistic experiments that also include visual input to the subject. It is known that the auditory and visual system work together to localize objects within the environment (see review by King, 2008) and so studying both senses in isolation can provide only limited insight on how we localize objects within our environment. Here, we "probed this" in chapter 6, where we delivered spatial auditory stimulus together with a simple visual environment that did not provide cues about the sound's location over Oculus VR headset. We believe, that using VR technology, which undergoes rapid development, might be particularly suitable for this as it allows precisely controlling both audio and video input while in real-time compensating for subjects movement (see Bohil *et al.*, 2011)

## 7.3. Practical applications

Recently, there have been attempts to implement cognitively steered hearing aids, which would automatically (without user's manual control) detect which aspect (speaker) of the auditory scene is attended and amplify it while filtering-out the irrelevant sounds. Currently, one of the most promising ways to decode selective attention is principally based on the stimulus envelope reconstruction method (Mesgarani & Chang, 2012; O'Sullivan *et al.*, 2015; Mirkovic *et al.*, 2016). Here, in chapter 4, we demonstrated that the attended source trajectory has a stronger cortical representation than the unattended source and we also showed that the reconstructed trajectory of the attended source is relatively unaffected by the distractor's location. There are two ways how this can be utilized in order to help with selective amplification of the attended source: (1) One could possibly use the trajectory decoder output (source azimuth) directly to steer the microphone array on the hearing aid e.g. using a beam-former to be more sensitive to sound coming from the direction corresponding to attended sound source. (2) Assuming we know the location of sound sources within the environment, one could use the trajectory decoder to determine which speaker is attended and then amplify the corresponding sound using the same pipeline that is used in the case of envelope based methods. Finally, we speculate that it might be beneficial to integrate both types of decoders, as we showed that a decoder that combined envelope and trajectory reconstruction had slightly higher decoding accuracy than each method separately.

Another possible practical application of our trajectory reconstruction method could be used in assessing spatial fidelity of virtual acoustic environments. The objective evaluation of virtual environments is typically done using psychophysically-based methods i.e. listeners are asked to determine the location of test stimuli or to note differences between sounds (Carlile, 1996; Xie, 2013). This approach is time-consuming and requires active participation of the listener. We speculate that our stimulus trajectory approach could be theoretically used as an alternative to these methods. Specifically, one could compare the trajectory reconstruction performance between the reference (free-field) sound delivery system and the tested virtual environment and any mismatch between the reconstruction models would signalise reduced fidelity of the tested environment. The advantage of this method would be the passive role of the listener since no active feedback would be required.

## 7.4. Summary

This thesis provides a novel linear regression-based framework to study spatial hearing. In our first study, we showed that the azimuth of a continuously moving noise source can be reconstructed from EEG. In particular, we identified two distinct frequency components, namely, delta (0-2Hz) and the alpha power (8-12Hz) of EEG that track the sound location. Importantly, by replicating the results using spatially-impoverished pulse stimuli, we showed that this method does not rely on a particular acoustic cue and is independent from the previously described sound envelope tracking by the cortex.

In our second study we employed a more naturalistic acoustic scenario with multiple moving sound sources. We showed that the attended sound source trajectory can be reliably reconstructed from EEG even in the presence of other competing sources and demonstrated that the trajectory tracking works for noise as well as more complex speech stimuli. Interestingly, we also observed weak tracking of the unattended source location for the speech stimuli. Finally, we demonstrated that the trajectory reconstruction approach can be theoretically used to decode selective attention, however, in comparison to envelope reconstruction-based methods, this method has lower decoding performance.

The study in chapter 5 tested whether the cortex is sensitive to other motion characteristics of sound besides location. We showed that sound speed but not velocity can be reconstructed from EEG independently from sound azimuth. Surprisingly, our results also indicate that sound acceleration might be independently represented at the cortex, which has not been reported before.

Finally, in the last study, we deployed our reconstruction method in a naturalistic scenario where subjects were allowed to move their heads and received visual input over a virtual

reality headset. We were primarily interested in whether sound is cortically encoded using cranio- or allo-centric coordinates. Although our initial analysis using the raw data indicated a cranio-centric representation of sound location, the subsequent analysis on the data with the motion-related artefact removed led to below-chance level reconstruction accuracies. Therefore, we were unable to find strong evidence for cortical encoding in either frame of reference. Our secondary goal was to test the feasibility of using the Oculus Rift headset together with high-density EEG recording. We found it is possible to perform experiments using this setup, however, several practical issues need to be resolved, mainly how to increase comfort of the subjects to allow for longer recording time.

# Chapter 8. Bibliography


Ahveninen, J., Huang, S., Belliveau, J.W., Chang, W.-T. & Hämäläinen, M. (2013) Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. *Journal of cognitive neuroscience*, **25**, 1926-1943.

Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., Witzel, T. & Belliveau, J.W. (2006) Task-modulated "what" and "where" pathways in human auditory cortex. *Proc Natl Acad Sci U S A*, **103**, 14608-14613.

Ahveninen, J., Kopco, N. & Jaaskelainen, I.P. (2014) Psychophysics and neuronal bases of sound localization in humans. *Hear Res*, **307**, 86-97.

Akbari, H., Khalighinejad, B., Herrero, J.L., Mehta, A.D. & Mesgarani, N. (2019) Towards reconstructing intelligible speech from the human auditory cortex. *Scientific Reports*, **9**, 874.

Alink, A., Euler, F., Kriegeskorte, N., Singer, W. & Kohler, A. (2012) Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum. Brain Mapp.*, **33**, 969-978.

Altman, J., Vaitulevich, S.P., Shestopalova, L. & Petropavlovskaia, E. (2010) How does mismatch negativity reflect auditory motion? *Hearing Research*, **268**, 194-201.

Altmann, C.F., Bledowski, C., Wibral, M. & Kaiser, J. (2007) Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage*, **35**, 1192-1200.

Altmann, C.F., Getzmann, S. & Lewald, J. (2012) Allocentric or craniocentric representation of acoustic space: an electrotomography study using mismatch negativity. *PLoS One*, **7**, e41872.

Altmann, C.F., Henning, M., Doring, M.K. & Kaiser, J. (2008) Effects of feature-selective attention on auditory pattern and location processing. *NeuroImage*, **41**, 69-79.

Altmann, C.F., Ueda, R., Bucher, B., Furukawa, S., Ono, K., Kashino, M., Mima, T. & Fukuyama, H. (2017) Trading of dynamic interaural time and level difference cues and its effect on the auditory motion-onset response measured with electroencephalography. *NeuroImage*, **159**, 185-194.

Altmann, C.F., Wilczek, E. & Kaiser, J. (2009) Processing of Auditory Location Changes after Horizontal Head Rotation. *J. Neurosci.*, **29**, 13074-13078.

Arnott, S.R., Binns, M.A., Grady, C.L. & Alain, C. (2004) Assessing the auditory dual-pathway model in humans. *NeuroImage*, **22**, 401-408.

Banerjee, S., Snyder, A.C., Molholm, S. & Foxe, J.J. (2011) Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? *Journal of Neuroscience*, **31**, 9923-9932.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M.G., Heinze, H.-J. & Scheich, H. (1999) A movement-sensitive area in auditory cortex. *Nature*, **400**, 724-726.

Bednar, A., Boland, F.M. & Lalor, E.C. (2017) Different spatio-temporal electroencephalography features drive the successful decoding of binaural and monaural cues for sound localization. *European Journal of Neuroscience*, **45**, 679-689.

Bednar, A. & Lalor, E.C. (2018) Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *NeuroImage*, **181**, 683-691.

Best, V., van Schaik, A. & Carlile, S. (2004a) Separation of concurrent broadband sound sources by human listeners. *The Journal of the Acoustical Society of America*, **115**, 324-336.

Best, V., van Schaik, A. & Carlile, S.J.T.J.o.t.A.S.o.A. (2004b) Separation of concurrent broadband sound sources by human listeners. **115**, 324-336.

Bizley, J.K., Nodal, F.R., Parsons, C.H. & King, A.J. (2007) Role of Auditory Cortex in Sound Localization in the Midsagittal Plane. **98**, 1763-1774.

Blauert, J. (1997) *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Bohil, C., Alicea, B. & Biocca, F. (2011) *Virtual reality in neuroscience research and therapy*.

Borst, A. & Egelhaaf, M. (1989) Principles of visual motion detection. *Trends in Neurosciences*, **12**, 297-306.

Bregman, A.S. (1994) *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Briley, P.M., Goman, A.M. & Summerfield, A.Q. (2016) Physiological Evidence for a Midline Spatial Channel in Human Auditory Cortex. *J. Assoc. Res. Otolaryngol.*, **17**, 331-340.

Briley, P.M., Kitterick, P.T. & Summerfield, A.Q. (2013) Evidence for opponent process analysis of sound source location in humans. *J Assoc Res Otolaryngol*, **14**, 83-101.

Brungart, D.S. & Rabinowitz, W.M. (1999) Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America*, **106**, 1465-1479.

Burke, K.A., Letsos, A. & Butler, R.A. (1994) Asymmetric performances in binaural localization of sound in space. *Neuropsychologia*, **32**, 1409-1417.

Carl, D. & Gutschalk, A.J.E.b.r. (2013) Role of pattern, regularity, and silent intervals in auditory stream segregation based on inter-aural time differences. **224**, 557-570.

Carlile, S. (1996) *Virtual Auditory Space: Generation and Applications*. Landes Bioscience.

Carlile, S. & Best, V. (2002) Discrimination of sound source velocity in human listeners. *The Journal of the Acoustical Society of America*, **111**, 1026-1035.

Carlile, S., Fox, A., Orchard-Mills, E., Leung, J. & Alais, D.J.J.o.t.A.f.R.i.O. (2016) Six degrees of auditory spatial separation. **17**, 209-221.

Carlile, S. & Leung, J. (2016) The perception of auditory motion. *Trends in hearing*, **20**, 2331216516644254.

Caruso, V.C., Mohl, J.T., Glynn, C., Lee, J., Willett, S.M., Zaman, A., Ebihara, A.F., Estrada, R., Freiwald, W.A. & Tokdar, S.T. (2018) Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nature communications*, **9**, 2715.

Chandler, D.W. & Grantham, D.W. (1992) Minimum Audible Movement Angle in the Horizontal Plane as a Function of Stimulus Frequency and Bandwidth, Source Azimuth, and Velocity. *Journal of the Acoustical Society of America*, **91**, 1624-1636.

Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, **25**, 975-979.

Clarke, S. & Geiser, E. (2015) Roaring lions and chirruping lemurs: How the brain encodes sound objects in space. *Neuropsychologia*, **75**, 304-313.

Combrisson, E. & Jerbi, K. (2015) Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J Neurosci Methods*, **250**, 126-136.

Crosse, M.J., Liberto, D., M, G., Bednar, A. & Lalor, E.C. (2016) The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci.*, **10**.

Day, M.L. & Delgutte, B.J.J.o.N. (2013) Decoding sound source location and separation using neural population activity patterns. **33**, 15837-15847.

Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, **134**, 9-21.

Deouell, L.Y., Heller, A.S., Malach, R., D'Esposito, M. & Knight, R.T. (2007) Cerebral responses to change in spatial location of unattended sounds. *Neuron*, **55**, 985-996.

Derey, K., Valente, G., de Gelder, B. & Formisano, E. (2016) Opponent Coding of Sound Location (Azimuth) in Planum Temporale is Robust to Sound-Level Variations. *Cerebral Cortex*, **26**, 450-464.

Ding, N. & Simon, J.Z. (2012a) Emergence of neural encoding of auditory objects while listening to competing speakers. *PNAS*, **109**, 11854-11859.

Ding, N. & Simon, J.Z. (2012b) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, **107**, 78-89.

Dingle, R.N., Hall, S.E. & Phillips, D.P.J.H.r. (2010) A midline azimuthal channel in human spatial hearing. **268**, 67-74.

Dingle, R.N., Hall, S.E. & Phillips, D.P.J.T.J.o.t.A.S.o.A. (2012) The three-channel model of sound localization mechanisms: interaural level differences. **131**, 4023-4029.

Dong, C.-J., Swindale, N.V., Zakarauskas, P., Hayward, V. & Cynader, M.S. (2000) The auditory motionaftereffect: Its tuning and specificity in the spatial and frequency domains. *Perception & Psychophysics*, **62**, 1099-1111.

Edmonds, B.A. & Krumbholz, K. (2014) Are interaural time and level differences represented by independent or integrated codes in the human auditory cortex? *J Assoc Res Otolaryngol*, **15**, 103-114.

Feddersen, W., Sandel, T., Teas, D. & Jeffress, L. (1957) Localization of high-frequency tones. *the Journal of the Acoustical Society of America*, **29**, 988-991.

Feng, W., Störmer, V.S., Martinez, A., McDonald, J.J. & Hillyard, S.A. (2017) Involuntary orienting of attention to a sound desynchronizes the occipital alpha rhythm and improves visual perception. *NeuroImage*, **150**, 318-328.

Fitzpatrick, D.C. & Kuwada, S. (2001) Tuning to interaural time differences across frequency. *Journal of Neuroscience*, **21**, 4844-4851.

Fujiki, N., Riederer, K.A.J., Jousmaki, V., Makela, J.P. & Hari, R. (2002) Human cortical representation of virtual auditory space: differences between sound azimuth and elevation. *European Journal of Neuroscience*, **16**, 2207-2213.

Gamble, M.L. & Luck, S.J. (2011) N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology*, **48**, 1057-1068.

Getzmann, S. (2008) Effects of velocity and motion-onset delay on detection and discrimination of sound motion. *Hearing Research*, **246**, 44-51.

Getzmann, S. (2009) Effect of auditory motion velocity on reaction time and cortical processes. *Neuropsychologia*, **47**, 2625-2633.

Getzmann, S. (2011) Auditory motion perception: onset position and motion direction are encoded in discrete processing stages. *European Journal of Neuroscience*, **33**, 1339-1350.

Getzmann, S. & Lewald, J. (2010) Shared Cortical Systems for Processing of Horizontal and Vertical Sound Motion. *Journal of Neurophysiology*, **103**, 1896-1904.

Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P. & Liegeois-Chauvel, C. (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin Neurophysiol*, **112**, 1850-1859.

Grantham, D.W. (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the Acoustical Society of America*, **79**, 1939-1949.

Grantham, D.W. (1989) Motion aftereffects with horizontally moving sound sources in the free field. *Perception & psychophysics*, **45**, 129-136.

Grantham, D.W. (1997) Auditory motion perception: Snapshots revisited. *Binaural and spatial hearing in real and virtual environments*, 295-313.

Grantham, D.W. & Wightman, F.L. (1979) Auditory motion aftereffects. *Perception & Psychophysics*, **26**, 403-408.

Griffiths, T.D. & Green, G.G. (1999) Cortical activation during perception of a rotating wide-field acoustic stimulus. *NeuroImage*, **10**, 84-90.

Griffiths, T.D., Green, G.G., Rees, A. & Rees, G. (2000) Human brain areas involved in the analysis of auditory movement. *Hum. Brain Mapp.*, **9**, 72-80.

Griffiths, T.D., Rees, G., Rees, A., Green, G.G., Witton, C., Rowe, D., Büchel, C., Turner, R. & Frackowiak, R.S. (1998) Right parietal cortex is involved in the perception of sound movement in humans. *Nat Neurosci*, **1**, 74.

Grothe, B., Pecka, M. & McAlpine, D. (2010a) Mechanisms of Sound Localization in Mammals. *Physiological Reviews*, **90**, 983-1012.

Grothe, B., Pecka, M. & McAlpine, D. (2010b) *Mechanisms of Sound Localization in Mammals*.

Harrington, I.A., Stecker, G.C., Macpherson, E.A. & Middlebrooks, J.C.J.H.r. (2008) Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. **240**, 22-41.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B. & Bießmann, F. (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, **87**, 96-110.

Henning, G.B. (1974) Detectability of interaural delay in high-frequency complex waveforms. *The Journal of the Acoustical Society of America*, **55**, 84-90.

Higgins, N.C., McLaughlin, S.A., Rinne, T. & Stecker, G.C. (2017) Evidence for cue-independent spatial representation in the human auditory cortex during active listening. *PNAS*, 201707522.

Jenkins, W.M. & Masterton, R.B. (1982) Sound localization: effects of unilateral lesions in central auditory system. *Journal of Neurophysiology*, **47**, 987-1016.

Johnson, B.W. & Hautus, M.J. (2010) Processing of binaural spatial information in human auditory cortex: neuromagnetic responses to interaural timing and level differences. *Neuropsychologia*, **48**, 2610-2619.

Joris, P.X. (2003) Interaural time sensitivity dominated by cochlea-induced envelope patterns. *Journal of Neuroscience*, **23**, 6345-6350.

Kaas, J.H. & Hackett, T.A. (2000) Subdivisions of auditory cortex and processing streams in primates. *PNAS*, **97**, 11793-11799.

Kaczmarek, T. (2005) Auditory perception of sound source velocity. *The Journal of the Acoustical Society of America*, **117**, 3149-3156.

Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H. & Birbaumer, N. (2000) Right-hemisphere dominance for the processing of sound-source lateralization. *J Neurosci*, **20**, 6631-6639.

Kelly, S.P., Gomez-Ramirez, M. & Foxe, J.J. (2009) The strength of anticipatory spatial biasing predicts target discrimination at attended locations: a high-density EEG study. *European Journal of Neuroscience*, **30**, 2224-2234.

Kerlin, J.R., Shahin, A.J. & Miller, L.M. (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, **30**, 620-628.

King, A.J. (2008) Visual influences on auditory spatial learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 331-339.

Krumbholz, K., Hewson-Stoate, N. & Schönwiesner, M. (2007) Cortical Response to Auditory Motion Suggests an Asymmetry in the Reliance on Inter-Hemispheric Connections Between the Left and Right Auditory Cortices. *Journal of Neurophysiology*, **97**, 1649-1655.

Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G.R. & von Cramon, D.Y. (2005a) Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience*, **21**, 230-238.

Krumbholz, K., Schonwiesner, M., von Cramon, D.Y., Rubsamen, R., Shah, N.J., Zilles, K. & Fink, G.R. (2005b) Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cereb Cortex*, **15**, 317-324.

Kuhn, G.F. (1977) Model for the interaural time differences in the azimuthal plane. *The Journal of the Acoustical Society of America*, **62**, 157-167.

Lalor, E.C. & Foxe, J.J. (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *The European journal of neuroscience*, **31**, 189-193.

Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G. & Foxe, J.J. (2006) The VESPA: a method for the rapid estimation of a visual evoked potential. *NeuroImage*, **32**, 1549-1561.

Lalor, E.C., Power, A.J., Reilly, R.B. & Foxe, J.J. (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of neurophysiology*, **102**, 349-359.

Lewald, J. & Getzmann, S. (2011) When and where of auditory spatial processing in cortex: a novel approach using electrotomography. *PLoS One*, **6**, e25146.

Lewald, J. & Getzmann, S. (2015) Electrophysiological correlates of cocktail-party listening. *Behav Brain Res*, **292**, 157-166.

Lewald, J., Hanenberg, C. & Getzmann, S. (2016) Brain correlates of the orientation of auditory spatial attention onto speaker location in a "cocktail-party" situation. *Psychophysiology*, **53**, 1484-1495.

Locke, S.M., Leung, J. & Carlile, S. (2016) Sensitivity to auditory velocity contrast. *Scientific reports*, **6**, 27725.

Luck, S.J. & Kappenman, E.S. (2011) *The Oxford handbook of event-related potential components*. Oxford university press.

Lunner, T., Rudner, M. & Rönnberg, J. (2009) Cognition and hearing aids. *Scandinavian journal of psychology*, **50**, 395-403.

Luo, H., Liu, Z. & Poeppel, D. (2010) Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation. *PLOS Biology*, **8**, e1000445.

Macpherson, E.A. & Middlebrooks, J.C. (2002) Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, **111**, 2219-2236.

Magezi, D.A., Buetler, K.A., Chouiter, L., Annoni, J.-M. & Spierer, L. (2013) Electrical neuroimaging during auditory motion aftereffects reveals that auditory motion processing is motion sensitive but not direction selective. *Journal of Neurophysiology*, **109**, 321-331.

Magezi, D.A. & Krumbholz, K. (2010) Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J Neurophysiol*, **104**, 1997-2007.

Makela, J.P. & McEvoy, L. (1996) Auditory evoked fields to illusory sound source movements. *Exp Brain Res,* **110**, 446-454.

Makous, J.C. & Middlebrooks, J.C. (1990) Two-dimensional sound localization by human listeners. *The journal of the Acoustical Society of America*, **87**, 2188-2200.

McAlpine, D., Jiang, D., Shackleton, T.M. & Palmer, A.R. (2000) Responses of neurons in the inferior colliculus to dynamic interaural phase cues: evidence for a mechanism of binaural adaptation. *Journal of Neurophysiology*, **83**, 1356-1365.

McLaughlin, S.A., Higgins, N.C. & Stecker, G.C. (2015) Tuning to Binaural Cues in Human Auditory Cortex. *J Assoc Res Otolaryngol,* 1-17.

Meng, J.-A., Saberi, K. & Hsieh, I.-H. (2016) Velocity Selective Networks in Human Cortex Reveal Two Functionally Distinct Auditory Motion Systems. *PloS one*, **11**, e0157131.

Mesgarani, N. & Chang, E.F. (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, **485**, 233.

Mesgarani, N., David, S.V., Fritz, J.B. & Shamma, S.A. (2009a) Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of neurophysiology*, **102**, 3329-3339.

Mesgarani, N., David, S.V., Fritz, J.B. & Shamma, S.A. (2009b) Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J Neurophysiol*, **102**, 3329-3339.

Middlebrooks, J.C. (2015) Sound localization *Handbook of clinical neurology*. Elsevier, pp. 99-116.

Middlebrooks, J.C. & Bremen, P. (2013) Spatial stream segregation by auditory cortical neurons. *Journal of Neuroscience*, **33**, 10986-11001.

Middlebrooks, J.C. & Green, D.M. (1991) Sound localization by human listeners. *Annu Rev Psychol*, **42**, 135-159.

Middlebrooks, J.C. & Onsan, Z.A. (2012) Stream segregation with high spatial acuity. *J Acoust Soc Am,* **132**, 3896-3911.

Middlebrooks, J.C. & Pettigrew, J.D. (1981) Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *J. Neurosci.*, **1**, 107-120.

Mills, A.W. (1958) On the Minimum Audible Angle. *Journal of the Acoustical Society of America*, **30**, 237-246.

Mirkovic, B., Bleichner, M.G., De Vos, M. & Debener, S. (2016) Target Speaker Detection with Concealed EEG Around the Ear. *Frontiers in Neuroscience*, **10**.

Mirkovic, B., Debener, S., Jaeger, M. & De Vos, M. (2015) Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *J Neural Eng*, **12**, 046007.

Młynarski, W. (2015) The Opponent Channel Population Code of Sound Location Is an Efficient Representation of Natural Binaural Sounds. *PLOS Computational Biology*, **11**, e1004294.

Mognon, A., Jovicich, J., Bruzzone, L. & Buiatti, M. (2011) ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, **48**, 229-240.

Moore, B.C. & Gockel, H.J.A.A.U.w.A. (2002) Factors influencing sequential stream segregation. **88**, 320-333.

Naatanen, R., Paavilainen, P., Rinne, T. & Alho, K. (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*, **118**, 2544-2590.

Naatanen, R. & Picton, T. (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, **24**, 375-425.

Neuhoff, J.G. (2004) Auditory motion and localization. *Ecological psychoacoustics*, 87-111.

O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A. & Lalor, E.C. (2015) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb Cortex*, **25**, 1697-1706.

Ortiz-Rios, M., Azevedo, F.A., Kusmierek, P., Balla, D.Z., Munk, M.H., Keliris, G.A., Logothetis, N.K. & Rauschecker, J.P. (2017) Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. *Neuron*, **93**, 971-983 e974.

Palomaki, K., Alku, P., Makinen, V., May, P. & Tiitinen, H. (2000) Sound localization in the human brain: neuromagnetic observations. *Neuroreport*, **11**, 1535-1538.

Palomaki, K.J., Tiitinen, H., Makinen, V., May, P.J. & Alku, P. (2005) Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques. *Brain Res Cogn Brain Res*, **24**, 364-379.

Pavani, F., Macaluso, E., Warren, J.D., Driver, J. & Griffiths, T.D. (2002) A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr Biol*, **12**, 1584-1590.

Pecka, M., Zahn, T.P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., Klug, A., Pollak, G.D. & Grothe, B. (2007) Inhibiting the inhibition: a neuronal network for sound localization in reverberant environments. *Journal of Neuroscience*, **27**, 1782-1790.

Perrott, D.R., Costantino, B. & Ball, J. (1993) Discrimination of moving events which accelerate or decelerate over the listening interval. *The Journal of the Acoustical Society of America*, **93**, 1053-1057.

Perrott, D.R. & Marlborough, K. (1989) Minimum audible movement angle: marking the end points of the path traveled by a moving sound source. *The Journal of the Acoustical Society of America*, **85**, 1773-1775.

Perrott, D.R. & Saberi, K. (1990) Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, **87**, 1728-1731.

Petkov, C.I., Kang, X., Alho, K., Bertrand, O., Yund, E.W. & Woods, D.L. (2004) Attentional modulation of human auditory cortex. *Nat Neurosci*, **7**, 658.

Picton, T. (2013) Hearing in time: evoked potential studies of temporal processing. *Ear Hear*, **34**, 385-401.

Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., Sun, L., Rees, A., Petkov, C.I., Thiele, A. & Griffiths, T.D. (2017) Auditory motion-specific mechanisms in the primate brain. *PLOS Biology*, **15**, e2001379.

Poirier, C., Collignon, O., DeVolder, A.G., Renier, L., Vanlierde, A., Tranduy, D. & Scheiber, C. (2005) Specific activation of the V5 brain area by auditory motion processing: an fMRI study. *Cognitive Brain Research*, **25**, 650-658.

Rauschecker, J.P. (2017) Where, When, and How: are they all Sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*.

Rauschecker, J.P. & Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *PNAS*, **97**, 11800-11806.

Rayleigh, L. (1875) On our perception of the direction of a source of sound. *Proceedings of the Musical Association*, **2**, 75-84.

Rayleigh, L. (1907) XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **13**, 214-232.

Recanzone, G.H., Makhamra, S.D. & Guard, D.C. (1998) Comparison of relative and absolute sound localization ability in humans. *J Acoust Soc Am*, **103**, 1085-1097.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S. & Rauschecker, J.P. (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*, **2**, 1131-1136.

Salminen, N.H., Takanen, M., Santala, O., Lamminsalo, J., Altoe, A. & Pulkki, V. (2015) Integrated processing of spatial cues in human auditory cortex. *Hear Res*, **327**, 143-152.

Schechtman, E., Shrem, T. & Deouell, L.Y. (2012) Spatial localization of auditory stimuli in human auditory cortex is based on both head-independent and head-centered coordinate systems. *J Neurosci*, **32**, 13501-13509.

Schnupp, J., Nelken, I. & King, A. (2011) *Auditory neuroscience: Making sense of sound*. MIT Press.

Shestopalova, L.B., Petropavlovskaia, E.A., Vaitulevich, S.P., Vasilenko, Y.A., Nikitin, N.I. & Altman, J.A. (2012) Discrimination of auditory motion patterns: the mismatch negativity study. *Neuropsychologia*, **50**, 2720-2729.

Shiell, M.M., Hausfeld, L. & Formisano, E. (2018) Activity in Human Auditory Cortex Represents Spatial Separation Between Concurrent Sounds. *Journal of Neuroscience*, **38**, 4977-4984.

Shinn-Cunningham, B.G., Schickler, J., Kopčo, N. & Litovsky, R. (2001) Spatial unmasking of nearby speech sources in a simulated anechoic environment. *The Journal of the Acoustical Society of America*, **110**, 1118-1129.

Smith, K.R., Okada, K., Saberi, K. & Hickok, G. (2004) Human cortical auditory motion areas are not motion selective. *Neuroreport*, **15**, 1523-1526.

Smith, K.R., Saberi, K. & Hickok, G. (2007) An event-related fMRI study of auditory motion perception: No evidence for a specialized cortical system. *Brain Research*, **1150**, 94-99.

Spitzer, M.W. & Semple, M.N. (1991) Interaural phase coding in auditory midbrain: influence of dynamic stimulus features. *Science*, **254**, 721-724.

Spitzer, M.W. & Semple, M.N. (1998) Transformation of Binaural Response Properties in the Ascending Auditory Pathway: Influence of Time-Varying Interaural Phase Disparity. *Journal of Neurophysiology*, **80**, 3062-3076.

Stecker, G. & Gallun, F. (2012) Binaural hearing, sound localization, and spatial hearing. *Translational perspectives in auditory neuroscience: Normal aspects of hearing*, 383-434.

Stecker, G.C., Harrington, I.A. & Middlebrooks, J.C. (2005) Location coding by opponent neural populations in the auditory cortex. *PLoS Biol*, **3**, e78.

Stecker, G.C., Mickey, B.J., Macpherson, E.A. & Middlebrooks, J.C. (2003) Spatial Sensitivity in Field PAF of Cat Auditory Cortex. *Journal of Neurophysiology*, **89**, 2889-2903.

Störmer, V.S., Feng, W., Martinez, A., McDonald, J.J. & Hillyard, S.A. (2016) Salient, irrelevant sounds reflexively induce alpha rhythm desynchronization in parallel with slow potential shifts in visual cortex. *Journal of cognitive neuroscience*.

Thompson, S.P. (1878) L. Phenomena of binaural audition—Part II. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **6**, 383-391.

Tian, B., Reser, D., Durham, A., Kustov, A. & Rauschecker, J.P. (2001) Functional Specialization in Rhesus Monkey Auditory Cortex. *Science*, **292**, 290-293.

Town, S.M., Brimijoin, W.O. & Bizley, J.K. (2017) Egocentric and allocentric representations in auditory cortex. *PLOS Biology*, **15**, e2001878.

Trapeau, R. & Schönwiesner, M. (2018) The Encoding of Sound Source Elevation in the Human Auditory Cortex. *J. Neurosci.*, **38**, 3252-3264.

Ungan, P., Yagcioglu, S. & Goksoy, C. (2001) Differences between the N1 waves of the responses to interaural time and intensity disparities: scalp topography and dipole sources. *Clinical Neurophysiology*, **112**, 485-498.

van Drongelen, W. (2006) *Signal Processing for Neuroscientists: An Introduction to the Analysis of Physiological Signals*. Elsevier Science.

Wallach, H. (1939) On sound localization. *The Journal of the Acoustical Society of America*, **10**, 270-274.

Wallach, H. (1940) The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, **27**, 339.

Wang, X., Liu, J. & Zhang, J. (2019) Chronic Unilateral Hearing Loss Disrupts Neural Tuning to Sound-Source Azimuth in the Rat Primary Auditory Cortex. *Frontiers in Neuroscience*, **13**.

Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P. & Griffiths, T.D. (2002) Perception of Sound-Source Motion by the Human Brain. *Neuron*, **34**, 139-148.

Webster, J.G. (2009) *Medical Instrumentation Application and Design*. John Wiley & Sons.

Werner-Reiss, U. & Groh, J.M. (2008) A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *J. Neurosci.*, **28**, 3747-3758.

Wightman, F.L. & Kistler, D.J. (1989) Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, **85**, 868-878.

Wightman, F.L. & Kistler, D.J. (1999) Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, **105**, 2841-2853.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D. & Bloom, F.E. (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. *PNAS*, **90**, 8722-8726.

Wood, K., Town, S. & Bizley, J. (2018) *Primary auditory cortex represents the location of sound sources in a cue-invariant manner*.

Worden, M.S., Foxe, J.J., Wang, N. & Simpson, G.V. (2000) Anticipatory biasing of visuospatial attention indexed by retinotopically specific-band electroencephalography increases over occipital cortex. *J Neurosci*, **20**, 1-6.

Wöstmann, M., Herrmann, B., Maess, B. & Obleser, J. (2016) Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 3873-3878.

Wostmann, M., Vosskuhl, J., Obleser, J. & Herrmann, C.S. (2018) Opposite effects of lateralised transcranial alpha versus gamma stimulation on auditory spatial attention. *Brain stimulation*, **11**, 752-758.

Xiang, J., Daniel, S.J., Ishii, R., Holowka, S., Harrison, R.V. & Chuang, S.J.B.t. (2005) Auditory detection of motion velocity in humans: a magnetoencephalographic study. **17**, 139-149.

Xie, B. (2013) *Head-Related Transfer Function and Virtual Auditory Display*. J. Ross Publishing, Incorporated.

Yang, M., Sheth, S.A., Schevon, C.A., Ii, G.M.M. & Mesgarani, N. (Year) Speech reconstruction from human auditory cortex with deep neural networks. Sixteenth Annual Conference of the International Speech Communication Association. City.

Yin, T.C. (2002) Neural mechanisms of encoding binaural localization cues in the auditory brainstem *Integrative functions in the mammalian auditory pathway*. Springer, pp. 99-159.

Yost, W.A. & Brown, C.A.J.T.J.o.t.A.S.o.A. (2013) Localizing the sources of two independent noises: Role of time varying amplitude differences. **133**, 2301-2313.

Zatorre, R.J., Bouffard, M., Ahad, P. & Belin, P. (2002) Where is' where'in the human auditory cortex? *Nat Neurosci*, **5**, 905-909.

Zatorre, R.J. & Penhune, V.B. (2001) Spatial localization after excision of human auditory cortex. *J Neurosci*, **21**, 6321-6328.

Zhang, X., Zhang, Q., Hu, X. & Zhang, B. (2015) Neural representation of three-dimensional acoustic space in the human temporal lobe. *Front Hum Neurosci*, **9**, 203.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D. & Schroeder, C.E. (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, **77**, 980-991.

Zündorf, I.C., Karnath, H.-O. & Lewald, J. (2014) The effect of brain lesions on sound localization in complex acoustic environments. *Brain*, **137**, 1410-1418.

Zündorf, I.C., Lewald, J. & Karnath, H.-O. (2013) Neural Correlates of Sound Localization in Complex Acoustic Environments. *PLOS ONE*, **8**, e64259.