KU LEUVEN

FACULTEIT GENEESKUNDE

LOGOPEDISCHE EN AUDIOLOGISCHE WETENSCHAPPEN
ON2 | HERESTRAAT 49 - BUS 721
B-3000 LEUVEN

The effect of speaker separation and noise level on auditory attention detection

Tine Arras

Verhandeling aangeboden tot het behalen van de graad van
Master in de Logopedische en Audiologische Wetenschappen

Promotor: Prof. Dr. T. Francart
Copromotoren: Drs. N. Das, Prof. Dr. A. Bertrand

– juni 2018 –

KU LEUVEN

FACULTEIT GENEESKUNDE

LOGOPEDISCHE EN AUDIOLOGISCHE WETENSCHAPPEN
ON2 | HERESTRAAT 49 - BUS 721
B-3000 LEUVEN

The effect of speaker separation and noise level on auditory attention detection

Tine Arras

Verhandeling aangeboden tot het behalen van de graad van
Master in de Logopedische en Audiologische Wetenschappen

Promotor: Prof. Dr. T. Francart
Copromotoren: Drs. N. Das, Prof. Dr. A. Bertrand

– juni 2018 –

Word of thanks

Writing a thesis involves a lot of hard work. Luckily, I received a lot of support - so I would like to thank everyone who helped me along the way. I'm very proud of the result, I hope you all are too. There are some people I would like to thank personally, because they contributed in a special way to this thesis. Firstly, my promotor, professor Tom Francart, who gave me very useful feedback and helped me make this thesis better than I could have hoped. Secondly, my co-promotors Alexander Bertrand and Neetha Das, for their aid during the past year. Especially Neetha, who supported me through the entire process and who rescued me when I got lost in the maze of research questions and R commands. Thirdly, Hans Lambaerts, who proofread this thesis and helped me gain an insight into correct grammar and a careful choice of words in English. I have myself to blame for all remaining errors.

Next, I want to extend my gratitude to all the subjects who participated in the experiment. Not only did they volunteer to wear a cap full of gel and electrodes on their head, they also sat through what was probably the longest four hours of their life without complaining. I couldn't have done this without you! I want to thank my parents too, for believing in me, supporting me, and gracefully accepting my mood swings when I couldn't get R to make the right graphs. Apart from that, my dad also helped me handle the massive amount of data, by rearranging them into a neat row of well-behaved numbers. Thanks a lot! And finally, my friends, who patiently listened to me when I was complaining again about how difficult it all was and helped me realise there is more to life than statistically significant results. You kept me sane - or at least, not crazier than I was before. Thank you!

Abstract

In a cocktail party scenario, with multiple talkers in a setting with a high amount of background noise, listeners with normal hearing can direct their attention to one speech stream and ignore the other sounds. This remarkable ability is based on the separate encoding of different speech streams, allowing attention-based gain control to enhance the attended stream. The organisation of the concurrent speech streams in the cortex affects the emergence of auditory evoked potentials (AEPs) and can be measured using brain imaging techniques. Auditory attention detection (AAD) uses these brain recordings to decide which speaker a person is listening to.

So far, most AAD researchers have decoded attention in a two-talker scenario. However, many acoustic environments contain more than two concurrent speech streams. Furthermore, only a few studies investigate the effect of background noise or other interfering factors. To start filling these gaps, this thesis applies AAD in a two-talker scenario with babble noise at different levels. Additionally, the separation angle between speakers is varied. The first two research questions assess the effect of both spatial separation and background noise on the decoding accuracy. The third research question examines a possible interaction effect between spatial separation and noise. The fourth research question studies the effect of individual differences on AAD performance. Finally, the fifth research question connects the decoding accuracy to the subjective speech intelligibility ratings.

The brain activity of 15 young adults with normal hearing was measured using 64-channel EEG. The experiment was split into different trials, during which the subjects listened to a story while ignoring a competing talker. Both separation angles and noise levels were varied so that all subjects experienced every combination of these factors. After each trial, subjects estimated the intelligibility of the story part they just heard. Afterwards, subject-specific decoders were constructed using the leave-one-out method.

High decoding accuracies were achieved, with a mean accuracy of 81.9% across all subjects (range 67.7-92.0%). Analyses showed that both separation angle and noise level had a significant effect on the decoding accuracy, just like the interaction between them. A smaller separation angle resulted in a lower AAD performance. For trials with babble noise, more noise resulted in lower decoding accuracies. Furthermore, individual differences accounted for much variance in AAD performance, although the decoding accuracy was not correlated with the subjects' performance on a speech-in-noise task. Finally, the subjective intelligibility was not significantly correlated with the decoding accuracy, yet it was similarly affected by separation angles, noise levels,

and the interaction between them. Intelligibility was lowest for trials with a small separation angle and/or a high amount of noise.

In conclusion, different acoustic parameters affected AAD performance. Decoding accuracy was not linked to subjective intelligibility, although the same parameters influenced these ratings. Individual differences indicated that the success of AAD is not the same for everyone. Future research should include more subjects, preferably adding hearing-impaired listeners and people from different age ranges, and focus on lifelike conditions that include different speaker orientations, varying noise levels, and realistic stimuli.

Samenvatting

In situaties waarin verschillende sprekers door elkaar heen praten, zijn normaal horende personen in staat om hun aandacht op één spreker te richten en de andere geluiden te negeren. Deze opmerkelijke vaardigheid steunt op het vermogen van de hersenen om elke geluidsbron apart te verwerken, en de corticale representatie van de gekozen geluidsbron te versterken. De mentale organisatie van de geluidsbronnen heeft een effect op de auditieve hersenrespons en kan daardoor via beeldvormingstechnieken worden opgemeten. Auditieve aandachtsdetectie (AAD) gebruikt zulke technieken om te bepalen naar welke spreker een persoon luistert.

Tot nog toe hebben de meeste onderzoekers AAD toegepast in een scenario met twee sprekers. De meeste akoestische omgevingen bevatten echter meer dan twee geluidsbronnen. Bovendien houdt slechts een beperkt aantal studies rekening met het effect van achtergrondgeluid of andere storende factoren. Om tegemoet te komen aan die bezwaren, past deze thesis AAD toe op een situatie met twee sprekers en verschillende niveaus van spraakruis. Bovendien werd de hoek tussen beide sprekers gevarieerd. De eerste twee onderzoeksvragen peilen naar het effect van directioneel horen en achtergrondgeluid op de accuratesse van de AAD. De derde onderzoeksvraag gaat in op de interactie tussen die factoren. De vierde onderzoeksvraag bekijkt het belang van individuele verschillen in het kader van AAD. De vijfde vraag verbindt de accuratesse met de subjectieve spraakverstaanbaarheid.

De hersenactiviteit van 15 normaal horende jongvolwassenen werd gemeten via EEG, gebruik makend van 64 elektroden. Het experiment bestond uit verschillende stukjes of ‘trials’, waarin de deelnemers telkens naar een verhaal moesten luisteren terwijl ze de andere spreker negeerden. Zowel de hoek tussen beide sprekers als het niveau van de achtergrondruis veranderden tijdens het experiment, zodat elke combinatie van factoren aan bod kwam. Na elke trial moesten de deelnemers inschatten hoeveel spraak ze hadden verstaan. Achteraf werden de EEG-resultaten voor elke deelnemer apart verwerkt.

De gemiddelde accuratesse was 81.9% over alle deelnemers heen (bereik 67.7-92.0%). Zowel de hoek tussen de sprekers als het ruisniveau had een significant effect op de accuratesse, net als de interactie tussen beide. AAD was minder succesvol bij een kleinere scheidingshoek of bij een hoger ruisniveau. Individuele verschillen zorgden voor veel variatie in de scores, maar de gemiddelde accuratesse per deelnemer correleerde niet met hun score op een spraak-in-ruis-test. De subjectieve spraakverstaanbaarheid correleerde ook niet met de accuratesse, maar werd wel op een gelijkaardige manier door de scheidingshoek, het ruisniveau en de interactie ertussen

beïnvloed. De spraakverstaanbaarheid was het laagst in situaties met een kleine scheidingshoek en/of een hoog ruisniveau.

Het experiment toonde aan dat verschillende akoestische parameters een invloed hebben op AAD. De accuratesse was niet gelinkt aan de subjectieve maat voor spraakverstaan, maar ze werden wel door dezelfde factoren beïnvloed. Individuele verschillen toonden aan dat AAD niet voor iedereen even succesvol is. Toekomstig onderzoek zou meer deelnemers moeten testen, met bij voorkeur ook slechthorende patiënten en oudere personen. Bovendien moet het onderzoek gebaseerd zijn op realistische situaties, met verschillende posities van de sprekers, wisselende ruisniveaus en uiteenlopende stimuli.

Table of contents

List of abbreviations

This list provides the reader with an overview of the abbreviations used in the text.

AAD	Auditory attention detection
ABR	Auditory brainstem response
AEP	Auditory evoked potential
AERS	Auditory event representation system
ASA	Auditory scene analysis
ASSR	Auditory steady-state response
BCD	Bone conduction device
BCI	Brain-computer interface
CI	Cochlear implant
CT	Competing talker
dB HL	Decibel hearing level
ECoG or ECochG	Electrocochleography
EEG	Electro-encephalography
ERP	Event-related potential
FFR	Frequency-following response
HA	Hearing aid
HRTF	Head-related transfer function
ILD	Interaural level difference
ITD	Interaural time difference
LEA	Left ear advantage
MEG	Magneto-encephalography

MEI	Middle ear implant
MLR	Middle-latency response
MMN	Mismatch negativity
PTA	Pure tone average
REA	Right ear advantage
RON	Reorienting negativity
SNR	Signal-to-noise ratio
SRT	Speech reception threshold
SSA	Stimulus-specific adaptation

Introduction

The topic of this thesis is auditory attention detection (AAD). AAD is a fairly new technology, which analyses brain activity to decide which speaker a listener is paying attention to. In the future, this method may be used to improve the way a HA handles noise, by selectively amplifying the speech stream the user is listening to. So far, research has shown that AAD can achieve robust results; however, it has mainly been tested in two-talker scenarios without background noise. This thesis fits in with a doctoral thesis by Neetha Das, who investigates the effect of more realistic conditions on AAD performance. Ultimately, the goal is to expand the applicability of AAD across all realistic listening conditions, so that it can successfully be implemented in HA technology.

This thesis implements a two-talker scenario with varying levels of background noise and different separation angles between both speakers. The purpose of this experiment is to discover whether spatial separation and background noise influence AAD performance. Furthermore, the relationship between decoding accuracy and subjective speech intelligibility under these circumstances is explored. Finally, it is important to know to what extent individual differences contribute to the performance of AAD. To investigate these questions, an EEG experiment is set up to perform AAD on 15 healthy subjects with normal hearing. Subjects are asked to listen to a story, while ignoring a competing story. They must each listen to eight stories, with varying separation angles between both speakers and with different levels of babble noise in the background. After each part of the story, subjects must estimate what percentage of the narration they understood, to measure the subjective speech intelligibility.

When analysing the experiment results, important differences between conditions can be found. This indicates that AAD is very context-dependent, and that earlier research on two-talker scenarios was probably overestimating the success rates of AAD. However, for most conditions, high decoding accuracies were achieved. An important side note is related to between-subject variability, which seriously affects AAD performance. These individual differences in decoding results are very relevant when considering the future application of AAD in HA technology and should therefore be explored in more detail.

The social relevance of this work should not be underestimated. As very little evidence exists on the effectiveness of AAD in realistic acoustic environments, the results of this thesis could act as a starting point for future inquiries. To really map out the possibilities of this technology, lifelike situations should receive more attention in AAD research.

1 Literature review

The first part of this thesis relates the present work to recent insights and developments in the field of audiology research. The first section considers the mechanisms underlying auditory perception and attention, together with the interaction between them. It also includes information about hearing aids and their shortcomings. In the second section, the concept of auditory attention detection is explained, together with its potential to enhance hearing aid technology. The concluding section contains the research questions that are investigated in this work.

1.1 Auditory perception and attention

In this section, three elements relating to auditory perception and attention will be discussed: the mechanisms involved in the processing of auditory information, the important role of both bottom-up and top-down attention, and the differences between processing general auditory stimuli and understanding speech. The final paragraph discusses hearing loss, along with the use of hearing aids and the associated problems.

1.1.1 Processing auditory information

By working side by side with the brain, the human ear helps people understand the world around them and communicate with their environment. When a sound is generated, the vibrating air pushes against the eardrum, which converts acoustic vibrations to mechanical energy. Next, the ossicles in the middle ear transmit the movement to the oval window. When the oval window starts vibrating, the fluid inside the cochlea is set in motion, causing the basilar membrane inside the cochlea to vibrate as well. The vibrations have different amplitudes along the membrane, depending on the frequency of the sound. This frequency-specific movement evokes an electrical response in the corresponding hair cells inside the cochlea. As a result, neural signals are sent through the auditory nerve towards the auditory cortex (McFarland, 2009; MED-EL, 2012). These signals implicitly carry information about the location of the hair cells, indicating the frequency of the sound. The higher levels of the central auditory system analyse the auditory information, integrating it over time and from different locations at once to determine the intensity of the sound and the temporal pattern or rhythm. At the end of this process, the sound is recognized as a banging door, a musical note, a dripping tap, or a fragment from a speech stream.

The successive steps of auditory processing are reflected in the listener's brain activity. It is therefore possible to map the processing of sound by measuring auditory evoked potentials (AEPs). AEPs are event-related potentials (ERPs) that can be measured in the first 500 ms after the onset of a sound. Depending on the time range they fit into, they are considered fast (0-20 ms), middle-latency (20-80 ms) or slow (80-500 ms) responses (Picton, 2013). First, the sound is processed in the cochlea and brainstem. The resulting early or fast responses are generated within 20 ms after the onset of a sound. A typical example of fast responses are so-called auditory brainstem responses (ABRs), which are commonly used in neonatal hearing screening programs (Mason & Herrmann, 1998; van Straaten, 1999). When the auditory information reaches the auditory cortex, after about 20-80 ms, middle-latency responses (MLRs) are generated. The most important MLR-waveforms are labelled Na, Pa, and Nb. During further processing, late or slow responses emerge. In this time range, the P1, N1, P2, and N2 response components can be measured, as well as alpha, beta, and theta waves (Picton, 2013).

Aside from this latency-based classification, AEPs can also be classified based on the type of stimulus that evokes them. A transient response marks a change in the stimulus, while a sustained response lasts throughout the entire stimulus. When a stimulus changes repetitively, following responses emerge. Frequency-following responses (FFRs) track the frequency of the stimulus, whereas envelope-following responses arise when the sound is modulated. If the stimulus undergoes periodic changes, like when a certain stimulus is rapidly repeated, the following response becomes an auditory steady-state response (ASSR) (Picton, 2013; Picton, John, Dimitrijevic, & Purcell, 2003).

1.1.2 The role of attention

Attention is a complex process and can be split up into two parts: bottom-up attention, driven by the characteristics of the sound itself, and top-down attention, based on task demands and conscious selection.

1.1.2.1 Bottom-up auditory attention

In a complex stimulus, the most salient parts trigger a sensory-driven selection mechanism. This mechanism steers perception toward the most striking part of the stimulus and is called 'bottom-up attention' (Kaya & Elhilali, 2014). To decide which stimuli are salient and should therefore receive attention, the brain uses predictive coding. It integrates information about the acoustic scene or stimulus over time to find regularities (Kaya & Elhilali, 2014; Winkler, Denham, & Nelken, 2009). If the actual stimulus deviates from what is predicted, it is marked as 'deviant'. This irregularity is reflected in two types of brain responses: mismatch negativity (MMN), which is an ERP, and stimulus-specific adaptation (SSA), its single-neuron counterpart (Khouri & Nelken, 2015). Both responses are part of the same deviance detection system in the auditory pathway. There is, however, a difference in their level of complexity. When the deviance corresponds to a change in a simple acoustic feature, like frequency or location, it elicits a response in the middle-latency range (SSA) as well as an MMN. Changes in more complex regularities do not result in SSA but will still elicit MMN responses (Escera, Leung, & Grimm, 2014).

The detection of an acoustic change in the stimulus thus leads to an MMN, which is a frontally distributed, negative component of the human ERP. The MMN peaks between 100 and 200 ms after the onset of a deviant sound, even in the absence of attention or when attention is directed toward another part of the stimulus (Näätänen & Escera, 2000), and it can even be elicited by relatively abstract changes, like phonological or grammatical errors (Näätänen, Paavilainen, Rinne, & Alho, 2007; Paavilainen, 2013). The MMN is often associated with a simultaneous enhancement of the N1 response and followed by an involuntary switch of attention to the detected target. This attentional shift can be measured as a positive deflection in the ERP called 'P3a' and is larger for new or rare sounds than for small changes in the stimulus (Berti, 2012; Escera et al., 2014; Rinne, Särkkä, Degerman, Schröger, & Alho, 2006). If the attentional shift is conflicting with the task at hand, it is followed by a reorienting negativity (RON) toward the relevant stream (Escera, Alho, Schröger, & Winkler, 2000; Schröger & Wolff, 1998). In some cases, changes in the acoustic scene occur without being noticed, which is called 'change deafness' (Vitevitch, 2003). The changes elicit middle-latency (Nb) or slow responses but do not result in an MMN or subsequent P3a response. In other words, they fail to activate the 'normal' change detection system (Puschmann et al., 2013; Sohoglu & Chait, 2016b).

In complex acoustic scenes, containing multiple sound sources, a process called 'auditory scene analysis' (ASA) takes place (Bregman, 1990; Gutschalk & Dykstra, 2014). Different sound elements are separated, based on their frequency, location, or other features; this is called 'segregation'. Isolated sounds showing similar acoustic features over time are then joined together to form auditory streams (integration) (Deike, Denham, & Sussman, 2014; Gandras, Grimm, & Bendixen, 2017). The mechanisms underlying ASA are activated automatically, whether listeners are directing their attention toward the auditory scene or not (Sohoglu & Chait, 2016a). Schröger and his colleagues have proposed a conceptual framework, the auditory event representation system (AERS), which considers predictive coding to be the underlying mechanism for both the formation of auditory streams and the detection of deviant sounds (Schröger et al., 2014). The auditory streams are based on the identified patterns of regularity, and any difference between the predicted and actual stimulus is flagged as deviant (Winkler et al., 2009). This regularity encoding occurs at different levels of the auditory pathway, including the auditory brainstem and auditory cortex, and is reflected by electrophysiological responses like the FFR, SSA, MLR and MMN (Escera, 2011).

1.1.2.2 Top-down auditory attention

The bottom-up processes linked to ASA, leading to a separate encoding of the attended and the unattended speech stream, can both precede and follow top-down attention. After the segregation and integration processes take place, a listener can direct attention to one stream, while ignoring the others. This so-called ‘foreground-background selection’ is attention-driven and is reflected by a long-lasting ERP (Gandras et al., 2017). The selective attention can then alter the organisation of sensory input in the early stages of acoustic processing. It changes the way streams are organized, enabling the automatic change detection system to identify deviant events, even if they occur in the unattended stream. This detection is reflected in the emergence of an MMN (Sussman, Ritter, & Vaughan, 1998).

An MEG study showed that concurrent auditory objects, like two speech streams, are separately encoded in the brain, allowing the listener to direct attention to one of them while suppressing the other (Ding & Simon, 2012b). This process is referred to as neural entrainment to the speech envelope, or envelope entrainment. Selective attention to a specific stream can trigger a selective gain mechanism in the auditory cortex (Kerlin, Shahin, & Miller, 2010). Thus, when multiple speakers are talking simultaneously, it is possible to focus on one talker and understand what he or she is saying, even if the background noise is much louder than the speech itself. This phenomenon is called the ‘cocktail party effect’ (Cherry, 1953). In a multi-talker environment, the auditory system manages to successfully encode the attended speech stream, even with a competing talker nearby. The neural representation of this attended stream is processed separately, regardless of any acoustic changes in the unattended stream (Ding & Simon, 2012a). The neural responses to the attended stream are enhanced, while responses to other sounds are suppressed (Christison-Lagay, Gifford, & Cohen, 2015; Ding & Simon, 2012b; Kong, Mullangi, & Ding, 2014). This selection results in increased stimulus processing speed (Folyi, Fehér, & Horváth, 2012).

Attention affects the emergence of AEPs. For example, researchers have identified an endogenous negative component (Nd) with an onset latency between 50 and 150 ms, which is linked to attention and is often referred to as ‘processing negativity’ (Näätänen, 1982; Power, Foxe, Forde, Reilly, & Lalor, 2012). For children, the Nd has longer onset and peak latencies than for adults, indicating that the attention allocation process is slower (Gomes, Duff, Barnhardt, Barrett, & Ritter, 2007). The Nd can overlap with the N1 wave, resulting in a larger N1 amplitude (Hansen & Hillyard, 1980), although both responses are independently generated (Alho, Teder, Lavikainen, & Näätänen, 1994; Michie, Bearpark, Crawford, & Glue, 1990). The Nd should not be confused with the MMN, which is mainly based on sensory memory and not caused by selective attention. The MMN is a pre-conscious, automatic process, whereas the Nd is the result of a post-conscious, controlled process (Jemel, Oades, Oknina, Achenbach, & Röpcke, 2003). Furthermore, in a complex acoustic environment with multiple sound sources, the AEP is organized differently compared to a single-source situation. This is reflected in a distinct negativity (N2d, 260 ms after stimulus onset) and an anterior contralateral subcomponent (N2ac, 360 ms after onset) in the difference waveform (Lewald & Getzmann, 2015).

1.1.3 Processing speech

Certain stimulus properties can activate specific processing mechanisms in the brain. Evidence suggests that some regions in the brain are more selective to vocalisations than to other sounds (Fukushima, Saunders, Leopold, Mishkin, & Averbeck, 2014; Ghazanfar & Eliades, 2014), although the localisation of these regions remains unclear (Bizley & Walker, 2009). This selectiveness, which can be seen in other animals as well, applies to complex auditory objects that are important to the individual, based on learning processes (Poremba, Bigelow, & Rossi, 2013). In humans, it applies to speech (Vouloumanos, Kiehl, Werker, & Liddle, 2001).

A well-known model describing the cortical processing of speech is the dual-stream model (Hickok & Poeppel, 2007). It splits the processing of speech into two pathways: the ventral stream, activating the right lexical conceptual representations, and the dorsal stream, linking the acoustic signals to the articulatory networks. The ventral stream is described as a bilaterally organised pathway, whereas the dorsal stream would be more left-hemisphere dominant. Other researchers have confirmed that speech is processed in both hemispheres, involving many different brain regions (de Heer, Huth, Griffiths, Gallant, & Theunissen, 2017). Some studies describe which brain regions are involved in specific parts of speech processing, like lexical or semantical processing (e.g. Steinschneider et al., 2014). There is evidence for an asymmetry between both hemispheres, based the so-called ‘right ear advantage’ (REA). This means that speech arriving at the right ear is processed faster than speech coming from the left. However, the effect can be modulated by attention, by memory demands, or by a change in the stimulus properties (D’Anselmo, Marzoli, & Brancucci, 2016; Hiscock & Kinsbourne, 2011; Hugdahl & Westerhausen, 2016).

1.1.3.1 External factors influencing speech perception

Certain environmental factors can influence the processing of speech. For example, the presence of background noise or a high amount of reverberation in the room can interfere with the speech signal, reducing the amount of information that reaches the listener. Of these two factors, the noise level is the most important in determining speech intelligibility (Bradley, Reich, &

Norcross, 1999). Researchers found that increasing reverberation or background noise decreased speech understanding. The presence of noise also increased listening effort (Picou, Gordon, & Ricketts, 2016) and affected the amplitude of AEP's (Billings, Tremblay, Stecker, & Tolin, 2009; Koerner & Zhang, 2015; Maamor & Billings, 2017).

The direction of the speaker, especially relative to noise sources, is an important factor too when considering speech understanding. Due to the head shadow effect, speech coming from the front is easier to understand than speech coming from the back. However, a head-orientation benefit can be achieved by partially turning away from the speech source (Grange & Culling, 2016). This benefit is due to spatial release of masking, allowing the listener to exploit binaural cues to enhance speech understanding (Culling, Hawley, & Litovsky, 2004). These binaural cues are based on the head-related transfer function (HRTF). By travelling around the head, sounds arrive in the opposite ear with a short delay and a lower amplitude. Additionally, each individual pinna has a different shape that influences the spectrum of the sound. The auditory system can detect and interpret these small interaural time (ITD) and level (ILD) differences to localise sounds. By spatially separating target and noise sounds, the ITD and ILD values increase, improving stream segregation and speech understanding (Dubno, Ahistrom, & Horwitz, 2002).

In addition to acoustical factors, some speaker-dependent characteristics can also have a major influence on speech perception. For example, speakers can have a high-pitched or low-pitched voice, which sounds either loud or soft, and they can articulate either well or poorly. Different studies have shown that voice characteristics, dialect, and even voice familiarity can affect speech understanding (Ericson, Brungart, & Simpson, 2004; Jacewicz & Fox, 2013; Johnsrude et al., 2013). Furthermore, speakers use prosody to clarify or emphasize the meaning of their sentences. Listeners use this prosodic information to interpret how words fit together in a sentence and which are the most important parts. Bögels and her colleagues found that erratic prosodic breaks can affect the processing of speech. This influence is worse in the case of superfluous prosodic breaks then when breaks are missing (Bögels, Schriefers, Vonk, Chwilla, & Kerkhofs, 2013). In addition, the ERPs evoked by the onset of syllables can be affected by prosody, both during early perceptual processing and during the following lexical retrieval (Breen, Dilley, Devin McAuley, & Sanders, 2014). Others found that the processing of accentuation interacts with the retrieval of information from long term memory (Li & Yang, 2013). Thus, a clear and consistent prosody might help listeners to understand the message.

1.1.3.2 Internal factors influencing speech perception

Apart from the external factors in the previous paragraph, speech perception is affected by individual differences in hearing abilities or cognitive skills as well. Listeners who are hearing-impaired or experience cognitive difficulties perform worse than listeners with normal hearing, especially in noise. Age may act as a mediating factor, as both hearing and cognitive abilities decline with age. For instance, stream segregation decelerates with aging listeners, which is reflected by increased MMN latencies for concurrent speech streams (Getzmann & Näätänen, 2015). Elderly listeners also show reduced MMN amplitudes (Woods, 1992), along with delayed attentional control and reduced speech processing (Getzmann, Wascher, & Falkenstein, 2015). Furthermore, they expend more listening effort than younger listeners when listening to speech in noise (Anderson & Gagné, 2011). This is probably due to a reduced working memory capacity and a lower processing speed in elderly listeners (Desjardins & Doherty, 2012; Zekveld, Rudner, Johnsrude, & Rönnberg, 2013).

Individual experience and training may also influence speech perception. Researchers have found that understanding speech in noise is harder for bilingual than for monolingual listeners (Tabri, Chacra, & Pring, 2011). In addition, bilingual listeners achieve better performance when listening to a speaker with an accent similar to their own, although this effect interacts with experience (Pinet, Iverson, & Huckvale, 2011). Musicians, on the other hand, show improved speech understanding in noise compared to listeners without musical training (Coffey, Mogilever, & Zatorre, 2017), possibly due to enhanced phonological representations and functional connectivity in the brain (Du & Zatorre, 2017). This auditory processing benefit shows as soon as at primary school level (Habibi, Cahn, Damasio, & Damasio, 2016; Strait, Parbery-Clark, Hittner, & Kraus, 2012).

1.1.4 Hearing difficulties

1.1.4.1 Hearing loss

Many things can go wrong in the auditory system. In the outer ear, a malformation of the ear canal or an excessive amount of earwax may obstruct the way for sound waves. In the middle ear, fluid can accumulate, reducing the movement of the tympanic membrane, or a congenital defect of the ossicles can block the transmission of the sound waves to the inner ear. On the level of the cochlea, congenital abnormalities or noise damage can impair the working of the hair cells. Any of these problems can result in hearing loss (Hearing Loss Association of America, 2017).

Hearing loss can be either conductive, when the problem lies in the outer or middle ear, or sensorineural, if something is wrong with the inner ear or with the auditory nerve. If the hearing loss is caused by problems in both the outer or middle and the inner ear, it is referred to as a mixed hearing loss. Hearing loss can also result from a damaged or missing auditory nerve. In this case, the patient suffers from retro-cochlear hearing loss (MED-EL, 2017; Sataloff, Sataloff, & Vassallo, 1980). When the hearing threshold of a patient is elevated by 25 dB or more, the patient is diagnosed with a hearing loss in that ear. If the hearing loss in the better ear exceeds 40 dB for adults or 30 dB for children, it is referred to as a disabling hearing loss. Worldwide, more than 360 million people or 5% of the global population suffers from a disabling hearing loss (WHO, 2017).

1.1.4.2 Hearing aids

Hearing loss can be traced back to different problems, so different solutions exist as well. The most well-known type of hearing aid (HA) consists of four parts: a microphone, an amplifier, a receiver (or loudspeaker), and a battery. By leading a signal through this HA, sounds are amplified and presented at a higher intensity level. The amplification level is adjusted separately for each frequency band, depending on the hearing difficulties of the individual patient. Often, the amplifier also uses digital signal processing to improve the sound. This ‘basic’ HA technology exists in devices of different sizes, ranging from a relatively large body-worn HA to one that completely fits in the ear canal. These aids can be used by patients with different types of hearing loss, like congenital hearing loss, presbycusis (hearing loss caused by aging) or noise-induced hearing loss (Dillon, 2012; Hougaard et al., 1995; Rodenburg, Huizing, Kapteyn, & Wanink, 1979).

Next to the ‘standard’ device, as described above, other types of hearing aids are available. For example, problems in the outer and middle ear can be bypassed by using a bone conduction device (BCD) or a middle ear implant (MEI). With this type of HA, an implant presents the sound by vibrating either the bone of the skull itself, the ossicles, or the oval window. Afterwards, the cochlea processes the sound in a normal way. A BCD or MEI helps patients with a variety of problems, like a large conductive hearing loss, recurring middle ear infections, or a malformation in the middle or outer ear (Dillon, 2012).

Finally, a cochlear implant (CI) overcomes sensorineural hearing loss. A CI converts sound waves into electrical signals, directly stimulating the hair cells inside the cochlea. This stimulation results in the excitation of the auditory nerve and sends a signal to the brain. However, this signal inevitably contains less information than a signal originating from a normal cochlea. For example, while healthy listeners can use thousands of inner hair cells to cover the entire audible frequency range (roughly from 20 Hz to 20 kHz), the CI reduces the signal to a limited number of frequency bands, each activated by one electrode. Thus, the accurate tonotopy of the auditory system, enabling listeners to hear the exact pitch of a sound, is lost. CI users perceive different pitches, but only a limited number (Laneau & Wouters, 2004). Despite the obvious shortcomings, this type of HA is often the only solution for patients with severe hair cell damage or congenital deafness. However, if a patient suffers from retro-cochlear deafness, a CI is not an option either, as the problem originates from the connection to the brain itself.

1.1.4.3 Problems associated with hearing aids

Despite the recent tremendous improvements on HA technology, there is still a lot of work to do. HA users have reported multiple issues involving the adjustment to their HA (Bennett, Laplante-Lévesque, Meyer, & Eikelboom, 2017; McCormack & Fortnum, 2013). Inadequate instructions received from caregivers present a major problem. A lot of information needs to be conveyed, and it can take some time for new users to get accustomed to the HA (Dawes & Munro, 2017). Adequate instructions and a well-organised follow-up system could address this problem (Solheim, Kværner, Sandvik, & Falkenberg, 2012; WHO, 2004). Furthermore, many reported problems relate to wrong expectancy patterns of HA candidates. They assume the HA will substitute the normal hearing most of them once had, obviously a wrong assumption. Other researchers confirm the influence of expectancy patterns on general HA satisfaction (Dashti, Khiavi, Sameni, & Bayat, 2015; Meyer, Hickson, Khan, & Walker, 2014). The third major issue concerns the HA itself. Users sometimes find it too small to handle, they experience difficulties when changing the batteries or operating the buttons, or they forget to replace the batteries causing the HA not to work properly. Finally, and despite many breakthroughs in HA research, the HA does not perform well in difficult situations. For example, speech perception is difficult in the presence of background noise, music often sounds distorted, and directional hearing can be very challenging.

One of those difficulties is linked to the way a HA handles noise. At the earliest stage of HA developments, the devices aimed at suppressing noise by maintaining level differences between background and nearby sounds. Later, a newly developed system selectively suppressed low frequency sounds with high intensities. Now, noise suppression is mainly regulated by complex processing algorithms analysing the acoustical environment and adaptively suppressing the noise. Unfortunately, the use of noise-reduction algorithms also reduces speech intelligibility in some listening environments (Hilkhuysen, Gaubitch, & Huckvale, 2013; Hu & Loizou, 2007), by eliminating important acoustic cues. For example, it was found that a commonly used noise suppression algorithm, based on adaptive directional microphones, could distort the binaural information in sounds (Van den Bogaert, 2008). Various solutions have been proposed to enhance speech intelligibility. In an experimental design based on multichannel Wiener filters, including a communication link between both HA's, better noise reduction and binaural cue preservation was achieved (Van den Bogaert, Doclo, Wouters, & Moonen, 2009). Another study used a trainable algorithm that adapted to the preferences of each individual user, based on the preferred level of amplification and the degree of beamforming, for different acoustic conditions. Given the variation in preferred settings between subjects, the use of trainable algorithms may improve individual HA satisfaction (Yoon et al., 2017). In CI users, researchers successfully improved speech intelligibility by stimulating only those channels with a positive signal-to-noise ratio (SNR) (Hu & Loizou, 2010).

Apart from risking distortion of the speech signal during noise suppression, the current algorithms face another important problem: the identification of noise. When the sound is a common type of noise, like a car engine or a ventilation system, suppression mechanisms often succeed. However, in ambiguous situations where the distinction between 'noise' and 'target' is unclear, like when two people are talking simultaneously, the algorithm could accidently suppress the wrong part of the acoustic environment. The inconsistency in the classification of noise makes it a difficult problem to tackle. In these ambiguous situations, one possible solution is the use of intelligent enhancement, which could be accomplished by using a brain-computer interface (BCI). A BCI steers the noise suppression algorithms towards the correct interpretation of sounds by using real-time brain activity. Determining which part of the auditory environment the user is listening to, and considering this part as 'target', can be achieved through auditory attention detection (AAD).

1.1.5 Summary

The processing of auditory information is organized in an ascending auditory pathway. The successive steps are reflected in AEPs at different times, from fast (0-20 ms) to slow (80-500 ms) responses. When listening to a complex stimulus, the auditory system uses regularity encoding to divide the auditory scene into separate streams (ASA). Using predictive coding, the system can also apply the information about regularities to detect changes. Both ASA and the deviance detection system can lead to bottom-up attention: redirecting the focus of the listener based on the characteristics of a stimulus. In the case of top-down attention, the focus change results from a conscious decision. This selective attention influences the cortical organisation of the streams, enhancing the chosen stream and suppressing the others, which is reflected in the electrophysiological responses as well. During the processing of speech, a dedicated system is activated, which is assumed to be left-hemisphere dominant. Specific characteristics of the acoustical environment, the speech signal, or the listener can influence speech perception. For patients with hearing loss, different types of hearing aids are available, depending on the cause and severity of the hearing loss. Unfortunately, hearings aids cannot replace a normal auditory system, which is why many HA users experience difficulties in everyday listening situations.

1.2 Auditory attention detection

This section tackles auditory attention detection (AAD). First, the concept of AAD and the underlying principles are explained. The second paragraph deals with recent findings in AAD research. Finally, the future use of AAD is discussed.

1.2.1 Detecting attention

As auditory stimuli elicit neural responses, the processing of these stimuli can be studied using brain imaging methods. Most of these studies use electro-encephalography (EEG), magneto-encephalography (MEG) or electrocorticography (ECoG). Because the neural representation of stimuli is stronger for stimuli receiving selective attention, brain activity can be recorded to achieve auditory attention detection (AAD). In other words, the analysis of a person's brain activity can help to determine what part of the stimulus he or she is paying attention to.

When choosing a method to record brain activity for AAD purposes, the temporal resolution, staying within the range of milliseconds, plays the most important role. Three popular methods in AAD research are EEG, MEG, and ECoG. In EEG, on one hand, surface electrodes measure the electric activity of the brain. It is a non-invasive method with a high temporal resolution and a relatively low cost. The location of the electrodes is based on international placement systems. MEG, on the other hand, measures the magnetic fields generated by brain activity. MEG is a non-invasive method with a high temporal resolution as well, and it has a better spatial resolution than EEG. However, the method is more expensive than EEG and it is very sensitive to external noise, making it less suitable for everyday use. It also relies on heavy equipment, which is impractical for BCI purposes. Finally, ECoG is an invasive method placing electrodes directly on the brain surface, which means it can only be performed during brain surgery. Compared to EEG, the spatial resolution is better and it results in a clearer signal, but it is unsuitable for the use in healthy patients (Bera, 2015; Srinivasan & Nunez, 2012). As every method has it shortcomings, careful selection based on the specific research goals is recommended (Lee, Larson, Maddox, & Shinn-Cunningham, 2014). Most AAD research is based on EEG and MEG recordings.

After recording the brain activity, for example using EEG, the information is analysed with a so-called 'decoder'. Two types of decoders exist: subject-specific decoders, which are trained using data from a single subject, and pre-trained decoders, using combined data from different subjects. If less than 15 minutes of data are available for each subject, a pre-trained decoder often shows the highest accuracy (Mirkovic, Debener, Jaeger, & De Vos, 2015). To train a decoder, a large amount of data is required. These data include EEG output as well as the stimuli used during the EEG recording. As the auditory processing mechanisms result in a slight delay in the EEG signal, a time lag of typically 200-250 ms is added to the speech signals. The decoder learns to use the EEG data to reconstruct the envelope of the attended speech. To this end, it attributes a relative weight to each EEG channel, adjusting the different weights to maximise the correlation between the reconstructed envelope and the attended speech. Ultimately, a set of decoder weights is formed, which helps the decoder to decode a new trial. By feeding the EEG data of the new trial into the decoder, an estimated envelope is formed. The reconstructed envelope can then be compared to both the attended and unattended speech stream and a decision can be made based on the highest correlation. The corresponding stream is then labelled as the 'attended' speech. If the algorithm correctly identifies the attended talker, the decoding process is considered successful. The decoding accuracy indicates the percentage of correct decisions by the decoder (O'Sullivan et al., 2015).

Most decoders are trained using the leave-one-out method. For a subject-specific decoder, this means a new set of decoding weights is calculated for each trial of the subject. Thus, in a data set containing 100 trials, the decoding weights for the first trial are created using trial 2 to 100. For the analysis of the 100th trial, the decoding weights are based on trial 1 through 99. Using this approach, the decoding weights vary slightly across trials. For a pre-trained decoder, which is used across subjects, the leave-one-out method means the decoding weights are based on the data of all subjects but one, before analysing the data of the remaining subject.

1.2.2 Status of AAD research

Since the first attempts to detect auditory attention, researchers have applied AAD in different situations, using a wide variety of stimuli and paradigms. Taking a closer look at their results shows that research choices can significantly influence AAD performance. For example, a study comparing different envelope extraction methods found relatively large differences in mean decoding accuracy (74.5-87.5%) between various extraction methods and recording times

(Biesmans et al., 2015). Choosing generic or subject-specific decoders can affect decoding results as well (e.g. 76.0-87.2%, in Das, Biesmans, Bertrand, & Francart, 2016). Consequently, it is important to consider the applied methods when designing AAD experiments and interpreting the results.

In general, however, AAD can achieve robust results. While many experiments apply 64- or 128-channel EEG, recent studies obtained acceptable decoding accuracies using down to 25, 10 or 8 channels (Fuglsang, Dau, & Hjortkjær, 2017; Mirkovic et al., 2015; Zink, Baptist, Bertrand, Huffel, & Vos, 2016). Other researchers have experimented with shortened recording times, proving that 30 or even 10 seconds of data was enough to accurately identify the attended talker (Fuglsang et al., 2017; Zink et al., 2016). One study tested even shorter sample lengths, confirming that the accuracy increased with longer EEG recordings (Horton, Srinivasan, & D'Zmura, 2014).

When attempting to apply AAD to more realistic listening conditions, the classic paradigm using only two competing speech streams is insufficient. Most everyday environments contain multiple sound sources, including different speech streams and interfering noise sources, sometimes with additional distortions due to reverberation. It is known that factors like background noise affect the auditory processing mechanisms (Koerner & Zhang, 2015). Therefore, simply assuming that the experimental results from basic acoustic environments can be generalized to these situations would be a mistake. To meet these objections, recent studies incorporate more realistic conditions, by using lifelike stimuli or adding noise sources to the listening environment.

For example, Das and her colleagues presented both dichotic and HRTF-filtered stimuli, finding that the latter resulted in higher AAD performance. The HRTF-filtered presentation was more realistic, which may have led to more efficient processing by the auditory system. Furthermore, as the speech streams were mixed in the HRTF-filtered condition, the separate streams were less clear, thereby reducing speech intelligibility. The authors believe this increased difficulty triggered selective gain mechanisms to enhance the representation of the attended stream and suppress the unattended stream, explaining the higher decoding accuracies (Das et al., 2016). A similar explanation was given by other researchers, who investigated the effect of different reverberation levels and that of multiple talkers. They found noise-robust cortical tracking of the attended speech stream, while the representation of the unattended stream deteriorated when reverberation or babble noise was added (Fuglsang et al., 2017).

1.2.3 Possibilities of AAD

The development of AAD technology has the potential to solve some major issues in the field of audiology. For example, speech intelligibility can at present only be measured using a behavioral task, with listeners repeating the word or sentence they heard. However, this method is not usable with all patients, because it requires a high amount of attention and cooperation. Meeting these limitations, measures of envelope entrainment have been studied as a valid alternative to standard speech perception tests (Commers, 2017; Goris, 2016). Recently, researchers successfully connected neural speech representations to behavioral intelligibility measures using EEG (Vanthornhout, Decruy, Wouters, Simon, & Francart, 2018).

AAD is likely to solve other problems as well. As mentioned before, the current hearing aid technology encounters major issues in the way noise suppression algorithms work. Starting from an analysis of the acoustic environment, the algorithm can successfully suppress stationary noise. But in a more complex situation, involving competing talkers as 'noise', the system might suppress the wrong talker. However, if the EEG of HA users could be recorded and analysed in real time, determining which speech stream they are listening to, the algorithm could use this information to decide what parts of the auditory scene should be considered 'target' or 'noise'. Subsequently, the software could selectively amplify the attended speech stream, mimicking the attentional gain control in listeners with normal hearing (Kerlin et al., 2010). This type of brain-computer interface (BCI) can improve speech perception in difficult hearing situations, like a conversation at a dinner table or a crowded party.

Unfortunately, AAD technology is not yet ready for everyday use. Whereas most research is based on at least 64-channel EEG measurements and compares the reconstructed envelope to the clean speech signals, these conditions are unrealistic when considering everyday use of AAD. To implement AAD in HA software, brain responses should be measured and decoded instantaneously instead of afterwards, using a portable system instead of a traditional EEG set-up, and based on the actual acoustic signal instead of isolated streams. In view of these discrepancies, recent studies have investigated more practical and realistic applications of AAD technology. For example, Mirkovic and her colleagues tested a small EEG electrode grid that can be worn behind the ear (Mirkovic, Bleichner, De Vos, & Debener, 2016). Other researchers decoded the EEG using only mixed speech recordings, comparable to what a HA could register in real-time (Van Eyndhoven, Francart, & Bertrand, 2017). At the same time, possibilities for real-time decoding of the EEG are being explored (de Souza Ranaudo, de Sá, & Felix, 2012), along with real-time modification of the speech streams, simulating the effect of selective amplification by AAD-steered HA software (Bau, 2016).

Clearly, there is still a lot of work to be done before AAD can be implemented in everyday HAs. Since the ultimate goal of this implementation is to improve HA performance, by enabling adequate noise suppression and enhancing the attended speech stream, additional research is needed. For instance, when adjusting the gain for each stream to improve speech perception, it is important to keep the suppressed streams loud enough to allow successful attention switching. Yet it is currently unclear how large the gain difference should be to optimise speech understanding while leaving open the possibility to redirect attention. Furthermore, although AAD is mostly tested on young adults with normal hearing, little or no research is available for hearing impaired listeners or elderly people. However, given the age-related decline in cognitive skills, and the effect of hearing impairment on the processing of sounds, these populations need to be included in future AAD research.

1.2.4 Summary

In 'auditory attention detection' (AAD), electrophysiological correlates of attention are used to predict which speech stream a person is attending to. To this end, brain activity is recorded using brain imaging techniques like EEG and analysed by decoders. AAD can be used to successfully identify the attended speaker in different listening environments, although additional research is needed. Ultimately, BCI interfaces may improve HA performance by enabling selective gain adjustments for attended speech streams.

1.3 Research questions

This thesis investigates some factors that could influence the decoding accuracy when performing AAD with 64-channel EEG, in a two-talker scenario with background noise. The first factor is the position of both speakers, relative to the listener. To this end, four angular conditions are used, in which the attended and unattended speaker are each coming from different angles. The second factor is the amount of noise in the acoustic environment. To investigate this factor, babble noise is presented at different noise levels. Next, the interaction between both factors is investigated, along with individual differences in AAD performance. Finally, the link between speech intelligibility and decoding accuracy is explored.

1.3.1 Research question one: spatial separation

What is the effect of spatial separation on AAD?

a. How does the angular separation between speakers influence the effectiveness of AAD?
b. Is there evidence for a right ear advantage in AAD performance?

Spatial separation may influence the effectiveness of AAD. In general, a larger angle difference between both speakers might be linked to better AAD performance. Based on the concept of REA, overall performance might also be better if the attended speaker is situated at the right side of the listener. However, as REA findings so far are based on dichotic listening experiments (Hugdahl & Westerhausen, 2016), it may not impact the decoding accuracy in a diotic experiment paradigm.

1.3.2 Research question two: background noise

How does the presence of babble noise influence AAD? Is there a difference in AAD performance for different noise levels?

The presence of noise is expected to have a negative influence on AAD performance, with higher noise levels resulting in lower accuracy. The overall decoding accuracy is expected to be the highest when there is no noise present.

1.3.3 Research question three: interaction between separation angle and noise

How do speaker separation and noise levels interact when influencing AAD?

The level of background noise might interact with the angle difference between both speakers. Background noise may have different effects, depending on the separation angle.

1.3.4 Research question four: individual differences

How important are individual differences between subjects?

a. Are there individual differences in decoding accuracy?
b. Can the subjects' decoding accuracies be linked to their individual performances on a speech-in-noise task?

Given the differences between individuals, both in listening strategies and in physiological factors, a significant variability in AAD performance across subjects is expected. Additionally, subjects with good speech-in-noise results (i.e. low SNRs) on the speech-in-noise task might have higher overall decoding accuracies.

1.3.5 Research question five: intelligibility versus decoding accuracy

To what extent is the AAD performance connected to the estimated speech intelligibility?

a. Does spatial separation influence the subjective intelligibility?
b. What is the effect of background noise on the estimated speech intelligibility?
c. Do spatial separation and background noise interact when influencing the intelligibility?
d. How important are individual differences for the subjective intelligibility?
e. Is there evidence for a connection between the intelligibility ratings and the corresponding decoding accuracies?

The experienced speech intelligibility by listeners does not necessarily correlate with the actual AAD performance, since the intelligibility rating is a subjective measure, while the decoding accuracy is based on objective results. Estimated intelligibility is expected to be the lowest for a 10° separation angle and the largest for a 180° angle. Furthermore, the presence of background noise may affect speech intelligibility, with more noise resulting in lower estimated intelligibility rates. Both factors might interact when influencing the subjective intelligibility, in the same way or differently than for the decoding accuracy. There may also be individual differences in subjective intelligibility. Additionally, the intelligibility may be linked to the individual's threshold on a speech-in-noise task, with lower thresholds going together with higher intelligibility ratings. Finally, a correlation between intelligibility and AAD performance seems plausible; however, this correlation may be weak because understanding speech is no prerequisite for a cortical representation of the speech stream.

2 Method

The experiment sessions were scheduled between October 12th and November 16th, 2017. All sessions took place in the audio cabins in the Experimental Oto-Rhino-Laryngology (ExpORL) office of KU Leuven.

2.1 Participants

A total of 15 subjects were recruited. All participants were native Flemish speakers between 20 and 25 years old (range 20;01-25;02 yrs., mean age 22;05 yrs.). They received no financial reward for their participation. All subjects declared they had normal hearing, which was confirmed by calculating pure tone averages (PTAs) for both ears. In addition, their speech reception threshold (SRT) for Matrix-sentences in babble noise had to be below -7 dB SNR to participate. In Table 1, the age, gender, hearing thresholds, and SRT of each subject is displayed.

Table 1. *Subjects.*

Number	Age	Gender	PTA (left)	PTA (right)	SRT
1	20;01 yrs.	female	10 dB HL	12 dB HL	-8,6 dB SNR
2	23;07 yrs.	female	5 dB HL	12 dB HL	-8,8 dB SNR
3	22;09 yrs.	female	5 dB HL	8 dB HL	-9,0 dB SNR
4	21;10 yrs.	female	3 dB HL	5 dB HL	-9,1 dB SNR
5	22;06 yrs.	female	2 dB HL	2 dB HL	-8,8 dB SNR
6	20;11 yrs.	female	7 dB HL	7 dB HL	-7,2 dB SNR
7	23;01 yrs.	female	8 dB HL	5 dB HL	-7,7 dB SNR
8	21;00 yrs.	female	12 dB HL	8 dB HL	-9,8 dB SNR
9	20;04 yrs.	female	7 dB HL	2 dB HL	-8,9 dB SNR
10	22;04 yrs.	female	2 dB HL	0 dB HL	-8,8 dB SNR
11	22;04 yrs.	female	12 dB HL	10 dB HL	-7,9 dB SNR
12	24;09 yrs.	female	-4 dB HL	-4 dB HL	-8,1 dB SNR
13	25;02 yrs.	male	5 dB HL	7 dB HL	-9,1 dB SNR
14	22;05 yrs.	male	10 dB HL	3 dB HL	-8,1 dB SNR
15	24;00 yrs.	female	5 dB HL	10 dB HL	-9,3 dB SNR

2.2 Equipment

2.2.1 Stimuli

2.2.1.1 Pure tone audiometry

Standard sinusoidal tones of different frequencies were used for obtaining pure tone thresholds. The tested frequencies included 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, and 8000 Hz.

2.2.1.2 Speech recognition test

The speech stimuli for the speech-in-noise test consisted of standard Matrix test sentences, read by a female talker (Houben et al., 2014). The noise was made up of two parts: a competing talker (CT), narrating a story, and babble noise. Thus, it resembled the noise used in the AAD experiment.

2.2.1.3 AAD in noise experiment

The stimuli consisted of Flemish short stories, read by female speakers. All stories were adapted from *www.radioboeken.eu* and are listed in Table 2. After processing a part of the data, it became clear that one story, 2a ('De gamba'), had systematically lower decoding accuracies than the other stories. Therefore, it was replaced with a different story, 2b ('In de zon kijken'), which was used with subjects 12-15 only. Unfortunately, it became clear afterwards that the story chosen to replace it resulted in lower performance as well. A comparison between the accuracies for each story can be found in Appendix A. In a prior run of the experiment, some stories had to be excluded too, because of low intelligibility results, the extensive use of difficult or uncommon words, or non-Flemish parts in the story.

Table 2. *Stories*.

Number	Title	Author
1	Honing	Kristien Hemmerechts
2a	De gamba	Thomas Gunzig
2b	In de zon kijken	Anne Provoost
3	Een krokodil aan de tong trekken	Saskia De Coster
4	Lena	Charlotte Therssen
5	Het bestaat	Annelies Verbeke
6	Het meisje en de kat	Rachida Lamrabet
7	De tuin	Kamiel Vanhole
8	De volle schort	Diane De Keyzer

The stories were presented at 65 dB SPL. The presentation of each story was split up into four parts, hereafter called 'clips', lasting 2 to 5 minutes each. The root mean square of intensity (RMS) was normalized per clip. Silent gaps exceeding 300 ms were truncated to 300 ms. Each story was paired with another story and used twice: once as attended speech and once as unattended speech. Both stories were presented simultaneously with varying angle differences, using a head-related transfer function (HRTF) to make the listening task more realistic (Das et al., 2016). In addition to a competing talker (CT), babble noise was used in some conditions. The noise consisted of 36 speakers, divided over 9 directions (4 talkers each) using HRTF. The spectrum of the babble noise had its spectrum matched with that of the average spectrum of all the speakers. The power of each babble source was balanced so that the SNR equalled the power of a speaker, divided by *N* times the power of a babble source, with *N* = 36 (number of speakers in the babble).

Three different SNRs of varying speech intelligibility were used when babble noise was presented. The chosen levels were based on the results of earlier speech intelligibility behavioral tests. The story speech material from both the attended speaker and the CT was presented at SRT_{50}, SRT_{50} + 3 dB and SRT_{50} + 6 dB. The estimated value for SRT_{50} (speech reception threshold) was based on subjective speech intelligibility ratings for stories, averaging -8,470 dB SNR for the angular setup [-90°, 90°]. Based on previous research, a conversion term of +1,363 dB was added to account for the difference between objective and subjective SRT values, calculated from comparable measures using Matrix sentences. This process resulted in an estimated objective SRT_{50} value of -7,107 dB SNR, referred to as SNR3. For SNR2, 3 dB was added to improve speech intelligibility, so speech was presented at -4,107 dB SNR. The easiest noise condition, SNR1, used a signal-to-noise ratio of -1,107 dB SNR.

2.2.2 Software and hardware

For the pure tone audiometry, the equipment consisted of standard headphones and an Orbiter 922-2 audiometer (Madsen Ltd.). The PTA's were calculated manually. The SRT assessment used E-A-R-TONE 3A insert earphones (3M) and a Hammerfall DSP Multiface II (RME). The user interface for the speech recognition test was constructed using APEX software (Francart, 2008), which also computed the SRT (in dB SNR) after each list.

For the AAD experiment, the complete BioSemi ActiveTwo-system was used (BioSemi, 2002). Brain potentials were measured using two sets of pin-type active electrodes, with 32 electrodes each. Additionally, a CMS (common mode sense) and DRL (driven right leg) electrode formed a feedback loop, replacing the typical ground electrode of other EEG systems. All electrodes were fitted in 10/20-labelled head caps, using Signa Gel (Parker Laboratories) to connect the electrodes to the scalp. During the experiment, the sensor-signals were digitized with a 24-bit resolution, using an ActiveTwo AD-box. The optical data were converted to USB output by the USB2 Receiver, which also controlled the input the subject received. Finally, ActiVIEW software (BioSemi, 2016) displayed the signals from each of the 64 channels, while indicating the presence of auditory input with triggers. Thus, the researcher could track the progress of each story part.

The EEG was split up into 30-second intervals afterwards and decoded on a subject-specific leave-one-out basis, following the method described by Biesmans using MATLAB software (Biesmans, Das, Francart, & Bertrand, 2017). In addition to the ActiveTwo set-up, the audio cabin was equipped with a screen, a computer mouse, and a microphone. This set-up allowed the subject to read instructions, answer questions, control the beginning of each trial and communicate with the experiment leader. The user interface was again constructed using APEX software and stimuli were presented through ER-1 insert earphones (Etymotic Research).

2.3 Research method

After receiving information about the goal and progress of the experiment, subjects signed two informed consent forms. Both forms had been approved by the KU Leuven ethical committee.

2.3.1 Pure tone audiometry

To verify that all subjects had normal hearing, a standard pure tone audiometry was performed. Subjects were seated in a soundproof booth. They were fitted with headphones and asked to raise their hand when they heard a sound. Pure tone air conduction thresholds were recorded using the Hughson-Westlake method. Afterwards, the PTA was calculated as the average threshold for the 500, 1000 and 2000 Hz tones. Subjects with a PTA above 20 dB HL were excluded from the rest of the experiment. Subjects with normal PTAs but high thresholds on specific frequencies underwent additional otoscopy and tympanometry to confirm normal hearing.

2.3.2 Speech recognition test

In the same soundproof booth, subjects were seated in front of a computer screen. After receiving verbal instructions about the test, they were fitted with insert phones. Through the insert phones, subjects were presented with both CT (left ear) and babble noise. The target speech, consisting of Matrix sentences, was presented to the right ear at an initial SNR of -15 dB. After each sentence, subjects had to select the correct words on the screen. Using an adaptive procedure, the SNR was varied to obtain the SRT, which is the SNR level at which the subject could understand 50% of the target speech. Each subject completed at least two training lists and one test list, taking the learning effect for SRT's in fluctuating noise into consideration (Rhebergen, Versfeld, & Dreschler, 2008).

2.3.3 AAD in noise experiment

2.3.3.1 Preparation

Subjects were seated on a chair outside the audio cabin. After measuring their head circumference, a BioSemi head cap was placed on their head. The cap was positioned according to the international 10/20-system, by placing the Cz electrode in the middle between the nasion (bridge of the nose) and the inion (midline occipital protuberance) and between both preauricular points. Consequently, a conductive gel was injected in the electrode gaps on the cap and the electrodes were clicked in place.

After fitting the electrode cap, subjects were positioned in a comfortable chair in an electromagnetically shielded, sound proof room. After connecting the electrodes to the computer interface, the impedance of each electrode was verified, and some extra gel was added if necessary. Subjects were fitted with insert phones, which were clipped to the chair for stability, and received instructions about the task at hand.

2.3.3.2 AAD in noise

When the set-up was completed, subjects were presented with some basic instructions on the screen. Next, they could see a diagram indicating the angular position of two speakers: one of them green, the other red. They were requested to focus their attention on the story coming from the green speaker while ignoring everything else. First, they listened to a 5-second test scenario with only the green speaker, to direct their attention to the right angle. They were then sequentially presented with four trials with different SNRs, each comprising one out of four successive story clips. After each trial, subjects were asked to solve a multiple-choice question about the story part they had just listened to. The questions encouraged them to focus on the task, but no feedback was given about their responses. Subjects also had to estimate their speech intelligibility rate, answering the question "How many percent of the words did you understand correctly (e.g. 27%)?". They were presented with an illustration of a ruler ranging from 0 to 100, on which they could mark the estimated percentage. Finally, a screen indicated that subjects could make remarks or ask questions. The experiment leader could hear them through a microphone, installed next to the subjects, and answer by typing directly on the screen. If everything was well, subjects could proceed to the next trial when ready.

The experiment consisted of four angular conditions: -5° and +5° (a); +30° and +90° (b); -30° and -90° (c); -90° and +90° (d). There was one 'test block' for each angular condition. A test block consisted of two parts: in the first part, the attended talker came from one direction, while in the second part, the subject had to focus on the talker coming from the other direction. Each part started with a 5-second test scenario, followed by four trials: one without babble noise and three with varying SNR levels. After four trials, the story was finished, and the subject was instructed to switch attention to the other talker. At the end of each test block, there were two additional trials, which were in fact repetitions of the trials without babble noise from the first and second part. For these additional trials, participants did not have to answer any questions.

The presentation order was randomised between subjects for the test blocks (angular conditions) and the trials containing babble noise (SNR1, SNR2 and SNR3). More information on this randomisation process can be found in Appendix B. The duration of the trials varied over the angular conditions. Quiet trials took 4 minutes in condition (a) and (d), while lasting only 2 minutes in condition (b) and (c). Noise trials took 5 minutes or 2,5 minutes respectively.

A typical angular condition (test block) looked like this:

Example: angular condition a											
direction	-5°	-5°	-5°	-5°	-5°	+5°	+5°	+5°	+5°	+5°	-5°/+5°
session	test	quiet	SNR1	SNR2	SNR3	test	quiet	SNR1	SNR2	SNR3	quiet
duration	5′′	4′	5′	5′	5′	5′′	4′	5′	5′	5′	8′

Example: angular condition c											
direction	-30°	-30°	-30°	-30°	-30°	-90°	-90°	-90°	-90°	-90°	-30°/-90°
session	test	quiet	SNR1	SNR2	SNR3	test	quiet	SNR1	SNR2	SNR3	quiet
duration	5′′	2′	2,5′	2,5′	2,5′	5′′	2′	2,5′	2,5′	2,5′	4′

2.4 Data-analysis

2.4.1 Available data

Data were recorded for 64 channels, with a sample rate of 8192 Hz. The trials lasted 138 minutes in total, thus resulting in 4,341 x 10^9 data points for each subject. For analysis purposes, each trial was split up in a series of 30-second intervals. The intervals were bandpass-filtered between 1 to 9 kHz before analysing them separately, resulting in 276 decoding results (one for each 30-second interval) for each subject.

2.4.2 Analysis method

Data were analysed using the statistical software *R* (R Core Team, 2017) and *RStudio* (RStudio Team, 2016). The analysis methods were selected using the decision tree from *Discovering statistics using R* (Field, Miles, & Field, 2012). The assumptions for parametric tests are discussed in Appendix C.

2.4.2.1 Linear mixed-effects model for decoding accuracy

Based on the hypothesis, AAD performance might be affected by speaker positions and background noise. To investigate the effect of these factors, a linear mixed-effects model (LME) was constructed. This model incorporated separation angle and noise level as fixed factors, along with the interaction between them. To account for between-subject variability, subject was added as a random effect.

2.4.2.2 Post-hoc analyses

Based on significant effects, indicated by an ANOVA of the LME model, the performance difference within the corresponding variables was confirmed using Friedman's ANOVA. Next, Wilcoxon signed-rank tests were carried out to further explore the differences. To correct for multiple post-hoc analyses, adjusted p-values were calculated using the Benjamini-Hochberg correction. Thus, post-hoc analyses were carried out for separation angle (10°, 60°, or 180°) and noise level (∞, -1.1, -4.1, or -7.1 dB SNR). Additionally, the mean accuracy for each side of attended speaker (left or right) was compared using another Wilcoxon signed-rank test.

2.4.2.3 Analysing the effect of individual differences

To answer the fourth research question, considering individual differences in AAD performance across subjects, a Kruskal-Wallis test was performed. This test assigned ranks to all the accuracy scores in a data set before adding them together for each subject. Next, the total rank was compared across subjects. Additionally, Kendall's tau was used to see if the mean AAD accuracy for each subject could be linked to their performance on the speech-in-noise task.

2.4.2.4 The relationship between speech intelligibility and AAD performance

The final research question focussed on the relationship between the subjective speech intelligibility and the measured AAD accuracy. First, the effect of spatial separation and background noise on the estimated intelligibility was studied. To this end, the LME model was reconstructed, but this time with the estimated speech intelligibility as predicted variable. Second, the correlation between the intelligibility and the accuracy was calculated using Kendall's tau. The analyses for individual differences, as described above, were repeated for intelligibility ratings as well.

3 Results

The third part of this thesis contains the results of the statistical analyses. All analyses are described using both the statistical results and the corresponding graphs. The interpretation and discussion of these results follow in chapter four.

3.1 Factors influencing AAD performance

A linear mixed-effects model (LME) was constructed to test the effect of spatial separation and background noise on AAD performance. The model predicted the decoding accuracy based on the separation angle between both speakers and the amount of babble noise, while incorporating the interaction between them. After presenting the LME model, the effect of each factor is further explored using post-hoc Wilcoxon signed-rank tests.

3.1.1 Linear mixed-effects model for decoding accuracy

In general, high decoding accuracies were achieved. The mean accuracy across all subjects was 81.9%, with a range from 67.7% to 92.0%. To investigate what factors influenced these decoding accuracies, a linear mixed-effects model fit by restricted maximum likelihood was constructed. In this model, separation angle and noise level were considered fixed factors, along with the interaction between them, whereas the subject was treated as a random factor. The accuracy of this model was tested using ANOVA, of which the results are given in Table 1. Both the separate factors and the interaction between them are significant.

Table 1. *ANOVA for linear mixed-effects model to predict decoding accuracy*.

variable	df	F	p
(intercept)	1, 454	2306.478	p < 0.001
separation angle	2, 454	11.130	p < 0.001
noise level	3, 454	17.236	p < 0.001
angle * SNR	6, 454	3.611	p = 0.002

3.1.2 The effect of spatial separation

After determining a significant effect of spatial separation on AAD performance, post-hoc tests were used to explore the difference between separation angles and sides.

3.1.2.1 The effect of the separation angle

Based on the LME model described above, the separation angle between both speakers has a significant effect on AAD accuracy ($F_{2,454} = 11.130$, $p < 0.001$). In other words, the mean decoding accuracy varied across separation angles. This was confirmed using Friedman's ANOVA ($\chi^2_2 = 10$, p = 0.007). Figure 1 compares the AAD performance across separation angles, with each box containing the mean accuracy of every subject for the corresponding angle.

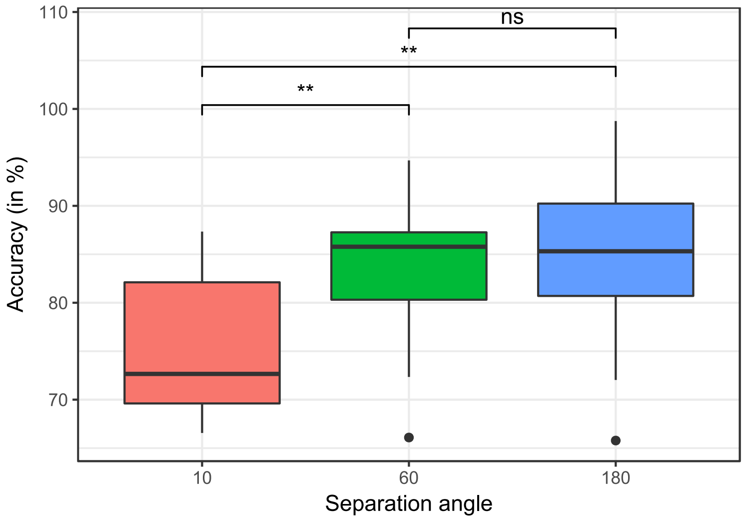

Figure 1. *Decoding accuracies for each separation angle.*

Multiple Wilcoxon signed-rank tests were performed post-hoc to compare the performance across different separation angles. The analysis results are listed in Table 2, along with the effect measures and the Benjamini-Hochberg adjusted p-values. Decoding accuracies were significantly worse for a 10° separation angle than for larger angles (60° or 180°). There is no evidence for different accuracies in 60° versus 180° separation angles.

Table 2. *Wilcoxon signed-rank tests for decoding accuracy across separation angles.*

pair of angles	W	p	r	corrected p
10°-60°	7	0.001	-0.593	0.003
10°-180°	13	0.008	-0.482	0.012
60°-180°	48	0.525	-0.116	0.525

3.1.2.2 The right ear advantage

Another Wilcoxon signed-rank test was performed to compare the decoding accuracy between trials with the attended speaker coming from either the right or the left side. As the trials with a 60° separation angle had both speakers on the same side, the data from these trials were excluded. Figure 2 shows the distribution of the mean accuracy for each side from every subject. Performance was worse for speech coming from the right side of the head (W = 113, p = 0.001, r = -0.593). Appendix E contains a further exploration of this difference.

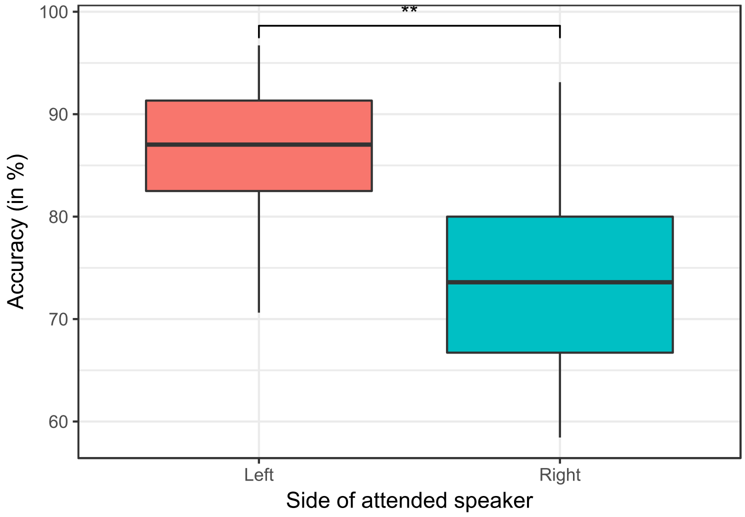

Figure 2. *Decoding accuracies when the attended speaker is on the left or on the right side.*

3.1.3 The effect of background noise

The second research question focussed on background noise as an influencing factor in AAD performance. The four conditions in the experiment are no noise (+∞ dB SNR), SNR1 (-1.1 dB SNR), SNR2 (-4.1 dB SNR), and SNR3 (-7.1 dB SNR). The difference in AAD accuracy between different noise levels is confirmed by the LME model ($F_{2,454} = 17.236$, $p < 0.001$), as well as by Friedman's ANOVA ($\chi^2_3 = 25.245$, $p < 0.001$). Figure 3 supports these analyses, by comparing the mean accuracy for every subject across the four noise levels.

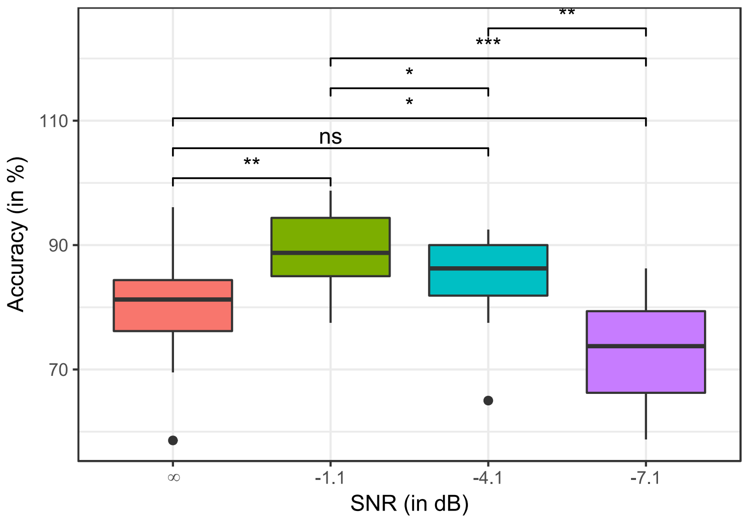

Figure 3. *Decoding accuracies for each noise level.*

Again, multiple Wilcoxon signed-rank tests were performed post-hoc to compare the performance across different noise levels. The results are listed in Table 3, along with the associated effect measures and the post-hoc corrected p-values. Decoding accuracies were highest in the condition with a low amount of noise (SNR1), even in comparison to the no noise condition. Performance in the SNR2-condition was comparable to that of the no noise condition and significantly better than that of the SNR3-condition. Finally, accuracies were significantly higher in the condition without noise than in the SNR3-condition.

Table 3. *Wilcoxon signed-rank tests for decoding accuracy across noise levels.*

pair of noise levels	W	p	r	corrected p
SNR0 – SNR1	3	0.001	-0.586	0.002
SNR0 – SNR2	32	0.121	-0.283	0.121
SNR0 – SNR3	100	0.022	-0.420	0.026
SNR1 – SNR2	82.5	0.011	-0.466	0.017
SNR2 – SNR3	120	0.001	-0.617	0.002
SNR2 – SNR3	105	0.001	-0.596	0.002

3.1.4 The interaction between separation angle and noise

The third research question studied the interaction between separation angles and noise levels. The LME model indicated that the interaction between both factors had a significant effect on decoding accuracies ($F_{6,454}$ = 3.611, p = 0.002). Figure 4 demonstrates this effect by comparing the accuracy distribution between the different noise levels, for each separation angle. Every box contains the mean accuracy for a specific combination of factors for each subject. The asterisks mark significant differences. Given that the pattern of significant comparisons is different for each separation angle, there must be an interaction effect between both factors.

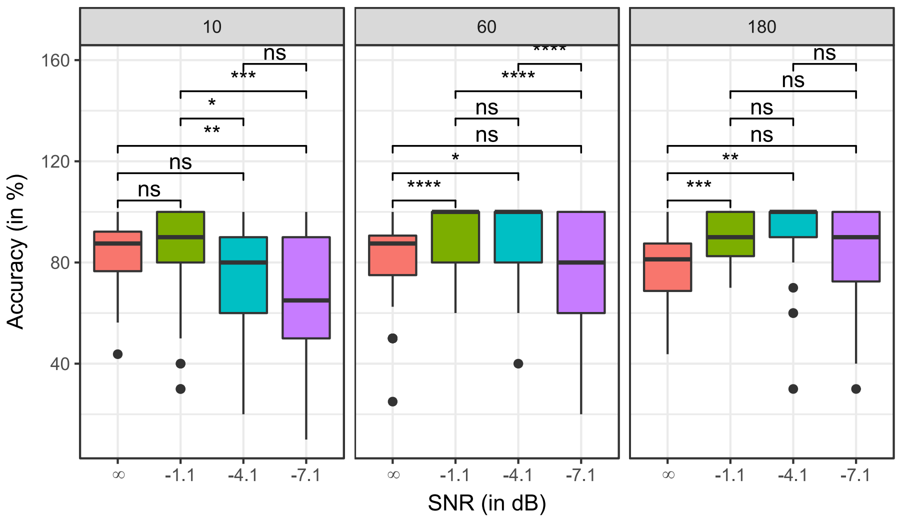

Figure 4. *Decoding accuracies for each noise level, compared across separation angles.*

Figure 5 shows the predicted decoding accuracies on a population level, based on the LME model. There is a clear effect of both separation angle (evidenced by different mean scores across the facets) and noise level (based on the differences in predicted accuracy across noise levels). Furthermore, it supports the existence of an interaction effect, as the shape of the predicted line is different for each facet. For example, whereas the accuracy is highest in the noise-free condition for the 10° angle, the accuracy is highest in the SNR1-condition (-1.1 dB SNR) for the larger angles.

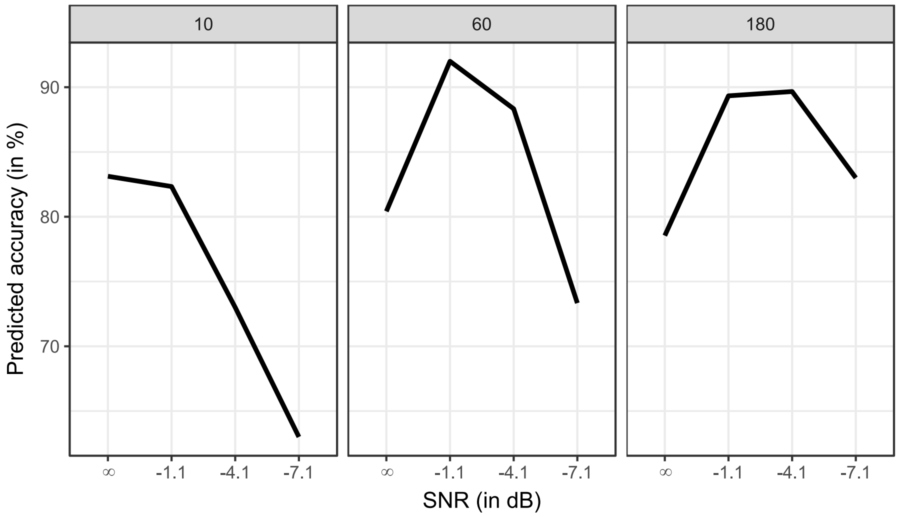

Figure 5. *Predicted decoding accuracies for each noise level, compared across separation angles.*

The relevance of the interaction between separation angles and noise levels can be seen in Figure 6. It shows the difference in the predicted decoding accuracy, when either accounting for the interaction effect or not. As a result, a positive value on the graph indicates that the decoding accuracy increased due to the interaction between both factors. Conversely, a negative value marks a lower predicted accuracy. The difference is compared across noise levels and separation angles. The graph shows that the decoding accuracy could be up to 10% lower or higher due to the interaction effect.

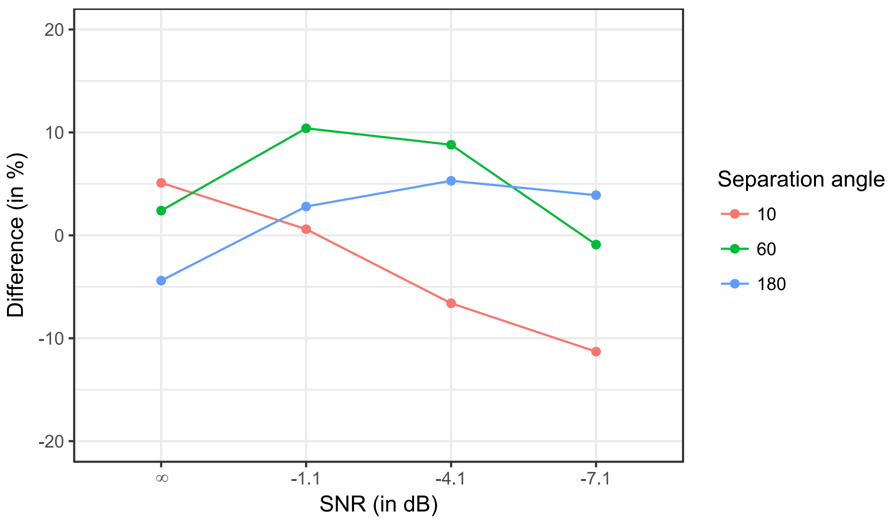

Figure 6. *Difference in predicted decoding accuracies with and without the interaction effect.*

3.2 Individual differences in AAD performance

Next, the individual performance of the subjects was explored. Both the variability in overall decoding accuracy and the potential link with the individual's performance on a speech-in-noise task were studied.

3.2.1 Variability among subjects

A Kruskal-Wallis test measured the difference between participants. The test indicated that the AAD performance varied significantly across subjects ($\chi^2_{14} = 60.732$, $p < 0.001$). Figure 7 shows this as well. It contains the decoding accuracy for each combination of separation angles and noise levels, compared across subjects.

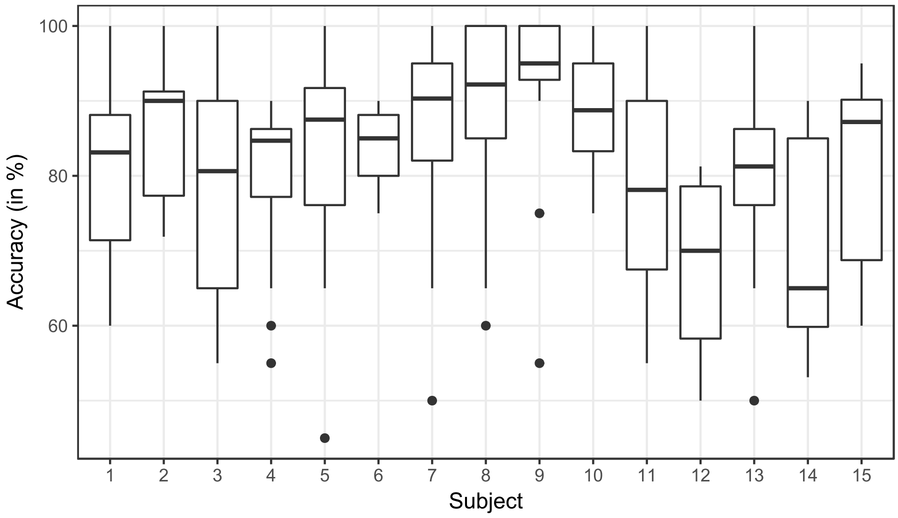

Figure 7. *Decoding accuracies for each subject*.

After applying the Kruskal-Wallis test, a multiple comparisons test was used to check for individual differences in performance. This test revealed that the accuracies of two subjects (number 12 and 14) were significantly worse than average. Subject 12 had lower accuracies than subjects 2, 7, 8, 9 and 10. The AAD performance for subject 14 was worse than for subjects 8 and 9. In other words: the highest average decoding accuracies were achieved for subjects 9 and 8, while the lowest average accuracies were found with subjects 12 and 14.

3.2.2 The connection between AAD performance and speech-in-noise thresholds

Kendall's tau was applied to tie the mean decoding accuracy for each subject to their performance on the speech-in-noise task. A small negative correlation was found, but it was not significant ($\tau = -0.105$, $p = 0.313$). The scatterplot in Figure 8 shows the relationship between the subjects' mean decoding accuracies across all conditions and their SRTs on the speech-in-noise task. The blue line indicates that in general, subjects with lower speech-in-noise thresholds had slightly higher decoding accuracies.

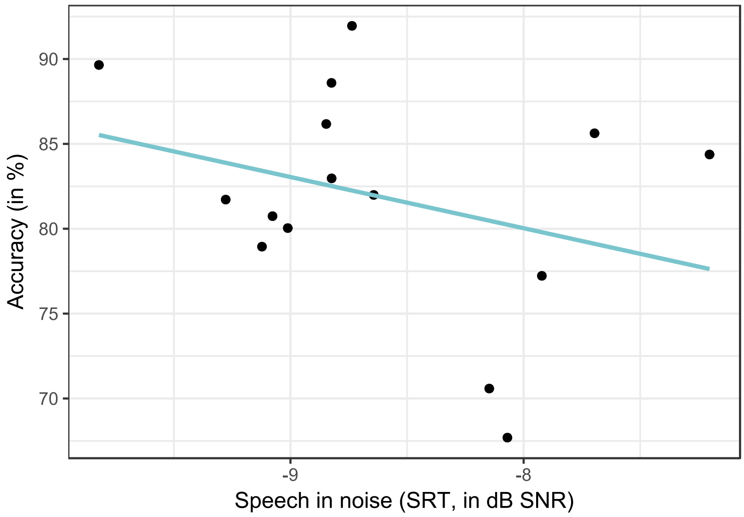

Figure 8. *Relationship between mean decoding accuracy and speech-in-noise threshold of individual subjects*.

To further explore this connection, the correlation was calculated for each separation angle. However, none of these correlations was significant either.

3.3 Factors influencing the subjective intelligibility

The final research question looked at the difference between AAD performance and the subjective speech intelligibility, as estimated by the subjects. To this end, the previous analyses were repeated, but this time with intelligibility as dependent variable. Furthermore, the relationship between accuracy and intelligibility was studied.

3.3.1 Linear mixed-effects model for subjective intelligibility

A new LME model was constructed to predict the subjective intelligibility, using separation angle, noise level and their interaction as fixed factors, and adding subject as a random factor. Again, the model was evaluated using ANOVA. Based on the results in Table 4, both the separate factors and the interaction between them are significant.

Table 4. *ANOVA for linear mixed-effects model to predict intelligibility.*

variable	df	F	p
(intercept)	1, 454	2249.130	$p < 0.001$
separation angle	2, 454	30.901	$p < 0.001$
noise level	3, 454	404.357	$p < 0.001$
angle * SNR	6, 454	3.892	$p < 0.001$

3.3.2 The effect of spatial separation and background noise

After determining a significant effect of both spatial separation and background noise on intelligibility, post-hoc tests were used to explore these effects.

3.2.2.1 The effect of the separation angle

The separation angle between speakers had a significant effect on the subjective intelligibility ($F_{2,454}$ = 30.901, p < 0.001). This was confirmed using Friedman's ANOVA (χ^2_2 = 7.600, p = 0.022) and is visualised in Figure 9, in which every box contains the mean accuracy for each subject for the corresponding separation angle.

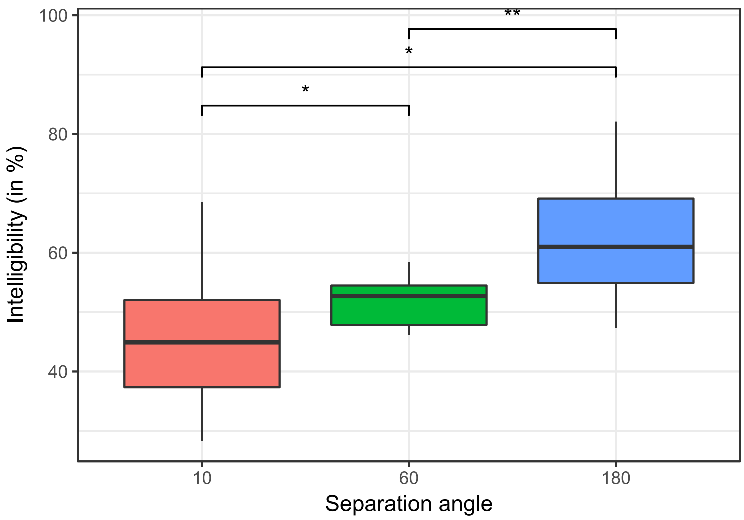

Figure 9. *Subjective intelligibility for each separation angle.*

Using multiple Wilcoxon signed-rank tests, the difference between separation angles was tested. Table 5 contains the results, indicating that the intelligibility rates significantly increased with larger separation angles.

Table 5. *Wilcoxon signed-rank tests for subjective intelligibility across separation angles.*

pair of angles	W	p	r	corrected p
10°-60°	24	0.041	-0.373	0.041
10°-180°	19	0.018	-0.432	0.027
60°-180°	13	0.005	-0.508	0.015

3.1.2.2 The right ear advantage

Another Wilcoxon signed-rank test was performed to compare the intelligibility between trials with the attended speaker coming from either the right or the left side. Figure 10 compares the mean intelligibility per subject for attended speech coming from the left or right side, indicating that there was no significant difference between both sides (W = 63, p = 0.890, r = -0.025).

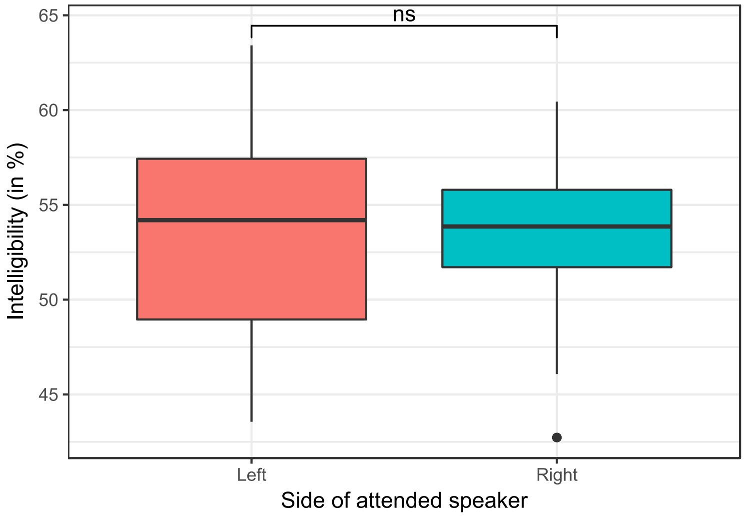

Figure 10. *Subjective intelligibility when the attended speaker is on the left or on the right side.*

3.2.2.3 The effect of the noise level

The amount of background noise had a significant effect on the subjective intelligibility ($F_{3,454}$ = 404.357, p < 0.001). This was confirmed using Friedman's ANOVA (χ^2_3 = 45, p < 0.001) and is demonstrated in Figure 11. Each box contains the mean accuracy per subject for the corresponding noise level.

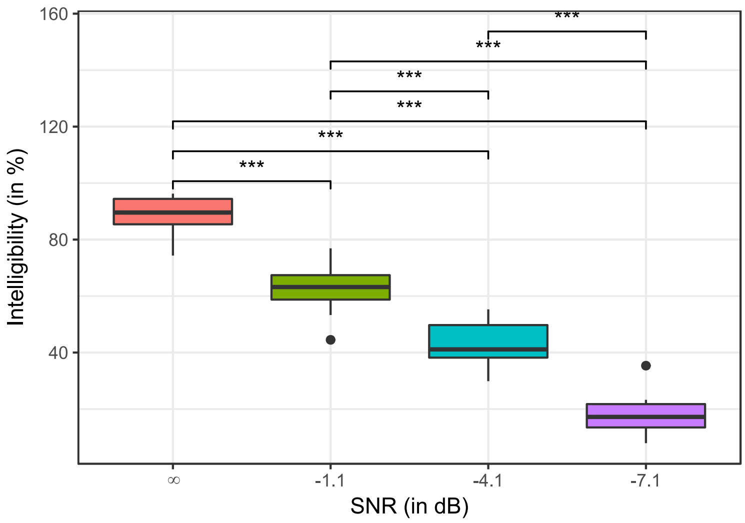

Figure 11. *Subjective intelligibility for each noise level.*

Multiple Wilcoxon signed-rank tests showed a significant decrease in intelligibility with every increase in noise level (W = 120, p < 0.001, r = -0.732, corrected p < 0.001, for all pairwise comparisons). Intelligibility ratings were highest in the condition without noise (mean = 89.4%) and lowest in the condition with the highest amount of noise (mean = 17.9%).

3.2.2.4 The interaction between separation angle and noise

The LME model indicated that the interaction between separation angles and noise levels significantly affects the subjective intelligibility ($F_{6,454} = 3.892$, $p < 0.001$). Figure 12 compares the intelligibility between noise levels for each separation angle. The boxes contain the mean decoding accuracy for each subject for the specific combination of factors.

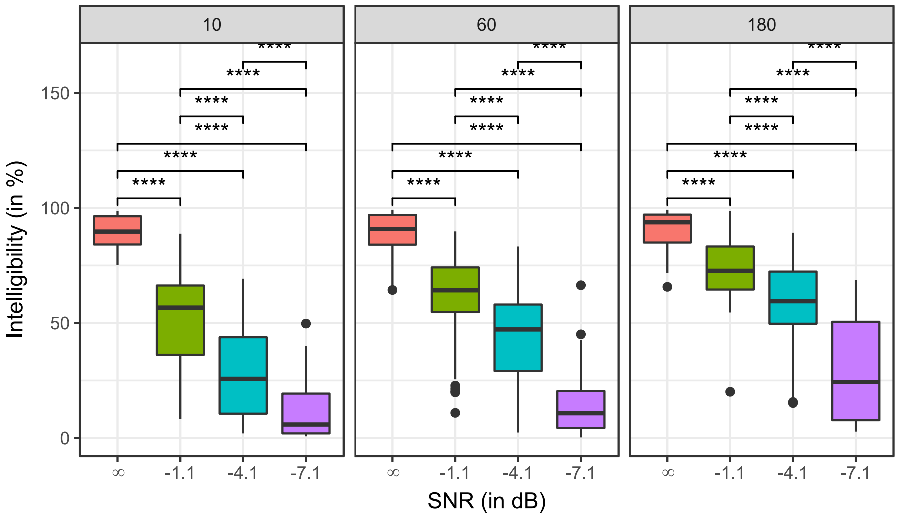

Figure 12. *Subjective intelligibility for each noise level, compared across separation angles.*

Based on the LME model, the intelligibility can be predicted for each combination of separation angles and noise levels. Figure 13 shows these predictions for each noise level, within each separation angle. The general shape of this graph is similar within each separation angle, with increasing noise levels resulting in decreasing intelligibility. However, small differences do exist between the graphs, consistent with an interaction effect.

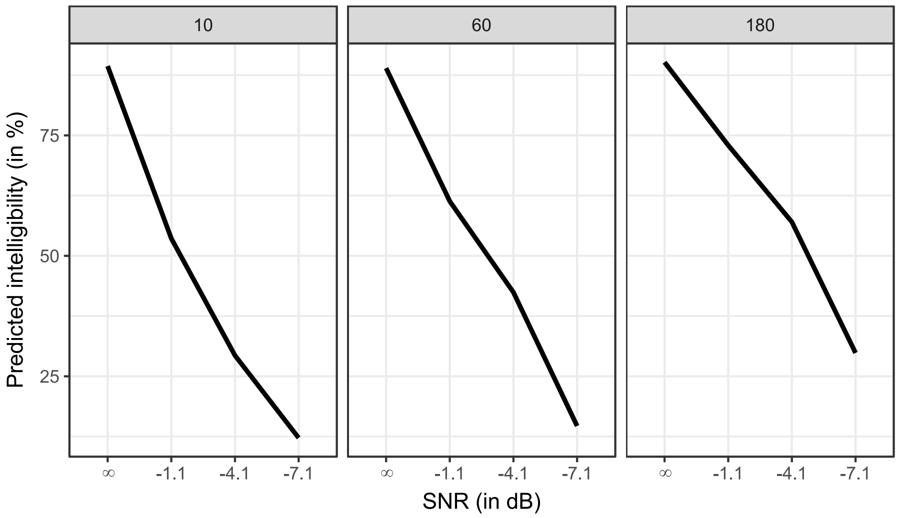

Figure 13. *Predicted intelligibility for each noise level, compared across separation angles.*

Again, the relevance of the interaction effect is explored using a line plot. Figure 14 marks the difference in the predicted intelligibility, when either considering the interaction effect or not. Positive differences represent an increase in the intelligibility ratings due to the interaction effect. Again, the differences are compared across noise levels and separation angles. Based on the graph, the interaction effect may result in up to 20% higher intelligibility ratings for trials without noise. For trials with the highest noise level, it shows a 10 to 20% decrease in intelligibility due to the interaction between both factors. This effect is stronger for trials with larger separation angles.

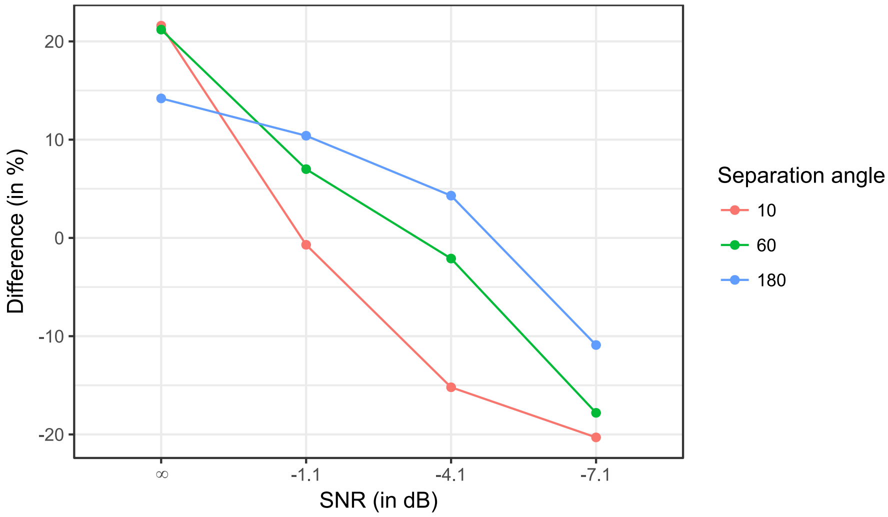

Figure 14. *Difference in predicted intelligibility with and without the interaction effect.*

3.3.3 Individual differences in intelligibility ratings

Given the relevance of individual differences for the decoding accuracy, these differences may also affect the subjective intelligibility. This was studied together with the potential link between the intelligibility ratings and the individual's performance on a speech-in-noise task.

3.3.3.1 Variability among subjects

To measure the difference between participants, a Kruskal-Wallis test was performed. There was no significant variation in intelligibility ratings between the subjects ($\chi^2_{14} = 7.958$, $p = 0.892$). Figure 15 shows the subjective intelligibility for each combination of separation angles and noise levels, compared across subjects. There are no relevant differences between subjects, which could be confirmed using a multiple comparisons test.

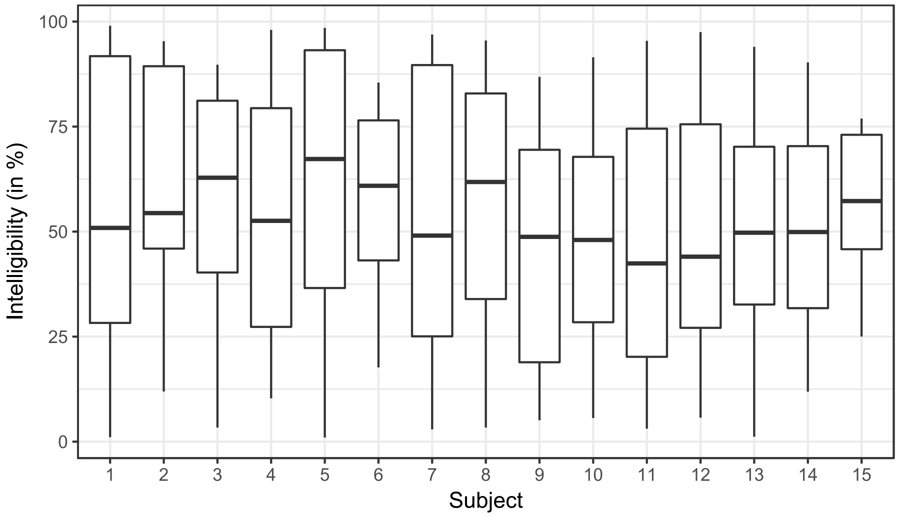

Figure 15. *Subjective intelligibility for each subject*.

3.3.3.2 The connection between intelligibility ratings and speech-in-noise thresholds

To find out whether the mean intelligibility rating for each subject could be tied to their performance on the speech-in-noise task, Kendall's tau was applied. This resulted in a small but insignificant negative correlation (τ = -0.295, p = 0.070). Figure 16 shows the relationship between the subjects' mean intelligibility ratings across all conditions and their SRT's on the speech-in-noise task. The blue line shows that on average, subjects with lower speech-in-noise thresholds rated the intelligibility slightly higher than subjects with higher thresholds. However, this trend was not significant.

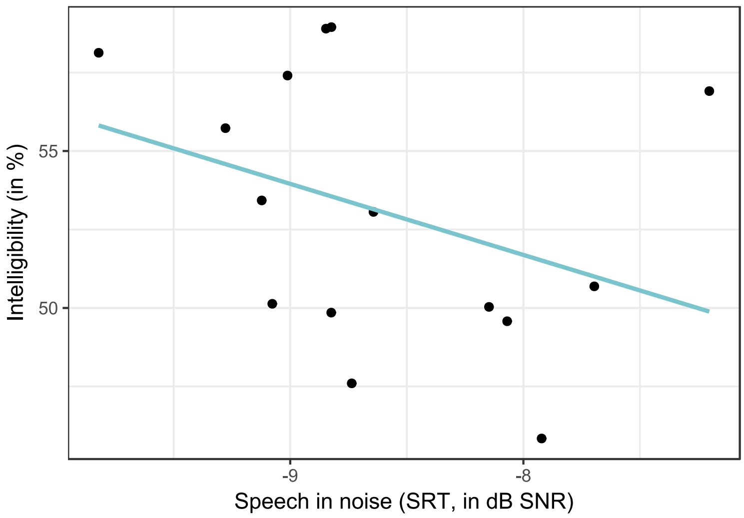

Figure 16. *Relationship between mean intelligibility rating and speech-in-noise threshold of individual subjects.*

The connection was further explored by calculating a correlation for each separation angle. Only for 60° angles, this produced a significant result (τ = -0.467, p = 0.008, corrected p = 0.023).

3.3.4 The connection between AAD performance and subjective intelligibility

Based on the analyses above, there are some clear differences between the decoding accuracies on one hand and the subjective intelligibility of speech on the other hand. The most important difference can be found in the effect of background noise, which affects accuracy and intelligibility differently. However, there might still be a connection between both measures. To test this hypothesis, Kendall's tau was calculated on a data set containing a mean decoding accuracy and an averaged intelligibility rating for each unique combination of a separation angle, a noise level, and a subject. These values are represented as 180 dots ('data points') in the scatterplot in Figure 17.

Although a positive correlation between both factors was expected, this correlation was not significant ($\tau = 0.209$, $p = 0.99$). Figure 17 indicates this weak connection with a blue line, showing that slightly higher intelligibility ratings were found in trials that were more successfully decoded. The figure also demonstrates that even with low intelligibility rates (below 50%), relatively high decoding accuracies could be achieved (mean accuracy 75%). For the lowest intelligibility rates (below 10%), decoding accuracies varied between 45% and 85% (mean accuracy 63%). Ultimately, about 3% of all data points (n = 5) represented decoding accuracies below chance level (55%). In general, decoding results are higher than could be expected based on the subjective intelligibility, given the many trials (n = 36) with intelligibility rates below 20%.

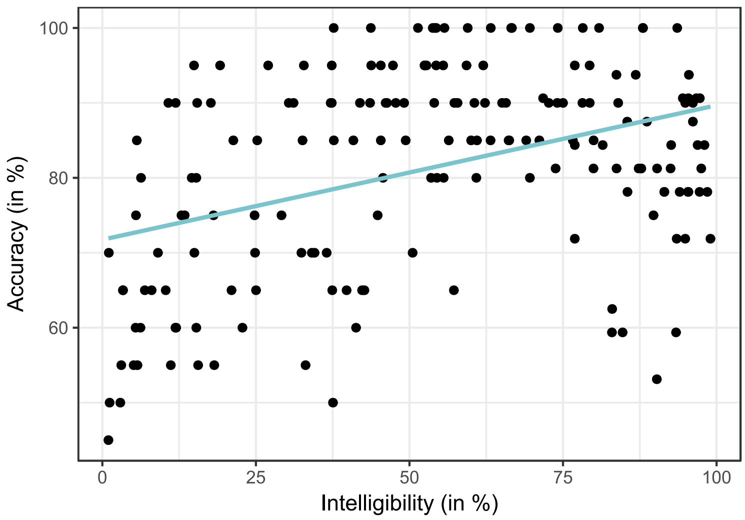

Figure 17. *Relationship between decoding accuracy and subjective intelligibility.*

4 Discussion

In this part of the thesis, the results from the different analyses are interpreted. After discussing the research questions separately, a general conclusion is drawn.

4.1 The effect of spatial separation

Both the decoding accuracy and the subjective speech intelligibility were influenced by spatial separation of the sound sources. Subjects found it easier to understand a speaker if the separation angle with the competing talker was large. Thus, intelligibility was higher for a 180° separation angle than for a 60° angle, and intelligibility rates were lowest for a 10° angle. This was reflected in the decoding accuracy as well: performance was worse when speakers were separated by a 10° angle, compared to a 60° or a 180° angle.

When listening to two or more concurrent speech streams, the brain segregates and integrates the auditory information from each stream using acoustic features. These features, like fundamental frequency or binaural cues, are essential to successfully encode each speech stream (Darwin, 2005). To prevent the use of fundamental frequency as a cue, the experiment included only stories read by female voices, making sure that the pitch was similar for each stream. Binaural cues, like ITDs and ILDs, were manipulated by varying the separation angle between both speakers. When the separation angle was small, spatial unmasking was limited due to small interaural differences, thus making it difficult to distinguish the sound sources. This was indicated by decreasing intelligibility ratings as the separation angle between the speakers diminished, in keeping with previous findings on spatial separation (Ericson et al., 2004).

As expected, the decoding accuracy decreased too with smaller separation angles, consistent with previous findings (Zhang, Lu, Wu, & Li, 2014). With the two sound sources closer together, the acoustic features of each stream are more similar, making it harder for the brain to separately encode each stream. Thus, as the encoded patterns were harder to distinguish, AAD became less successful. The difference between 60° and 180° separation angles was not significant, which might be related to the effort made by subjects. As the 180° condition was easiest in terms of speech intelligibility, maybe subjects did not have to pay close attention to understand the story. The 60° angle, however, had both speakers on the same side of the head, thereby reducing binaural cues and requiring more attentional compensation mechanisms from the subjects. Alternatively, the non-significance of this result could also be attributed to the limited number of participants (n = 15).

In addition to the separation angle, the effect of spatial separation was further explored in search of a potential REA. As the REA is associated with faster processing of speech coming from the right side, this might affect subjective intelligibility and AAD performance. Yet, previous research found evidence for a changing ear advantage depending on the focus of attention, which may influence the results (D'Anselmo et al., 2016). Given the diotic instead of dichotic presentation of the stories, however, the effect may not be measurable at all (Hiscock & Kinsbourne, 2011; Kimura, 1961). Indeed, no evidence for such an effect could be found when comparing the intelligibility ratings for each side of the attended speaker. With respect to the decoding accuracy, results were higher when the attended speech stream came from the left side, at least for trials with a 10° or 180° separation angle, indicating a possible left ear advantage (LEA). However, individual exploration of the decoding accuracies showed an LEA for twelve subjects and an REA for the three others, indicating that the difference could probably be attributed to individual variability and measurement artefacts. From that point of view, the 'evidence' for an overall left ear advantage is probably just a coincidental finding.

4.2 The effect of background noise

The amount of background noise had a significant effect on both decoding accuracy and subjective speech intelligibility. As the background noise increased, subjects found it harder to understand speech, resulting in lower intelligibility ratings. At the same time, decoding accuracies decreased with additional noise, although performance was better with a little background noise (-1.1 dB SNR) than with no noise at all.

The impact of noise on speech perception is obvious: it adds acoustical energy to the auditory scene, resulting in energetic masking. Increasing the relative intensity of the noise, compared to the target speech, reduces the intelligibility (Ericson et al., 2004). Some types of noise may also cause informational masking, meaning the noise itself distracts the listener. For example, whereas stationary noise only provides energetic masking, babble noise may also capture the listener's attention because of its meaningful content. Additionally, acoustic similarities between target and masking sounds may heighten the cognitive load with respect to streaming. Although all types of background noise interfere with speech processing, some types therefore divert a listener's attention more than others (Larsby, Hällgren, Lyxell, & Arlinger, 2005). A study by Zhang et al. (2016) showed an increased need for cortical processing in informationally complex backgrounds. Thus, attentional masking itself may influence brain activity (Manan, Yusoff, Franz, & Mukari, 2017; Wiegand, Heiland, Uhlig, Dykstra, & Gutschalk, 2018; Zhang et al., 2014). Finally, distracting background noise, like the babble noise used in the experiment, may disrupt the normal streaming processes. Whereas attention-based gain improves the representation of the attended stream (Ding & Simon, 2012a; Kerlin et al., 2010; Kong et al., 2014), salient features of the competing stream may lead to involuntary bottom-up attention, temporarily distracting the listener (Kaya & Elhilali, 2014).

Keeping this in mind, the effect of background noise on speech intelligibility is easy to comprehend. Additional noise resulted in more energetic and informational masking, making it harder to understand the attended speech stream. Furthermore, if the noise distracted the listener, the attended stream may have been lost for a short period of time, making it even more difficult afterwards to focus on the intended speaker. Thus, the intelligibility rates decreased with increasing noise levels, as expected.

To explain the effect of noise on the decoding accuracy, an analogous explanation can be given, at least when comparing the trials containing babble noise. With increasing energetic and informational masking, listeners struggled to focus on the attended speech stream. If they got distracted and lost their focus, this could have resulted in contamination of the speech representation in the brain. Additionally, the informational masking itself may have affected the cortical activity, making it harder for the decoder to accurately extract the attended speech envelope. However, this theory does not account for the slight decrease in accuracy for no-noise trials.

Gordon-Hickey and her colleagues found that the accepted noise level is higher in multi-talker backgrounds than with one competing talker. They attributed this finding to a relatively higher amount of informational masking in the single-talker background, compared to more energetic masking in a multi-talker background (Gordon-Hickey, Moore, & Estis, 2012). If their hypothesis is valid, the explanation above may also apply to the no-noise condition after all. Although trials in this condition contained no additional background noise, the presence of the competing talker resulted in informational masking that was perhaps more distracting than the babble noise in the other trials. This distraction may have interfered with the cortical representation of the attended stream. Apart from this masking effect, some alternative explanations may account for the lower decoding accuracy in the trials without background noise.

First, because subjects found no-noise trials easier in terms of speech intelligibility, they may have adapted different listening strategies that required less cognitive resources. This may have resulted in less attentional gain control to track the attended stream, making the cortical representations harder to decode. This theory is partially supported by a study from Papesh and her colleagues. They found that in some cases, a low amount of background noise may result in enhanced AEPs compared to a quiet condition (Papesh, Billings, & Baltzell, 2015). This finding may explain why the decoding accuracy is higher in trials with babble noise, compared to the no-noise trials.

Secondly, a different explanation may be found in the experiment design. Since the trials without noise always contained the first part of a new story, subjects may simply have found these parts less captivating. Additionally, as each story was narrated by a different speaker, subjects may have needed some time to adapt to this new voice, as voice familiarity can benefit speech-in-noise performance (Johnsrude et al., 2013). Furthermore, the analysis treated all trials without noise equally, although some trials were presented after each test block and contained a repetition of the first part of the story. These trials might have been less interesting to the listeners, receiving less attention than the other trials. In that case, their more passive listening attitude would have resulted in smaller AEPs, thus affecting the AAD performance (Zhang et al., 2016).

4.3 The interaction effect between separation angle and noise

As both the separation angle and the noise level influenced the subjective intelligibility and the decoding accuracy, it made sense to look for an interaction effect between these parameters. Indeed, the LME models indicated that the factors interacted for both intelligibility and accuracy.

The intelligibility ratings showed a steady decline with increasing noise levels or decreasing separation angles. The effect of the interaction between both factors was explored by comparing the predicted intelligibility rates from the LME, with or without the interaction effect. Figure 14 showed that for trials with low noise levels, the predicted intelligibility was higher when accounting for the interaction effect. Especially for large separation angles (60° and 180°), the intelligibility could be up to 20% higher for low noise levels, pointing to a combined positive effect of both factors. For the 10° separation angle, a similar increase in intelligibility could be found, possibly indicating that listeners benefit from the absence of noise in difficult listening situations. Conversely, the predicted scores were lower than expected for trials with a high noise level across all separation angles, showing the dominant impact of high-level background noise. Again, this effect was larger for large separation angels. A possible explanation is that compensational mechanisms (based on the spatial separation between speakers) were not sufficient in high noise levels, causing the intelligibility results to drop.

As discussed before, AAD accuracies were generally highest for trials with a small amount of babble noise (-1.1 dB SNR). Performance dropped for trials with a medium amount of babble noise (-4.1 dB SNR), resulting in comparable accuracies to the trials without noise. Performance was lowest for trials with a high amount of noise (-7.1 dB SNR). However, this pattern varied across separation angles, as shown in Figure 4. To explore this variation, Figure 6 showed the estimates from the LME are calculated with and without factoring in the interaction effect. In general, the interaction effect had less influence on the decoding accuracy than on the subjective intelligibility. For 10° trials, it resulted in an additional decrease in accuracy for increasing noise levels. This may indicate the existence of a cumulative negative effect of both a small separation angle and high noise levels on the cortical tracking of speech. For the 180° trials, the opposite trend could be found, although the difference equalled 5% at most. Trials with background noise had slightly higher decoding accuracies due to the interaction effect. Perhaps listeners could exploit the spatial separation of both speech streams to partially compensate for the effect of background noise. Finally, for 60° trials in low levels of background noise, the accuracy was up to 10% higher when accounting for the interaction effect. Again, the spatial separation between both speakers may have aided the speech-in-noise perception of the subjects.

In conclusion, spatial separation and background noise interacted when influencing both intelligibility and decoding accuracy. This is consistent with the previous finding that the interaction between babble noise and perceptual separation affected the cortical representation of speech, by promoting selective auditory attention (Zhang et al., 2014). Furthermore, changes in both the characteristics of the competing speech and the spatial separation between target and noise sources are known to interact when influencing speech perception and cognitive load (Zekveld, Rudner, Kramer, Lyzenga, & Rönnberg, 2014). It is also known that binaural cues can aid speech perception in noise (Shabtai, Nehoran, Ben-Asher, & Rafaely, 2017). Although these research findings support the results discussed above, caution is recommended when interpreting the results. As the number of subjects was limited (n = 15) and between-subjects variability was large, the measured values may not be reliable. Furthermore, many processing mechanisms in the auditory system are not yet fully understood, so that certain undetected effects may have been overlooked when explaining these results.

4.4 The effect of individual differences

The fourth research question focussed on individual differences in AAD performance. The use of subject-specific decoders was appropriate due to the large amount of available data for each subject. Even so, substantial accuracy differences could be found between subjects. In terms of subjective intelligibility, no individual differences were found. The individual differences in both accuracy and intelligibility could not be tied to the subjects' performance on the speech-in-noise task.

There were significant individual differences in AAD performance, with mean accuracies ranging from 68% to 92%, resulting in an overall mean accuracy of 82%. These results are similar to previous research findings from comparable experiments. For example, O'Sullivan reported a mean decoding accuracy of 89%, with individual means ranging from 57% to 100% (O'Sullivan et al., 2015). Horton mentioned a mean accuracy of 82%, with subject-specific results ranging from 73% to 93% (Horton et al., 2014). Another study found subject-specific accuracies between 65% and 95% (Van Eyndhoven et al., 2017). Mirkovic reported a mean decoding accuracy of 85% (Mirkovic et al., 2015). Other researchers found average accuracies of 82% (Biesmans et al., 2017), of 87% (Das et al., 2016), and of 92% (Zink et al., 2016), depending on the research design.

For two subjects, the decoding accuracies were significantly higher than average, and for two others they were lower. The source of this variability is unclear. The inter-subject differences may have been caused by varying levels of interest in the stories, variable impedances of the EEG electrodes, individual degrees of attention and fatigue, limb movement disturbing the EEG recordings and so on. Furthermore, each subject possibly used slightly different listening strategies to focus on the indicated speaker, which may have been compatible with AAD in varying degrees. Additionally, as the randomisation process ensured that the order of the conditions varied across subjects, this may also have influenced their learning curves, affecting their overall AAD performance. In general, research shows that BCI effectiveness differs across subjects, due to the adapted strategies, the task-related experience, and unexplained individual differences (Kober, Witte, Ninaus, Neuper, & Wood, 2013; Neumann & Birbaumer, 2003).

With regard to the intelligibility, no differences could be found between the subjects. This finding is unsurprising, as each subject would have interpreted the rating scale according to his or her own performance, using a broad range of ratings. Furthermore, the experiment design ensured that the listening task would be difficult for every subject, regardless of their speech-in-noise reception thresholds, by including high levels of background noise. As a result, all subjects would have understood almost nothing during the most difficult trials.

There was no evidence for a connection between the subjects' SRT and decoding accuracy. However, AAD appeared to be slightly more successful for subjects with lower SRTs. To check whether the separation angle influenced the significance of the result, the correlation was calculated separately for each separation angle. These analyses resulted in non-significant correlations as well. Given the limited variation in SRTs, this was only to be expected. The lack of evidence for a connection could be due to the small number of participants (n = 15), the fact that they were all young adults with normal hearing (i.e. thresholds below 20 dB HL), or the strict speech-in-noise selection criterion (requiring an SRT of -7 or lower to participate). It is also possible that the results are not connected at all, because the underlying tasks are quite different. During the speech-in-noise test, subjects had to listen to Matrix-sentences before selecting the appropriate words on a screen. The performance of each subject was tested using a behavioral measurement, and the results could be influenced by guessing or by accidently clicking on the wrong box on the screen. To obtain the decoding accuracy, the subject's EEG was analysed to reconstruct the attended speech envelope. Next, the speech envelopes of both speakers were compared to the reconstructed envelope, in a mathematical comparison following an objective measure. The results could unintentionally be influenced by electrode impedances, irrelevant brain activity due to limb movement or other thought processes, or distortion of the EEG signal in the transfer to the computer. In other words, the sources of 'unrelated' variability were different for each test.

Furthermore, a study by Houtgast & Festen (2008) showed that overall, a combination of PTA, age, and temporal processing skills can account for about 70% of variability in SRTs. A review paper from the same year, based on 20 experimental studies, concluded that the SRT was mainly influenced by hearing loss, with an additional effect of some cognitive abilities; probably working memory, although the results differed across the studies (Akeroyd, 2008). Thus, in young adults with normal hearing, like the subjects tested in the experiment, it is difficult to explain the between-subjects variance, since no cognitive skills were explicitly tested. As a result, connecting their SRT to their mean decoding accuracy is, at least for the time being, not plausible.

Finally, the individuals' SRTs could not be linked to their subjective intelligibility ratings. Only for 60° separation angles, a significant negative correlation could be found. It is unclear why this is the case. As described above, an SRT below -7 dB SNR was required to participate, making it harder to study individual differences. The significant result for the 60° separation angle may therefore reflect an accidental finding. Conversely, the lack of evidence for a general connection between speech-in-noise performance and speech intelligibility may be due to the limited number of subjects.

4.5 The connection between subjective intelligibility and decoding accuracy

The final research question compared the subjective intelligibility ratings to the AAD performance. Most intelligibility results have already been discussed above, indicating that spatial separation and background noise had a similar effect on the intelligibility as they had on the decoding accuracy. Individual differences were less relevant for intelligibility than for decoding accuracies, and the mean intelligibility ratings could not be linked to the subjects' performance on a speech-in-noise task. Thus, only the last part of this research question remains: is there a connection between the subjective intelligibility on one hand and the decoding accuracy on the other?

A small but insignificant positive correlation could be found, indicating that AAD performance tended to be higher for trials with high intelligibility ratings. Again, this could be explained by the small number of subjects, resulting in a dataset that may not be large enough to identify such a connection. Furthermore, the intelligibility ratings were very subjective, as each subject was free to interpret the scale as he or she wished. Some may have considered '0' to equal not understanding most of the story, why others regarded '0' as the absolute zero – the point at which they couldn't even properly track the attended speech stream anymore. Additionally, given the considerable amount of between-subjects variability in AAD performance, this may have concealed any connection between intelligibility and accuracy.

Alternatively, the non-significance may be explained by the inclusion of trials without background noise. These trials had lower decoding accuracies than the trials with a low amount of babble noise, although the intelligibility ratings were higher. To check whether this discrepancy could account for the lack of correlation between both values, the correlation was calculated again based on trials with background noise only. However, even though the magnitude of the correlation increased, it remained insignificant ($\tau = 0.441$, $p = 0.99$).

4.6 Methodological shortcomings of the experiment

Although several hypotheses could be supported with statistically significant test results, the current experiment has a few important shortcomings. They can be split up into two parts: the participant group and the experiment design.

The most important issue is the limited number of subjects that were tested. This limitation was due to the time schedule of the researchers involved, the availability of the sound proof cabins, and the scope of this thesis. However, it may have resulted in less accurate conclusions on the presence and relevance of certain effects. Furthermore, the subjects included only young adults with normal hearing. This strict selection of participants resulted in a very homogeneous group, making the interpretation of the results more straightforward, but leaving the generalisation of the results more problematic.

The second issue is related to the experiment design. The experiment allowed for a comparison of AAD performance for multiple subjects across different noise levels and separation angles. A randomized, balanced design made sure that each story could be used in two different angular conditions, depending on the subject number (see Appendix C). However, the speaker location of each story did not change within an angular condition. For a specific angular condition, a certain story would always come from the same angle. Furthermore, the short and long stories were not interchanged. Due to this incomplete randomisation, there might have been an undetected effect of the story that was used. In addition, post-experiment analyses revealed a difference in performance depending on the story that was used (see Appendix A). These design flaws may have influenced the results unnoticed, interfering with a successful and well-founded interpretation.

4.7 Suggestions for future research

To gain a better insight into the effectiveness of AAD in different circumstances, more research is needed. With an eye to the intended use of AAD in HA technology, some important advancements are necessary. Future research should focus on three important issues: technical constraints, realistic situations, and selection of the subjects.

The first issue is related to technical constraints associated with AAD. The experiment in this thesis used 64-channel EEG recordings, requiring careful preparation of the subjects and the use of gel to improve electrode contact. However, to make AAD available for everyday use, small and portable EEG alternatives should be developed. Furthermore, this study used offline decoders that compared two clean speech signals to the reconstructed envelope. Unfortunately, in realistic situations, clean speech signals will not be available. Decoders should instead be able to compare the reconstructed envelope to (a processed version of) the mixed signal. These technical issues are beyond an audiologist's perspective but should be dealt with nonetheless.

The second issue concerns the application of AAD in lifelike situations. So far, AAD research has mainly focussed on two-talker scenarios. By adding babble noise and varying the separation angle between both speakers, the experiment in this thesis attempted to create more realistic situations. Even so, additional variation is necessary to understand the effectiveness of AAD in everyday acoustic environments. For example, the three separation angles should be complemented with additional angles, preferably comparing absolute speaker positions within separation angles as well. Moreover, this thesis applied babble noise at four levels (no noise, -1.1, -4.1, -7.1 dB SNR), but this only serves as a first exploration of the effect of background noise. By using various noise levels, including positive SNRs, along with different types of noise, more comprehensive results could be found. Finally, this thesis ignored several other factors that could influence AAD performance, such as reverberation, additional concurrent speech streams, male speakers, foreign languages... These factors encompass many future research possibilities.

Finally, the third issue lies within the selection of the subjects. This experiment, as most of the other AAD research, exclusively employs young adults with normal hearing. However, as hearing difficulties and cognitive skills have a major influence on speech perception, they probably affect AAD as well. For example, studies show that AEPs are affected by hearing loss, but also by age, and that the effect of noise interacts with the individual's hearing status (Maamor & Billings, 2017; Oates, Kurtzberg, & Stapells, 2002). Based on these findings, it is very important to extend AAD research to other populations, like hearing impaired subjects and subjects with different ages. Including these populations may yield valuable insights into the effectiveness of AAD, which could be relevant to the population of young adults with normal hearing as well.

In conclusion, there is still room for a lot of research on the topic of AAD. The technology has great potential, but more information is needed before it can be used in everyday applications.

4.8 Summary

Spatial separation had an important effect on both speech intelligibility and decoding accuracy, with larger separation angles resulting in higher scores on both measures. The level of background noise had a significant effect as well: intelligibility ratings decreased as the noise level increased, and a similar influence could be found on the decoding accuracy. However, AAD performance was slightly lower for trials without babble noise. Spatial separation and background noise also interacted when affecting speech intelligibility and AAD performance, possibly reflecting additive positive or negative effects, or compensational mechanisms in difficult listening conditions. Individual differences between subjects accounted for a major amount of variance in AAD performance. Still, there was no evidence for a correlation between the SRT and accuracy from individual subjects. The intelligibility ratings were similar across subjects and could not be linked to their SRTs either. Subjective intelligibility and decoding accuracy were not correlated.

Conclusion

AAD is a relatively new topic in audiology research. Many studies only aimed to prove that detecting attention based on brain imaging is possible, and that decoders succeed at identifying the attended speaker in a two-talker scenario. Yet eventually, researchers want to apply AAD to HA systems to solve (or lessen) noise-related problems. With this in mind, studies using a two-talker scenario without background noise are not very realistic.

The goal of this thesis was to study the effectiveness of AAD in more lifelike listening conditions. To this end, AAD was performed in a two-talker scenario, but this time with varying separation angles between both speakers and using different levels of background noise. The objective was to answer multiple research questions: what is the effect on AAD performance of spatial separation, background noise, and the interaction between them? How important are individual differences when it comes to decoding accuracy? Can the decoding accuracy under different circumstances be linked to subjective intelligibility ratings?

During the experiment, brain activity from 15 subjects with normal hearing was recorded using 64-channel EEG. Subject had to listen to a story, narrated by a female speaker, while ignoring the story told by a competing talker. The speakers were spatially separated by a 10°, 60°, or 180° angle. In some trials, babble noise was added at varying levels (-1.1, -4.1, or -7.1 dB SNR). To complete every combination of separation angles and noise levels, all subjects listened to eight stories. After each story part, subjects had to estimate how much they had understood from the narration before the noise level changed. The EEG data were analysed separately for each subject, constructing a subject-specific decoder by the leave-one-out method. This was done for each 30-second interval, before summing the decoding accuracies within each story part. Afterwards, the decoding accuracies were combined with the intelligibility ratings for analysis purposes.

Using an LME model to predict the decoding accuracy, the effect of the separation angle, the level of background noise, and the interaction between them could be investigated. All three factors significantly influenced AAD performance. Larger separation angles resulted in higher decoding accuracies. Higher noise levels decreased the decoding accuracies; however, performance in the no-noise condition was worse than in trials with babble noise at -1.1 dB SNR. Decoding accuracies were slightly higher in trials with the attended speaker on the left side of the listener.

To compare these results to the intelligibility ratings, a new LME model was constructed to predict the subjective intelligibility. In this model, the same factors were significant, including the interaction between them. Larger separation angles had higher intelligibility ratings, and as noise levels increased, the reported intelligibility dropped accordingly. The side of the attended speaker did not influence the ratings. Although the correlation between the decoding accuracy and the subjective intelligibility was not significant, they are both influenced by the same characteristics of the acoustic environment.

Finally, individual differences in AAD performance across subjects were significant as well. Although decoding accuracies covered a large range for all subjects, some had markedly lower or higher performance than others. These differences could not be tied to their performance on a speech-in-noise task; however, given the small number of participants and the strict selection criteria including their speech-in-noise performance, this is not surprising. Intelligibility ratings were similar across subjects and could not be linked to their speech-in-noise performance either.

When reading about AAD and carrying out an experiment using this technology, it is impossible to not be amazed by this fascinating topic. It is almost incredible what the technology is already capable of right now: reading people's minds to know what or who they are listening to. For the time being, there certainly are some limitations, but researchers are working hard to overcome these issues. I believe it is only a matter of time until AAD can be done in real-time, with portable EEG set-ups, and with high success rates. At that point, the technology has the potential to overcome a large remaining problem in HA development: intelligent noise suppression.

Of course, there are yet some challenges that must be accepted. As mentioned above, current research mostly uses 64-channel EEG-caps, which is not suitable for everyday use. Furthermore, the high reported decoding accuracies are mainly based on two-talker scenarios without noise, in which decoders have the clean speech signals at their disposal to compare the reconstructed envelopes to. These results may paint a too positive picture. However, some promising results have been reported in more realistic conditions, preparing the way for more research in lifelike situations.

Bibliography

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(Suppl. 2), 53–71.

Alho, K., Teder, W., Lavikainen, J., & Näätänen, R. (1994). Strongly focused attention and auditory event-related potentials. *Biological Psychology*, *38*(1), 73–90.

Anderson, P. A., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*, *54*, 944–958.

Bau, J. (2016). *Real-time EEG based auditory attention detection with closed loop gain control*. KU Leuven.

Bennett, R. J., Laplante-Lévesque, A., Meyer, C. J., & Eikelboom, R. H. (2017). Exploring hearing aid problems: Perspectives of hearing aid owners and clinicians. *Ear and Hearing*, 1–16.

Bera, T. K. (2015). Noninvasive electromagnetic methods for brain monitoring: A technical review. In A. E. Hassanien & A. T. Azar (Eds.), *Brain-computer interfaces: Current trends and applications* (pp. 51–95). Cham, Switzerland: Springer.

Berti, S. (2012). Automatic processing of rare versus novel auditory stimuli reveal different mechanisms of auditory change detection. *NeuroReport*, *23*, 441–446.

Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2017). Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(5), 402–412.

Biesmans, W., Vanthornhout, J., Wouters, J., Moonen, M., Francart, T., & Bertrand, A. (2015). Comparison of speech envelope extraction methods for EEG-based auditory attention detection in a cocktail party scenario. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (pp. 5155–5158). Milan, Italy: IEEE.

Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, *254*(1–2), 15–24.

BioSemi. (2002). ActiveTwo. Amsterdam: BioSemi B.V.

BioSemi. (2016). ActiVIEW. Amsterdam: BioSemi B.V.

Bizley, J. K., & Walker, K. M. M. (2009). Distributed sensitivity to conspecific vocalizations and implications for the auditory dual stream hypothesis. *Journal of Neuroscience*, *29*(10), 3011–3013.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2013). Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia*, *51*(13), 2715–2728.

Bradley, J. S., Reich, R. D., & Norcross, S. G. (1999). On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(4), 1820–1828.

Breen, M., Dilley, L. C., Devin McAuley, J., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience*, *29*(9), 1132–1146.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT press.

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*(5), 975–979.

Christison-Lagay, K. L., Gifford, A. M., & Cohen, Y. E. (2015). Neural correlates of auditory scene analysis and perception. *International Journal of Psychophysiology*, *95*(2), 238–245.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49–69.

Commers, L. (2017). *De invloed van aandacht op het objectief meten van spraakverstaanbaarheid*. KU Leuven.

Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, *116*(2), 1057–1065.

D'Anselmo, A., Marzoli, D., & Brancucci, A. (2016). The influence of memory and attention on the ear advantage in dichotic listening. *Hearing Research*, *342*, 144–149.

Darwin, C. J. (2005). Pitch and auditory grouping. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (Vol. 24, pp. 278–305). New York: Springer.

Das, N., Biesmans, W., Bertrand, A., & Francart, T. (2016). The effect of head-related filtering and ear-specific decoding bias on auditory attention detection. *Journal of Neural Engineering*, *13*(5).

Dashti, R., Khiavi, F. F., Sameni, S. J., & Bayat, A. (2015). Elderly with different types of hearing loss and comorbidities: Satisfaction with hearing aids. *Jundishapur Journal of Health Sciences*, *7*(3), 5–8.

Dawes, P., & Munro, K. J. (2017). Auditory distraction and acclimatization to hearing aids. *Ear and Hearing*, *38*(2), 174–183.

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *The Journal of Neuroscience*, *37*(27), 6539–6557.

de Souza Ranaudo, F., de Sá, A. M. F. L. M., & Felix, L. B. (2012). Real time classification of selective attention on auditory steady-state evoked potentials. In *Biosignals and Robotics Conference (BRC), 2012 ISSNIP* (pp. 1–5). Manaus, Brazil: IEEE.

Deike, S., Denham, S. L., & Sussman, E. (2014). Probing auditory scene analysis. *Frontiers in Neuroscience*, *8*(293), 5–7.

Desjardins, J. L., & Doherty, K. A. (2012). Age-related changes in listening effort for various types of masker noises. *Ear and Hearing*, *34*(3), 261–272.

Dillon, H. (2012). *Hearing aids* (Second edi). New York: Thieme Medical Publishers Inc.

Ding, N., & Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 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*(1), 78–89.

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, *114*(51), 13579–13584.

Dubno, J. R., Ahistrom, J. B., & Horwitz, A. R. (2002). Spectral contributions to the benefit from spatial separation of speech and noise. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1297–1310.

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *The International Journal of Aviation Psychology*, *14*(3), 313–334.

Escera, C. (2011). The role of auditory brainstem in regularity encoding and deviance detection. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The frequency-following response: A window into human communication* (pp. 101–120). Cham, Switzerland: Springer.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, *5*(3–4), 151–166.

Escera, C., Leung, S., & Grimm, S. (2014). Deviance detection based on regularity encoding along the auditory hierarchy: Electrophysiological evidence in humans. *Brain Topography*, *27*(4), 527–538.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (First edi). London: SAGE Publications Ltd.

Folyi, T., Fehér, B., & Horváth, J. (2012). Stimulus-focused attention speeds up auditory processing. *International Journal of Psychophysiology*, *84*(2), 155–163.

Francart, T. (2008). APEX.

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, *156*, 435–444.

Fukushima, M., Saunders, R. C., Leopold, D. A., Mishkin, M., & Averbeck, B. B. (2014). Differential coding of conspecific vocalizations in the ventral auditory cortical stream. *Journal of Neuroscience*, *34*(13), 4665–4676.

Gandras, K., Grimm, S., & Bendixen, A. (2017). Electrophysiological correlates of speaker segregation and foreground-background selection in ambiguous listening situations. *Neuroscience*, 1–11.

Getzmann, S., & Näätänen, R. (2015). The mismatch negativity as a measure of auditory stream segregation in a simulated “cocktail-party” scenario: Effect of age. *Neurobiology of Aging*, *36*(11), 3029–3037.

Getzmann, S., Wascher, E., & Falkenstein, M. (2015). What does successful speech-in-noise perception in aging depend on? Electrophysiological correlates of high and low performance in older adults. *Neuropsychologia*, *70*, 43–57.

Ghazanfar, A. A., & Eliades, S. J. (2014). The neurobiology of primate vocal communication. *Current Opinion in Neurobiology*, *28*, 128–135.

Gomes, H., Duff, M., Barnhardt, J., Barrett, S., & Ritter, W. (2007). Development of auditory selective attention: Event-related potential measures of channel selection and target detection. *Psychophysiology*, *44*(5), 711–727.

Gordon-Hickey, S., Moore, R. E., & Estis, J. M. (2012). The impact of listening conditions on background noise acceptance for young adults with normal hearing. *Journal of Speech, Language, and Hearing Research*, *55*, 1356–1372.

Goris, L. (2016). *Neural responses to the speech envelope: Towards an objective measure for speech*. KU Leuven.

Grange, J. A., & Culling, J. F. (2016). The benefit of head orientation to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *139*(2), 703–712.

Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing Research*, *307*, 98–110.

Habibi, A., Cahn, B. R., Damasio, A., & Damasio, H. (2016). Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience*, *21*, 1–14.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*(3–4), 277–90.

Hearing Loss Association of America. (2017). Types, causes and treatment. Retrieved November 11, 2017, from http://www.hearingloss.org/content/types-causes-and-treatment

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews*, *8*, 393–403.

Hilkhuysen, G. L. M., Gaubitch, N., & Huckvale, M. (2013). Effects of noise suppression on intelligibility: Experts' opinions and naïve normal-hearing listeners' performance. *Journal of Speech, Language, and Hearing Research*, *56*(2), 404.

Hiscock, M., & Kinsbourne, M. (2011). Attention and the right-ear advantage: What is the connection? *Brain and Cognition*, *76*(2), 263–275.

Horton, C., Srinivasan, R., & D'Zmura, M. (2014). Envelope responses in single-trial EEG indicate attended speaker in a 'cocktail party.' *Journal of Neural Engineering*, *11*, 1–12.

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., Verschuure, H., & Dreschler, W. A. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. *International Journal of Audiology*, *53*(10), 760–763.

Hougaard, S., Jensen, O., Kristensen, M., Ludvigsen, C., Petersen, J., & Weis, P. (1995). *Sound & hearing* (Second edi). Widex ApS.

Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology*, *47*(6), 287–295.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America*, *122*(3), 1777–1786.

Hu, Y., & Loizou, P. C. (2010). Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *The Journal of the Acoustical Society of America*, *127*(6), 3689–3695.

Hugdahl, K., & Westerhausen, R. (2016). Speech processing asymmetry revealed by dichotic listening and functional brain imaging. *Neuropsychologia*, *93*, 466–481.

Jacewicz, E., & Fox, R. A. (2013). The effects of dialect variation on speech intelligibility in a multitalker background. *Applied Psycholinguistics*, *36*(3), 729–746.

Jemel, B., Oades, R. D., Oknina, L., Achenbach, C., & Röpcke, B. (2003). Frontal and temporal lobe sources for a marker of controlled auditory attention: The negative difference (Nd) event-related potential. *Brain Topography*, *15*(4), 249–262.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995–2004.

Kaya, E. M., & Elhilali, M. (2014). Investigating bottom-up auditory attention. *Frontiers in Human Neuroscience*, *8*, 1–12.

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*(2), 620–628.

Khouri, L., & Nelken, I. (2015). Detecting the unexpected. *Current Opinion in Neurobiology*, *35*, 142–147.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, *15*(3), 166–171.

Kober, S. E., Witte, M., Ninaus, M., Neuper, C., & Wood, G. (2013). Learning to modulate one's own brain activity: the effect of spontaneous mental strategies. *Frontiers in Human Neuroscience*, *7*, 1–12.

Koerner, T. K., & Zhang, Y. (2015). Effects of background noise on inter-trial phase coherence and auditory N1-P2 responses to speech stimuli. *Hearing Research*, *328*, 113–119.

Kong, Y. Y., Mullangi, A., & Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, *316*, 73–81.

Laneau, J., & Wouters, J. (2004). Multichannel place pitch sensitivity in cochlear implant recipients. *JARO - Journal of the Association for Research in Otolaryngology*, *5*(3), 285–294.

Larsby, B., Hällgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology*, *44*(3), 131–143.

Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, *307*, 111–120.

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

Li, X., & Yang, Y. (2013). How long-term memory and accentuation interact during spoken language comprehension. *Neuropsychologia*, *51*(5), 967–978.

Maamor, N., & Billings, C. J. (2017). Cortical signal-in-noise coding varies by noise type, signal-to-noise ratio, age, and hearing status. *Neuroscience Letters*, *636*, 258–264.

Manan, H. A., Yusoff, A. N., Franz, E. A., & Mukari, S. Z. S. (2017). Effects of aging and background babble noise on speech perception processing: An fMRI study. *Neurophysiology*, *49*(6), 441–452.

Mason, J. A., & Herrmann, K. R. (1998). Universal infant hearing screening by automated auditory brainstem response measurement. *Pediatrics*, *101*(2), 221–228.

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*(5), 360–368.

McFarland, D. H. (2009). Auditory system. In *Netter's atlas of anatomy for speech, swallowing, and hearing* (pp. 169–198). St. Louis: Mosby, Inc.

MED-EL. (2012). *Video about hearing and how it works*. Austria: YouTube.

MED-EL. (2017). Types of hearing loss. Retrieved November 10, 2017, from http://www.medel.com/hearing-loss/

Meyer, C., Hickson, L., Khan, A., & Walker, D. (2014). What is important for hearing aid satisfaction? Application of the expectancy-disconfirmation model. *Journal of the American Academy of Audiology*, *25*(7), 644–655.

Michie, P. T., Bearpark, H. M., Crawford, J. M., & Glue, L. C. T. (1990). The nature of selective attention effects on auditory event-related potentials. *Biological Psychology*, *30*(3), 219–250.

Mirkovic, B., Bleichner, M. G., De Vos, M., & Debener, S. (2016). Target speaker detection with concealed EEG around the ear. *Frontiers in Neuroscience*, *10*, 1–11.

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. *Journal of Neural Engineering*, *12*, 1–9.

Näätänen, R. (1982). Processing negativity: an evoked-potential reflection of selective attention. *Psychological Bulletin*, *92*(3), 605–640.

Näätänen, R., & Escera, C. (2000). Mismatch negativity: Clinical and other applications. *Audiology and Neuro-Otology*, *5*, 105–110.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590.

Neumann, N., & Birbaumer, N. (2003). Predictors of successful self control during brain-computer communication. *Journal of Neurology Neurosurgery and Psychiatry*, *74*(8), 1117–1121.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., … Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706.

Oates, P. A., Kurtzberg, D., & Stapells, D. R. (2002). Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing. *Ear and Hearing*, *23*(5), 399–415.

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology*, *88*(2), 109–123.

Papesh, M. A., Billings, C. J., & Baltzell, L. S. (2015). Background noise can enhance cortical auditory evoked potentials under certain conditions. *Clinical Neurophysiology*, *126*(7), 1319–1330.

Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear and Hearing*, *37*(1), 1–13.

Picton, T. W. (2013). Hearing in time: Evoked potential studies of temporal processing. *Ear and Hearing*, *34*(4), 385–401.

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. W. (2003). Human auditory steady-state responses. *International Journal of Audiology*, *42*(4), 177–219.

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–1662.

Poremba, A., Bigelow, J., & Rossi, B. (2013). Processing of communication sounds: Contributions of learning, memory, and experience. *Hearing Research*, *305*(1), 31–44.

Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, *35*(9), 1497–1503.

Puschmann, S., Sandmann, P., Ahrens, J., Thorne, J., Weerda, R., Klump, G., … Thiel, C. M. (2013). Electrophysiological correlates of auditory change detection and change deafness in complex auditory scenes. *NeuroImage*, *75*, 155–164.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2008). Learning effect observed for the speech reception threshold in interrupted noise with normal hearing listeners. *International Journal of Audiology*, *47*(4), 185–188.

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, *1077*(1), 135–143.

Rodenburg, M., Huizing, E. H., Kapteyn, T. S., & Wanink, A. (1979). *Revalidatie van slechthorenden*. Alphen aan den Rijn: Stafleu’s Wetenschappelijke Uitgeversmaatschappij.

RStudio Team. (2016). RStudio: Integrated development for R. Boston, MA: RStudio, Inc.

Sataloff, J., Sataloff, R. T., & Vassallo, L. A. (1980). *Hearing loss* (Second edi). Philadelphia and Toronto: J. B. Lippincott Company.

Schröger, E., Bendixen, A., Denham, S. L., Mill, R. W., Bohm, T. M., & Winkler, I. (2014). Predictive regularity representations in violation detection and auditory stream segregation: From conceptual to computational models. *Brain Topography*, *27*(4), 565–577.

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *NeuroReport*, *9*(15), 3355–3358.

Shabtai, N. R., Nehoran, I., Ben-Asher, M., & Rafaely, B. (2017). Intelligibility of speech in noise under diotic and dichotic binaural listening. *Applied Acoustics*, *125*, 173–175.

Sohoglu, E., & Chait, M. (2016a). Detecting and representing predictable structure during auditory scene analysis. *ELife*, *5*, 1–17.

Sohoglu, E., & Chait, M. (2016b). Neural dynamics of change detection in crowded acoustic scenes. *NeuroImage*, *126*, 164–172.

Solheim, J., Kværner, K. J., Sandvik, L., & Falkenberg, E. S. (2012). Factors affecting older adults' hearing-aid use. *Scandinavian Journal of Disability Research*, *14*(4), 300–312.

Srinivasan, R., & Nunez, P. L. (2012). Electroencephalography. *Encyclopedia of Human Behavior*, 15–23.

Steinschneider, M., Nourski, K. V., Rhone, A. E., Kawasaki, H., Oya, H., & Howard, M. A. (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. *Frontiers in Neuroscience*, *8*(240), 1–13.

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3), 191–201.

Sussman, E., Ritter, W., & Vaughan, H. G. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, *789*(1), 130–138.

Tabri, D., Chacra, K. M. S. A., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language and Communication Disorders*, *46*(4), 411–422.

Van den Bogaert, T. (2008). *Preserving binaural cues in noise reduction algorithms for hearing aids*. KU Leuven.

Van den Bogaert, T., Doclo, S., Wouters, J., & Moonen, M. (2009). Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids. *The Journal of the Acoustical Society of America*, *125*(1), 360–371.

Van Eyndhoven, S., Francart, T., & Bertrand, A. (2017). EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Transactions on Biomedical Engineering*, *64*(5), 1045–1056.

van Straaten, H. L. M. (1999). Automated auditory brainstem response in neonatal hearing screening. *Acta Paediatrica. Supplementum*, *88*(432), 76–79.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *Journal of the Association for Research in Otolaryngology*, *191*, 1–11.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 333–342.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, *13*(7), 994–1005.

WHO. (2004). *Guidelines for hearing aids and services for developing countries* (Second edi). Geneva: World Health Organization.

WHO. (2017). Deafness and hearing loss. Retrieved November 10, 2017, from http://www.who.int/mediacentre/factsheets/fs300/en/

Wiegand, K., Heiland, S., Uhlig, C. H., Dykstra, A. R., & Gutschalk, A. (2018). Cortical networks for auditory detection with and without informational masking: Task effects and implications for conscious perception. *NeuroImage*, *167*, 178–190.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540.

Woods, D. L. (1992). Auditory selective attention in middle-aged and elderly subjects: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, *84*(5), 456–468.

Yoon, S. H., Nam, K. W., Yook, S., Cho, B. H., Jang, D. P., Hong, S. H., & Kim, I. Y. (2017). A trainable hearing aid algorithm reflecting individual preferences for degree of noise-suppression, input sound level, and listening situation. *Clinical and Experimental Otorhinolaryngology*, *10*(1), 56–65.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225–2234.

Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., & Rönnberg, J. (2014). Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Frontiers in Neuroscience*, *8*, 1–11.

Zhang, C., Arnott, S. R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X., … Schneider, B. A. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hearing Research*, *331*, 119–130.

Zhang, C., Lu, L., Wu, X., & Li, L. (2014). Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. *Brain and Language*, *135*, 32–41.

Zink, R., Baptist, A., Bertrand, A., Huffel, S. Van, & Vos, M. De. (2016). Online detection of auditory attention in a neurofeedback application, 8–11.

List of tables

List of figures

Appendices

Appendix A: Comparison of the stories used in the AAD experiment

To check whether all stories were equally difficult, the mean intelligibility ratings for each subject were compared across stories. A Friedman's ANOVA indicated a significant variation in the intelligibility ratings (χ^2_7 = 63.622, $p < 0.001$). A multiple comparisons test showed that especially stories 2 and 3 were harder to understand than the other stories. Figure 18 supports this result.

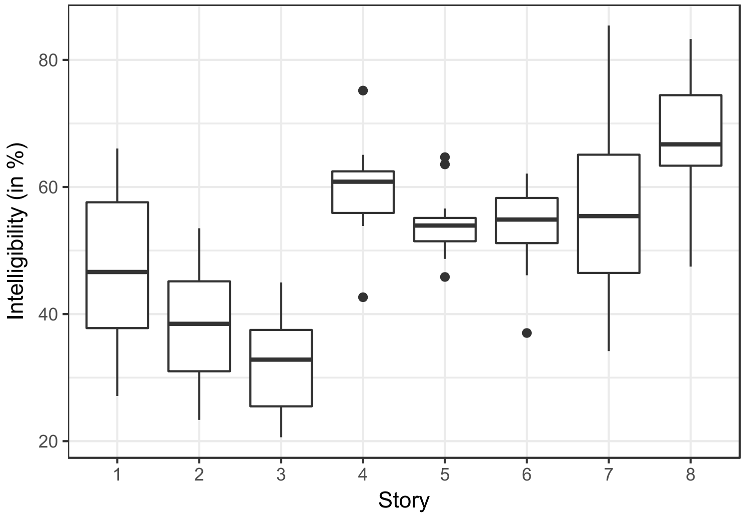

Figure 18. *Subjective intelligibility for each story.*

To identify an effect of the stories on the AAD performance, the subjects' mean decoding accuracies were also compared using Friedman's ANOVA. Again, the scores differed significantly across stories (χ^2_7 = 27.731, $p < 0.001$), as can be seen in Figure 19. A multiple comparisons test showed that performance was significantly worse for story 2 than for stories 1, 4, 5 and 8.

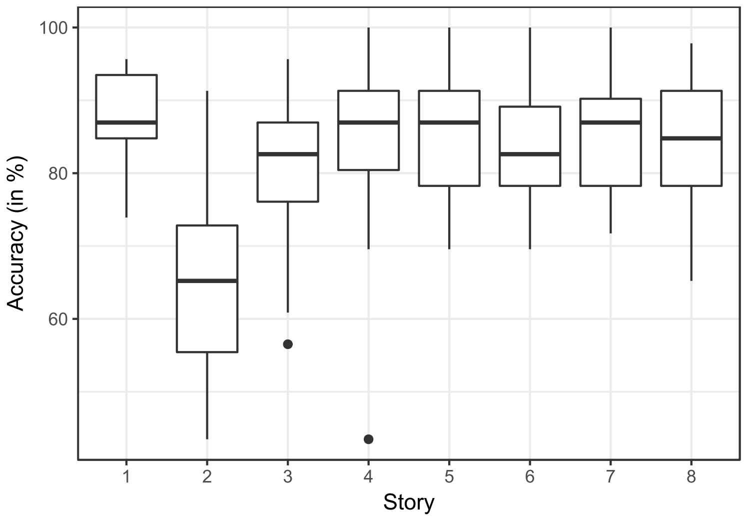

Figure 19. *Decoding accuracies for each story.*

Appendix B: Randomisation

The order of the angular conditions, the stories, and the signal-to-noise ratios were randomised for each subject. The result of this randomization is displayed in Table 6. A thorough exploration of this table shows that for every subject, the course of the experiment differed on one or more dimensions from the others. Each part of the randomisation process will be discussed below.

Table 6. *Order of the stimuli for each subject.*

Subject	Noise conditions	Angular conditions
1	1-2-3	2-3-1-4
2	2-3-1	3-2-4-1
3	3-1-2	4-2-3-1
4	1-2-3	1-3-4-2
5	2-3-1	2-1-3-4
6	3-1-2	4-1-3-2
7	1-2-3	4-2-1-3
8	2-3-1	3-1-2-4
9	3-1-2	4-2-1-3
10	1-2-3	2-1-4-3
11	1-2-3	1-3-2-4
12	3-1-2	3-1-4-2
13	1-2-3	1-2-3-4
14	2-3-1	1-3-2-4
15	3-1-2	1-2-3-4

First, the stories were combined into four pairs. Two of these pairs consisted of long stories (19 minutes each), the two others were short stories (9,5 minutes each). Two possible variations emerged: the presentation order of the stories within a pair could be inversed, and each story pair could be used in either of two angular conditions. To subjects with an odd number, the odd story was presented first (1-2, 3-4, 5-6, 7-8), whereas for subjects with an even number, this pattern was reversed (2-1, 4-3, 6-5, 8-7). This order also determined the angular condition of each story pair and each story's angle. For odd numbers, story 1 was used in angular condition 1, coming from a -5° angle. As a result, story 2 was presented coming from a 5° angle. In condition 2, stories 3 (30°) and 4 (90°) were presented. Condition 3 contained stories 5 (-30°) and 6 (-90°), whereas condition 4 presented stories 7 (-90°) and 8 (90°). For even numbers, this was reversed. In angular condition 1, stories 8 (5°) and 7 (-5°) were presented. Condition 2 used stories 6 (90°) and 5 (30°), condition 3 had stories 4 (-90°) and 3 (-30°), and condition 4 contained stories 2 (90°) and 1 (-90°). This distribution is shown by Table 7.

Table 7. *Distribution of the stories.*

Participant number	Angular condition	Story A	Story B
Odd	1	1 (-5°)	2 (5°)
	2	3 (30°)	4 (90°)
	3	5 (-30°)	6 (-90°)
	4	7 (-90°)	8 (90°)
Even	1	8 (5°)	7 (-5°)
	2	6 (90°)	5 (30°)
	3	4 (-90°)	3 (-30°)
	4	2 (90°)	1 (-90°)

Next, the sequence of the noise conditions was chosen. The first trial was always presented without babble noise, but for the second to fourth part of each story, babble noise of different levels was added. For the first subject, this sequence was SNR1-SNR2-SNR3, as shown in **Fout! Verwijzingsbron niet gevonden.** as '1-2-3'. For the second subject, it was changed to SNR2-SNR3-SNR1 ('2-3-1'). The third subject was assigned the sequence SNR3-SNR1-SNR2 ('3-1-2'). This combination of three sequences was then repeated for the other subjects. Due to a human error during the experiment, the noise sequence of subject 11 did not comply with this routine; instead, the subject followed the sequence from subject 10.

Finally, the angular condition order for each subject was defined. Four conditions, without any order restrictions, resulted in 24 possible sequences. The computer randomly assigned a sequence to each subject. As a result, some sequences were used twice.

Appendix C: Assumptions for parametric tests

To determine whether parametric tests could be used, four assumptions had to be checked: normal distribution, homogeneity of variance, interval data and independence. First, the mean decoding result was calculated for each trial, i.e. the performance of a subject for each combination of angular condition, attended speaker, and noise level. This resulted in a new variable, decoding accuracy, which contained 480 values and was used for all subsequent analyses. As the subjective intelligibility had been rated for each trial, no further manipulations were necessary.

Decoding accuracy

The decoding accuracy was not normally distributed, since the values could only vary between 0 and 100. The Shapiro-Wilk normality test confirmed that the decoding accuracy was significantly not-normal ($W = 0.843$, $p < 0.001$). This result is supported by the histogram in Figure 20, which contains the mean decoding accuracy for each unique combination of a separation angle, a noise level, an attended direction, and a subject (as explained above). The overall mean accuracy was 0.82, with a standard deviation of 0.20. The value for skewness was -1.411 and kurtosis equalled 1.870, indicating a heavily left-tailed distribution (left-skewed, leptokurtic distribution).

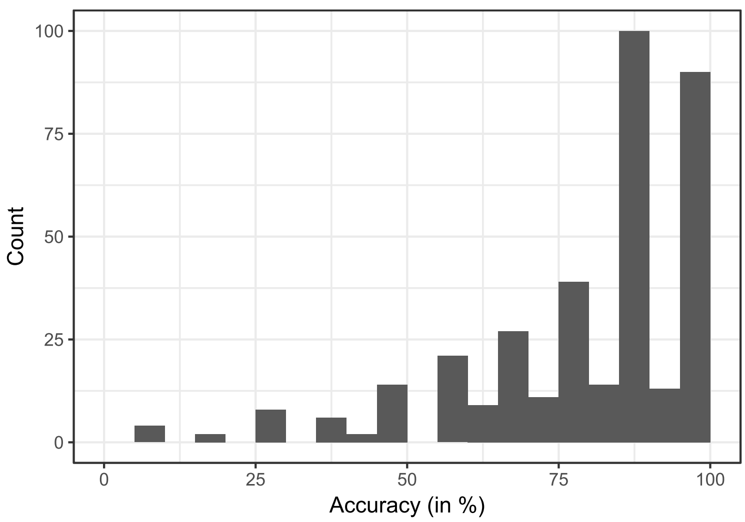

Figure 20. *Histogram of decoding accuracies*.

The distribution was also visualised separately for each factor. Figure 21 contains the mean decoding accuracy per subject for each separation angle. Figure 22, on the other hand, shows the mean decoding accuracy per subject for each noise level.

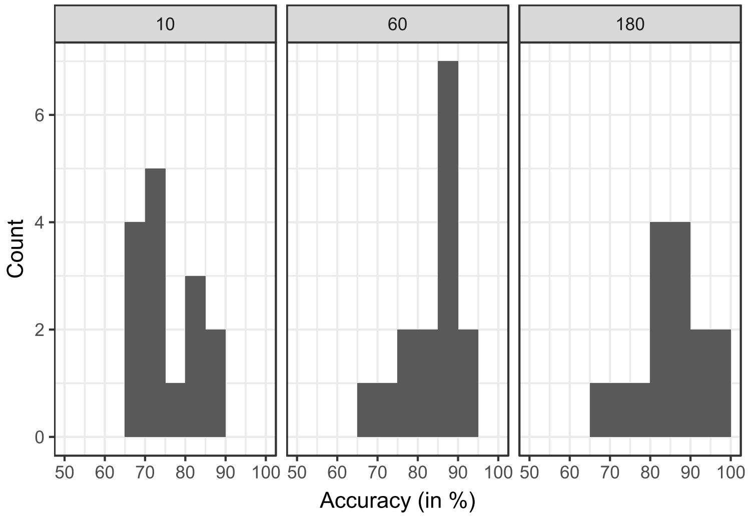

Figure 21. *Histogram of mean decoding accuracies per subject, for each separation angle.*

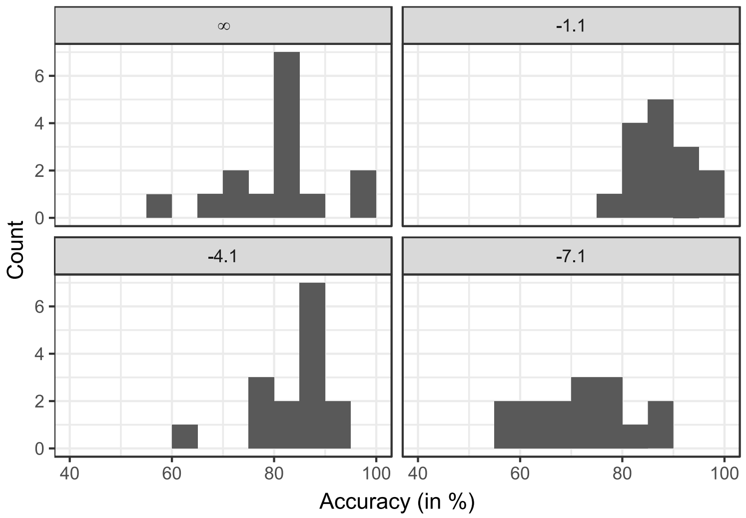

Figure 22. *Histogram of mean decoding accuracies per subject, for each noise level.*

Secondly, the homogeneity of variance was investigated. To this end, the Levene's test for homogeneity of variance was carried out. The variances were similar for the different subjects ($F_{14,465}$ = 1.507, p = 0.104). However, the variance in decoding accuracy was significantly different across separation angles ($F_{2,477}$ = 5.821, p = 0.003), across sides ($F_{1,478}$ = 10.594, p = 0.001), and across noise levels ($F_{3,476}$ = 12.453, p < 0.001).

Thirdly, the assumption of interval level data was checked. Although the original data points are binary, because the decoder either makes a right (1) or wrong (0) decision, the decoding accuracy is based on a mean value and can therefore be considered an interval variable. Thus, the third assumption was met. Finally, the independence of the data was determined. In a repeated-measures design, this means that there is independence across subjects. Since every subject was tested separately, there was no contact between subjects and the independence was assured.

In conclusion, as the accuracy data had a non-normal distribution and there was heterogeneity of variance, non-parametric tests had to be used.

Subjective intelligibility

For the subjective intelligibility, the values were again limited between 0 and 100, and could therefore not be normally distributed. This was confirmed by a Shapiro-Wilk normality test ($W = 0.934$, $p < 0.001$). The histogram in Figure 23 contains one intelligibility rating for each unique combination of a separation angle, a noise level, an attended direction, and a subject.

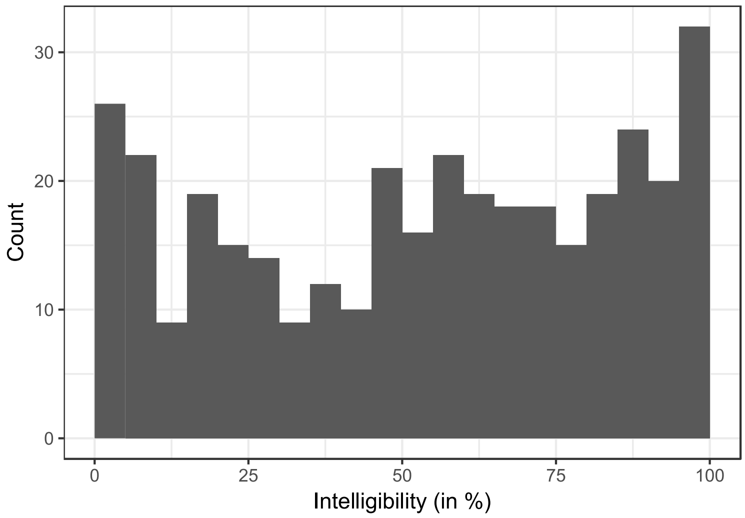

Figure 23. *Histogram of subjective intelligibility.*

Next, the distribution was visualised separately for both separation angle and noise level. The resulting histograms show a mean intelligibility rating per subject, either for each separation angle (Figure 24) or for each noise level (Figure 25).

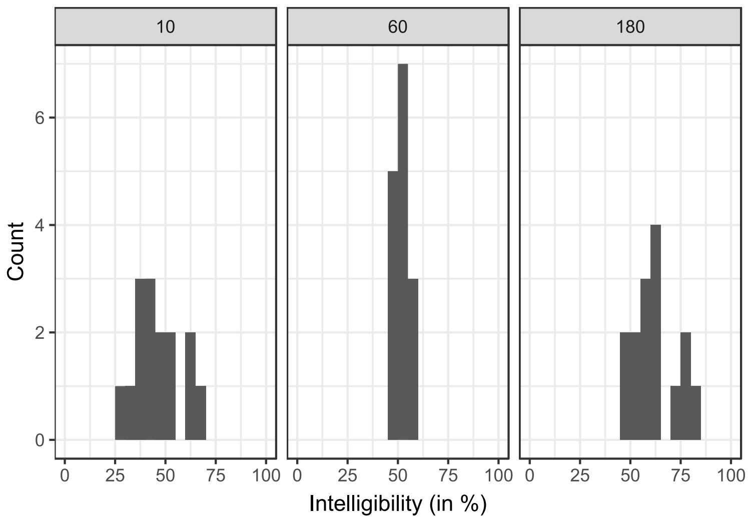

Figure 24. *Histogram of mean subjective intelligibility per subject, for each separation angle*.

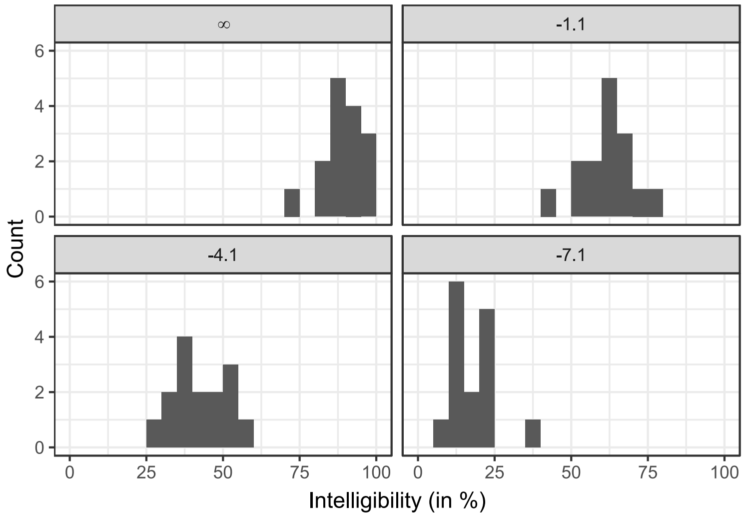

Figure 25. *Histogram of mean subjective intelligibility per subject, for each noise level.*

Secondly, the Levene's test for homogeneity of variance was carried out. The variances were similar for the different sides of the attended speaker ($F_{1,478} = 0.208$, $p = 0.649$). However, the variance in intelligibility was significantly different across separation angles ($F_{2,477} = 5.640$, $p = 0.004$), across noise levels ($F_{3,476} = 19.963$, $p < 0.001$), and across subjects ($F_{14,465} = 3.122$, $p < 0.001$).

Thirdly, the data had to be measured at interval level at least. Since subjects had to rate the intelligibility on a continuous scale from 0 to 100, this assumption was met. Finally, as each subject was tested separately, the independence of the results was assured.

In conclusion, based on the non-normal distribution of the intelligibility ratings and the heterogeneity of variance for some factors, non-parametric tests were used.

Appendix D: Coefficients for the LME model predicting accuracy

To predict the decoding accuracy, a linear mixed-effects model fit by restricted maximum likelihood was constructed. This model included separation angle, noise level, and the interaction between them as fixed factors. Furthermore, it considered subject as a random factor. Based on this model, mean predicted accuracies were constructed for each of the fixed factors. These estimates are given in Table 8, Table 9, and Table 10.

Table 8. *Accuracy estimates for separation angle.*

separation angle	estimate	95% confidence interval		standard error
		lower	*upper*	
10°	0.754	0.706	0.801	0.022
60°	0.835	0.795	0.876	0.019
180°	0.851	0.804	0.899	0.022

Table 9. *Accuracy estimates for noise level.*

noise level	estimate	95% confidence interval		standard error
		lower	*upper*	
SNR0	0.807	0.758	0.856	0.023
SNR1	0.879	0.830	0.928	0.023
SNR2	0.837	0.788	0.885	0.023
SNR3	0.731	0.682	0.780	0.023

Table 10. *Accuracy estimates for the interaction between separation angle and noise level.*

interaction	estimate	95% confidence interval		standard error
		lower	*upper*	
10° - SNR0	0.831	0.755	0.908	0.036
10° - SNR1	0.823	0.747	0.900	0.036
10° - SNR2	0.730	0.653	0.807	0.036
10° - SNR3	0.630	0.553	0.707	0.036
60° - SNR0	0.804	0.745	0.863	0.027
60° - SNR1	0.920	0.861	0.979	0.027
60° - SNR2	0.883	0.825	0.942	0.027
60° - SNR3	0.733	0.675	0.792	0.027
180° - SNR0	0.785	0.709	0.862	0.036
180° - SNR1	0.893	0.817	0.970	0.036
180° - SNR2	0.897	0.820	0.973	0.036
180° - SNR3	0.830	0.753	0.910	0.036

To further evaluate the model, the residuals were investigated. Figure 26 displays the standardized residual for each fitted value. Since these values are not randomly scattered, forming a pattern instead, the model may not be a perfect fit for the data.

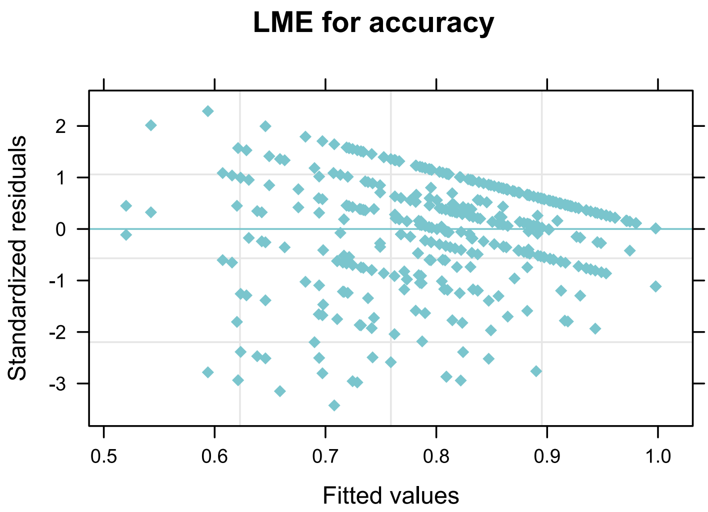

Figure 26. *Linear mixed-effects model for decoding accuracy*.

The median residual equalled 0.068, with an interquartile range of -0.098 to 0.109. A Shapiro-Wilk normality test, together with the histogram in Figure 27, indicated that the residuals are not normally distributed ($W = 0.938$, $p < 0.001$). However, the Q-Q plot in Figure 28 is acceptable. In conclusion, the model is not a perfect fit but can be used for analyses nonetheless. It may be possible to improve the model by treating the noise level as a numeric factor and/or excluding the trials without noise. However, this possibility was not explored in this work.

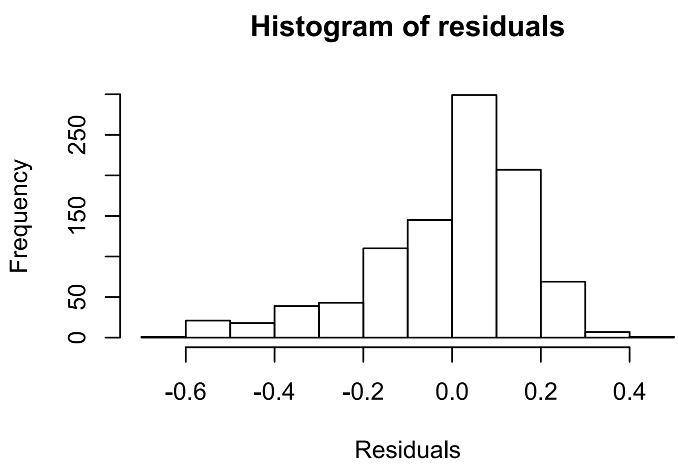

Figure 27. *Histogram with residuals from LME for predicting accuracy*.

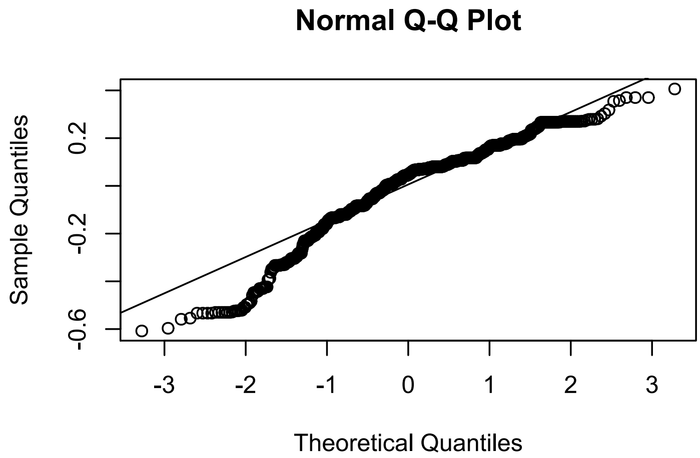

Figure 28. *Q-Q-plot with residuals from LME for predicting accuracy*.

Appendix E: Exploration of the right ear advantage

In the first research question, the effect of spatial separation was investigated. It was hypothesized that a right ear advantage (REA) might be represented in the data. However, global analysis of the AAD accuracies for the 10° and 180° separation angles revealed no REA; on the contrary, a small left ear advantage was found. To explore this finding, post-hoc analysis with multiple Wilcoxon signed-rank tests was used to check whether the effect was the same for each separation angle. The distributions of the mean decoding accuracy of each subject, labelled as belonging to a left- or right-sided attended speaker and grouped within each separation angle, are visualized in Figure 29.

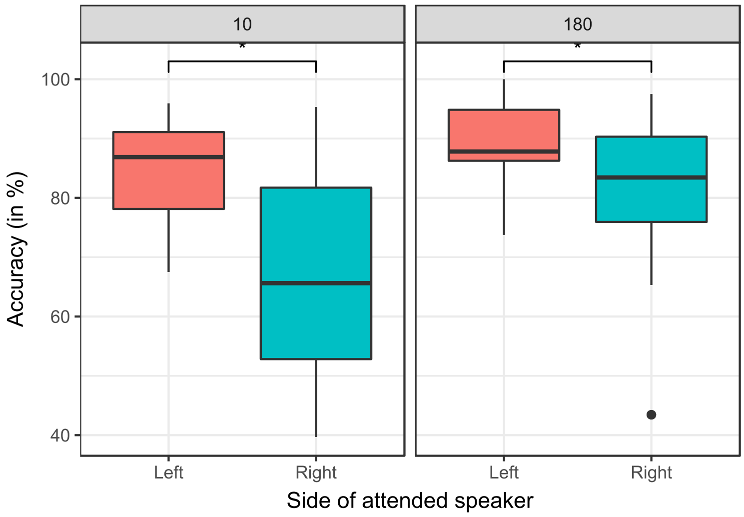

Figure 29. *Decoding accuracies for each attended side, compared across separation angles.*

Table 11 shows the statistical results for each separation angle. For both 10° and 180° separation angles, accuracies were significantly higher when the attended speaker was on the left side of the subject. The effect was larger for 10° angles.

Table 11. *Wilcoxon signed-rank tests for side of attended speaker.*

separation angles	W	p	r	corrected p
10°	101	0.021	-0.420	0.041
180°	96.5	0.041	-0.373	0.041

Appendix F: Coefficients for the LME model predicting intelligibility

A second LME model was constructed to predict the subjective intelligibility, again fit by restricted maximum likelihood. The model was based on the same fixed factors: separation angle, noise level, and the interaction between them. Subject was again added as a random factor. Table 12, Table 13, and Table 14 contain the mean predicted accuracies for each factor.

Table 12. *Intelligibility estimates for separation angle.*

separation angle	estimate	95% confidence interval		standard error
		lower	*upper*	
10°	0.461	0.424	0.498	0.017
60°	0.518	0.490	0.547	0.013
180°	0.625	0.588	0.662	0.017

Table 13. *Intelligibility estimates for noise level.*

noise level	estimate	95% confidence interval		standard error
		lower	*upper*	
SNR0	0.895	0.857	0.933	0.018
SNR1	0.626	0.588	0.665	0.018
SNR2	0.429	0.391	0.468	0.018
SNR3	0.189	0.151	0.228	0.018

Table 14. *Intelligibility estimates for the interaction between separation angle and noise level.*

interaction	estimate	95% confidence interval		standard error
		lower	*upper*	
10° - SNR0	0.894	0.827	0.961	0.031
10° - SNR1	0.536	0.469	0.603	0.031
10° - SNR2	0.293	0.226	0.360	0.031
10° - SNR3	0.122	0.055	0.189	0.031
60° - SNR0	0.890	0.841	0.939	0.023
60° - SNR1	0.613	0.564	0.662	0.023
60° - SNR2	0.424	0.375	0.473	0.023
60° - SNR3	0.147	0.098	0.196	0.023
180° - SNR0	0.902	0.835	0.969	0.031
180° - SNR1	0.729	0.662	0.796	0.031
180° - SNR2	0.570	0.503	0.637	0.031
180° - SNR3	0.298	0.232	0.365	0.031

The residuals were checked to evaluate the model. In Figure 30, the standardizes residuals are plotted for each fitted value. The data points are not scattered randomly, suggesting that the model may not be a perfect fit for the data.

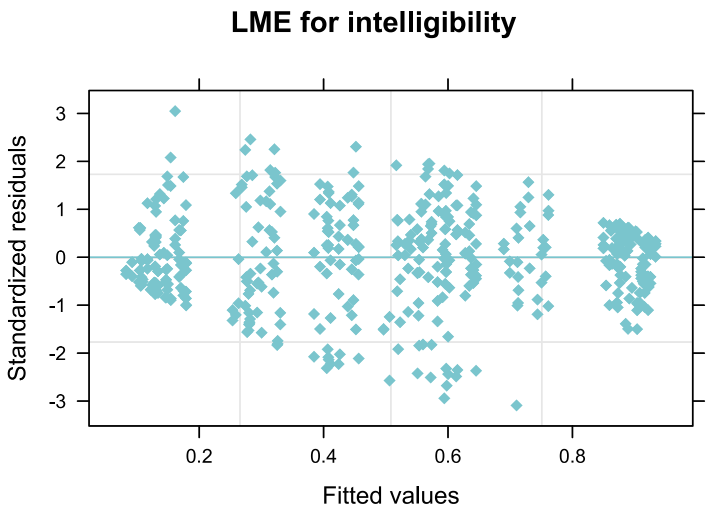

Figure 30. *Linear mixed-effects model for subjective intelligibility.*

The median of the residuals equalled 0.008, with an interquartile range of -0.098 to 0.095. A Shapiro-Wilk normality tests indicated that the residuals may not be normally distributed (W = 0.992, p < 0.001). However, the histogram in Figure 31 and the Q-Q plot in Figure 32 look acceptable. Thus, the applied model may not be perfect, but it is still useful for analysis purposes.

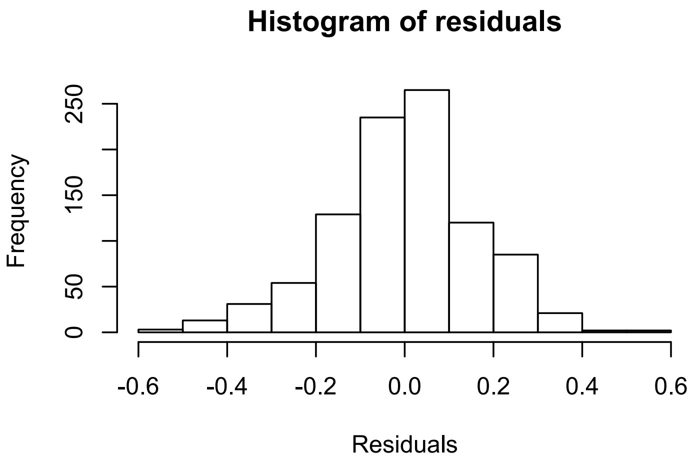

Figure 31. *Histogram with residuals from LME for predicting intelligibility.*

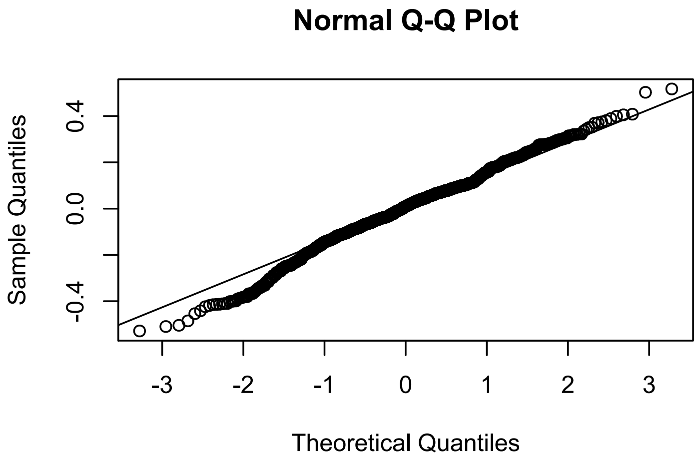

Figure 32. *Q-Q-plot with residuals from LME for predicting intelligibility.*

LOGOPEDISCHE EN AUDIOLOGISCHE WETENSCHAPPEN
Herestraat 49/721
3000 LEUVEN, België