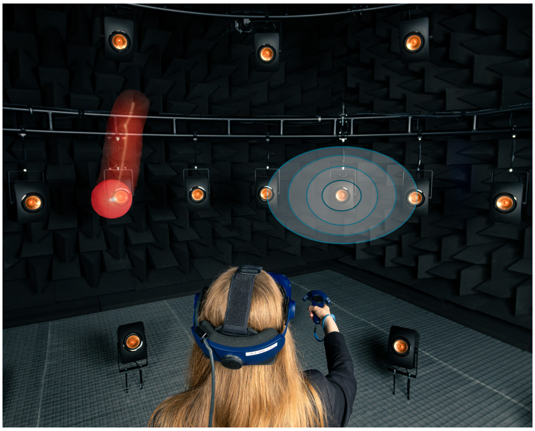
CONTRIBUTIONS TO
HEARING RESEARCH

Volume 48

Thirsa Huisman

The Influence of Vision on Spatial Localization in Normal-Hearing and Hearing-Impaired Listeners



DTU Health Tech
Department of Health Technology

The Influence of Vision on Spatial Localization in Normal-Hearing and Hearing-Impaired Listeners

PhD thesis by
Thirsa Huisman

Preliminary version: September 9, 2021

Centre for Applied Hearing Research

Technical University of Denmark

2021

Preprint version for the assessment committee.
Pagination will differ in the final published version.
Cover picture: © Niclas. A. Janßen, 2021

This PhD dissertation is the result of a research project carried out at the Hearing Systems Group, Department of Health Technology, Technical University of Denmark.

The project was partly financed by CAHR (2/3) and by the Technical University of Denmark (1/3).

Supervisors

Prof. Dr. Torsten Dau
Hearing Systems Group
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark
Assoc. Prof. Dr. Ewen N. MacDonald
Systems Design Engineering
University of Waterloo
Waterloo, Canada
Dr. Tobias Piechowiak
GN Hearing
Region Hovedsteden, Denmark

Abstract

Hearing-impaired listeners and aided hearing-impaired listeners have been shown to have degraded auditory localization abilities in auditory-only conditions, where information from other sensory modalities is not available to the listener. However, it is unclear how auditory localization performance in such listeners is affected in more realistic, daily-life, conditions, where they have access to additional cues that may aid localization, such as visual and self-motion cues. This thesis investigated how visual information affects spatial localization in normal and hearing-impaired listeners.

In the first study, a new analysis method was developed to distinguish between integration, i.e., a shift in perception, and response biases, i.e., a shift in decision making, in the spatial ventriloquist effect, a well-known phenomenon of audio-visual integration where the perceived location of an auditory stimulus is shifted towards the location of a visual stimulus. Response biases can result in an overestimation of both the shift in the perceived location of the auditory stimulus and the 'spatial integration window', i.e., the spatial distance in the horizontal plane between the auditory and the visual stimuli up to which they are integrated. Data of normal-hearing participants was gathered using this ventriloquist paradigm. A Gaussian clustering method was then used to cluster the localization data. These clusters were categorized into integrated, non-integrated and response bias clusters to allow for an unbiased analysis. With this new analysis method, the results showed that the spatial integration window is asymmetric, ranging from about -12 to +28 degrees, with a negative value indicating that the visual stimuli occurred closer to the center compared to the auditory stimulus.

The second study explored the effect of stimulus realism on the spatial ventriloquist effect, by comparing the visual bias evoked with various sets of stimuli, such as a 'non-realistic' noise burst and a light flash vs. a 'realistic' bouncing ball and an impact sound. As in the first study, it was found that the relative stimulus positioning affected the probability of integration. However, no effect of stimulus realism was found, i.e., the naturalness of the stimuli did not consistently affect the results. This is important as it suggests that the results from laboratory studies using non-natural stimuli will generalize to realistic situations with natural stimuli.

Virtual reality goggles have been shown to modify spatial localization cues and affect auditory localization. The third study investigated the effect of virtual

reality goggles on the perceived location of sounds that were reproduced using ambisonics with and without visual information about the position of the loudspeakers. Participants perceived sounds to be further outwards when wearing the virtual reality goggles. This effect was found to be larger in the right than in the left hemisphere and it was largest around ±52.5 degrees azimuth. When visual information was available, auditory localization was strongly biased towards the visual sources. This bias towards visual sources generally improved localization accuracy, as compared to blindfolded auditory localization, when the auditory stimulus was simulated at a loudspeaker location. However, when the auditory stimulus was simulated in between loudspeakers, participants localized the auditory sources more accurately without visual information.

The fourth study investigated spatial integration in young normal-hearing, older normal-hearing and older hearing-impaired listeners to explore how age and hearing loss affect the spatial integration window. For this, a modified version, i.e., using relative instead of absolute localization, of the ventriloquist's paradigm was used. The results demonstrated that the spatial integration window was increased in older listeners. However, no difference was found between older normal and older hearing-impaired listeners.

Finally, the last study explored congruent audio-visual localization behavior and how this is affected by the number of auditory distractors. When the number of auditory distractors was low, the audio-visual area localization time, i.e., the time it took participants to get the target within their field of view, was consistent with the audio-only area localization time. However, as the number of distractors increased, visual information became more important. Audio-visual area localization times were significantly shorter than in audio-only conditions. Moreover, head-motion data showed that participants modified their behavior as the number of auditory distractors increased. Audio-visual target localization times, i.e., the time it took participants to find the target when it was already within the field of view, were consistently smaller than both auditory-only and visual-only target localization times. These results show that, instead of audio-visual localization being a combination of auditory area localization and visual target localization, the auditory and visual system contribute to both the area localization and the target localization.

Together, the experiments in this thesis demonstrate that visual information strongly influences auditory localization. The occurrence of the shift in the perceived location of auditory stimuli as a result of visual stimuli was affected by both the absolute and relative stimulus positioning as well as the participants' age. However, realism, movement and hearing loss did not affect integration, at least when the stimuli were presented from the front direction. While auditory localization of hearing-impaired listeners was strongly biased towards visual information, the probability for this shift to occur was not higher than in normal-hearing listeners of the same age. Considering audio-visual localization behavior at increased angles, both the auditory and visual system were shown to contribute to finding the approximate area of a target and finding the target

when it was within the field of view. Overall, these results show a strong connection of the auditory and visual system that was, at least in the front, unaffected by a hearing loss. These results may guide future research on audio-visual localization in hearing-impaired and aided-hearing impaired listeners and are likely to help in the design of new hearing-aid processing algorithms or deciding between already existing algorithms.

Resumé

Tidligere studier har vist, at hørehæmmede personer og høreapparatsbrugere er dårligere til at lokalisere i rene auditive eksperimenter, hvor stimuli er auditivt og hvor der ikke er adgang til information fra andre sanser. Det er dog uvist, hvordan evnen til at bruge auditiv lokalisering er påvirket i sådanne personer under mere realistiske betingelser, der bedre afspejler deres dagligdag, hvor de også har adgang til anden information fra andre sanser (som f.eks. syn og selvbevægelse), som muligvis kan forbedre evnen til at lokalisere. Denne afhandling undersøgte, hvordan visuel information påvirker rumlig lokalisering i normalthørende og hørehæmmede personer.

I det første studie blev en ny analysemetode udviklet, der kan skelne mellem integration, dvs. en ændring i opfattelse, og "response bias", dvs. en ændring i beslutningstagning, i den rumlige bugtalereffekt (et velkendt fænomen indenfor audiovisuel integration, hvor den opfattede placering af en auditiv stimulus er flyttet mod placeringen af den visuelle stimulus). "Response bias" kan føre til en overvurdering af både ændringen af den opfattede placering af det auditive stimulus og det 'rumlige integrations vindue', dvs. den rumlige afstand i det horisontale plan mellem det auditive og den visuelle stimulus til det punkt, hvor de opfattes som værende integreret. Data fra normalthørende personer var indsamlet, hvor der blev gjort brug af dette bugtalerparadigme. En gaussisk klyngemetode blev benyttet til at gruppere disse lokaliseringsdata. Disse klynger blev kategoriseret som integreret, ikke-integreret og 'response bias' klynger for a kunne lave en objektiv analyse. Med denne nye analysemetode, viste resultaterne, at det rumlige integrationsvindue er asymmetrisk. Det spænder mellem -12 til +28 grader, og en negativ rumlig værdi indikerer, at den visuelle stimulus opfattes som værende tættere på midten end den auditive stimulus.

Det andet studie udforskede effekten af stimulus realisme på den rumlige bugtaler effekt ved at sammenligne den visuelle "bias" fremkaldt med forskellige sæt stimuli, så f.eks. 'ikke-realistisk' udbrud af støj og et lysglimt versus en 'realistisk' hoppende bold og en slaglyd. Som i det første studie, viste det, at den relative placering af stimulus påvirker sandsynligheden for at opnå integration. Dog var der ingen effekt af stimulus realisme, dvs. naturligheden af stimuli påvirkede ikke regelmæssigt resultaterne. Dette er vigtigt, da det antyder, at resultater fra laboratorie studier, der bruger ikke-realistiske stimuli kan generaliseres til realistiske situationer med realistiske stimuli.

Det er tidligere blevet vist, at 'Virtual reality' briller ændrer rumlig lokalisering og påvirker auditiv lokalisering. Det tredje studie undersøgte effekten af

virtual reality briller på den opfattede rumlige placering af lyd i rummet, der var reproduceret med 'Ambisonic' med og uden visuel information, der angiver højtalernes placering. Forsøgspersonerne opfattede lyden som værende længere væk i den udadgående retning, når de brugte virtual reality briller. Denne effekt var større i den højre end den venstre hemisfære og den var størst ved omkring ±52.5 graders i den horisontale retning. Når der var visuel information til rådighed, var auditiv lokalisering stærkt påvirket mod den visuelle kilde. Denne hældning mod visuelle kilder forbedrede lokaliseringsnøjagtighed i sammenligning med auditiv lokalisering med bind for øjnene ved de fleste horisontale vinkler, når det auditive stimulus var simuleret til at komme fra samme position som en af højtalerne. Derimod når det auditive stimulus var simuleret som kommende fra en position mellem højtalerne, lokaliserede forsøgspersonerne de auditive kilder mere nøjagtigt i forsøgsbetingelser uden visuel information.

Det fjerde studie undersøgte den rumlige integration i unge normalthørende, ældre normalt hørende og ældre hørehæmmede forsøgspersoner for at finde ud af, hvordan alder og høretab påvirkede det rumlige integrationsvindue. Her blev en modificeret udgave af bugtalerparadigmet brugt, der benyttede relativ i stedet for absolut lokalisering. Resultaterne viste, at det rumlige integrationsvindue var forøget i ældre personer. Dog var der ingen forskel mellem ældre normalthørende og ældre hørehæmmede personer.

Det sidste studie undersøgte kongruente audiovisuel omådelokaliseringsadfærd, og hvordan denne er påvirket af antallet af forstyrrende auditive kilder. Når antallet af forstyrrende lydkilder var lavt, var den audiovisuelle områdelokaliseringstid, dvs. den tid det tager forsøgspersoner at få målet indenfor deres synsfelt, var det samme som områdelokaliseringstiden for forsøgsbetingelser med kun auditiv information. Dog som antallet af forstyrrende lydkilder blev øget, blev visuel information mere vigtig. Her var audiovisuel områdelokaliseringstid væsentligt kortere end i forsøgsbetingelser med kun auditiv information. Desuden, viste hovedbevægelsesdata, at forsøgspersoner ændrede deres adfærd, når antallet af auditive forstyrrende lydkilder øges. Audiovisuel mållokaliseringstid, dvs. tiden det tager forsøgspersoner at finde målet, når det allerede var i deres synsfelt, var konsekvent mindre end for forsøgsbetingelser med kun auditiv information og forsøgsbetingelser med kun visuel information. Disse resultater viser at, i stedet for at audiovisuel lokalisering er en kombination af auditiv områdelokalisering og visuel mållokalisering, bidrager det auditive og visuelle system til både områdelokalisering og mållokalisering.

Tilsammen viser eksperimenterne i denne afhandling at visuel information har en stor indflydelse på auditiv lokalisering. Forekomsten af flytningen i den opfattede placering af auditive stimuli pga. visuelt stimuli blev påvirket af absolut og relativ placering af stimuli og af forsøgspersonens alder. Derimod påvirkede realisme, bevægelse og høretab ikke integration, i det mindste når stimuli blev præsenteret forfra. Derfor, mens auditiv lokalisering hos hørehæmmede forsøgspersoners er mest påvirket af visuel information, er sandsynligheden for at se en sådan ændring ikke væsentligt højere end hos normalthørende personer på den samme alder. Når man betragter audiovisuel lokaliseringsadfærd, når man øger den horisontale vinkel, bidrager både det auditive og det visuelle system til at finde det omtrentlige målområde og til at finde målet når det er indenfor synsfeltet. Samlet set viser disse resultater en stærk forbindelse mellem det auditive og det visuelle system som ikke var påvirket af høretab i det mindste når stimuli kommer forfra. Disse resultater kan guide fremtidig forskning i audiovisuel lokalisering i hørehæmmede personer med og uden høreapparat og kan være behjælpelige i forbindelse med design af nye høreapparats signalbehandlingsalgoritmer og i valget imellem allerede-eksisterede algoritmer.

Acknowledgments

“It’s better to take a step that’s suboptimal, then to never take a step at all”
-Ewen MacDonald

My Ph.D. studies have been an intense and, at times, challenging process where I got the opportunity to delve into the topic of audio-visual integration, to develop my scientific and technical skills, and to grow also personally. All of this would not have been possible without those around me, so here I would like to express my gratitude to those who have supported on my path.

First, I would like to sincerely thank my supervisors, Ewen MacDonald, Tobias Piechowiak and Torsten Dau, for giving me this opportunity and guiding me through the process. Although I am not sure we always took the optimal steps, looking back it is amazing how far we have gotten and how much we have learned along the way. None of this progress would have been possible without the discussions we have had and the feedback you gave me, so thank you. Although not officially a supervisor, I would also like to thank Axel Ahrens for his help throughout the years. Your help on how to work with AVIL was instrumental in getting me started with my project and your feedback and advice in the last year of my Ph.D. was instrumental in wrapping it up. Next, I would like to thank all my colleagues at the Hearing Systems for their support, discussions and generally just making this place such an amazing place to work, especially office 102. Special thanks go to Kasper Lund for his work on the VR environment and assistance with everything VR related, Rikke Skovhøj Sørensen for her help with the participants, measuring audiograms etc., and Sara Miay Kim Madsen for her help with the Danish abstract. Next, I would like to thank the Center for Applied Hearing Research, a consortium of GN Hearing, Oticon and WSA, for funding this work. Of course, all these studies would have been impossible without our participants, so to all those who participated; thank you for your time and effort. A big thank you also to Tobias Piechowiak and colleagues for my time at GN Hearing. Moreover, although the external stay

unfortunately could not happen there, I would like to sincerely thank Tifanie Bouchara from the Conservatoire National des Arts et Métiers and the Computer Science laboratory CEDRIC, hopefully we get to meet someday in person still.

I also would have never made it to where I am now, without the love and support of my friends and family. Thank you so much for always being there for me. Last, but definitely not least, I would thank my partner, Niclas, for his love and support throughout this process and for taking and editing the cover picture. Thank you so much for always helping me back up.

Related publications

Journal papers

- Huisman, T., Dau, T., Piechowiak, T., and MacDonald, E. (**2021**). "The ventriloquist effect is not consistently affected by stimulus realism", J. Percept. Imaging. (Accepted)
- Huisman, T., Ahrens, A., and MacDonald, E. (**2021**). "Ambisonics sound source localization with varying amount of visual information in virtual reality", Frontiers in Virtual Reality. (Under revision)
- Huisman, T., MacDonald, E., Piechowiak, T., and Dau, T. (**2021**). "Increase of the audio-visual spatial integration window with age but not hearing loss", Trends in Hearing. (Submitted)

Conference papers

- Huisman, T., Dau, T., Piechowiak, T., and MacDonald, E. (**2020**). "Audio-visual sound localization in virtual reality", Proceedings of the International Symposium on Auditory and Audiological Research. **7**, 349-356.

Contents

1

General introduction

Previous studies have demonstrated that hearing-impaired listeners typically show degraded spatial localization abilities; in particular, age-related hearing loss is associated with an increased susceptibility to front-back confusions and a decreased vertical acuity (e.g., Häusler et al., 1983; Noble et al., 1994; Otte et al., 2013; Rakerd et al., 1998). Moreover, while hearing aids are a common rehabilitation strategy for treating hearing loss, their effects on spatial perception remains somewhat unclear, with some studies finding that hearing aids improve unimodal localization performance (Chung et al., 2008; Jensen et al., 2013; Keidser et al., 2009, 2006), whereas other studies find that hearing aids degrade it further (Keidser et al., 2006; Van Den Bogaert et al., 2011; Van den Bogaert et al., 2006). However, people do not experience the world through just one sensory system; each object, person and thing produces auditory, visual, haptic information, etc. By combining information from the various sensory systems, referred to as multisensory integration, more complete, precise and accurate knowledge of our surrounding can be obtained (e.g., Ernst and Banks, 2002; Freeman et al., 2018; Lovelace et al., 2003; Odegaard et al., 2015). Thus, it is not necessarily clear how well results from unimodal localization studies generalize to more realistic, multi-modal situations. This thesis investigated how visual cues influence spatial hearing, both in normal-hearing and hearing-impaired listeners. Gaining a better understanding of the challenges that hearing-impaired listeners face with regards to spatial localization and how hearing aids affects this in realistic settings that include visual information is likely to help in the design of new hearing-aid processing algorithms or deciding between already existing algorithms.

1.1 Multisensory integration

By combining the information from the various sensory systems, localization accuracy and precision can be improved (e.g., Alais and Burr, 2004; Freeman et al., 2018; Odegaard et al., 2015), stimulus detection can be enhanced (e.g., Andersen and Mamassian, 2008; Frassinetti et al., 2002; Lovelace et al., 2003; Noesselt et al., 2008) and reaction times can decreased (e.g., Diederich and Colonius, 2004; Miller, 1982; Schröger and Widmann, 1998). While the benefits of integration are clear, there is still a lot of discussion about how stimuli are integrated. The current thinking is that integration works by weighting the stimuli relative to their reliability (e.g., Alais and Burr, 2004; Körding et al., 2007). Since the sensory systems process information differently some temporal, spatial or even informational misalignment between stimuli does not necessarily prevent integration. In fact, stimuli have been shown to be integrated over a range of temporal and spatial disparities, referred to as the temporal and spatial integration window (Chen and Vroomen, 2013; Lewald and Guski, 2003; Stenzel et al., 2019). When these misaligned stimuli are integrated, they are combined and perceived as one signal with one location, one point in time, etc. The relative weighting of stimuli means that each signal contributes to the perceived timing, location etc., relative to the inverse of their variance. This relative weighting is included in the current standard models: the maximum likelihood estimation (MLE) model (Ernst and Banks, 2002) and Bayesian Causal Inference (BCI) model (Körding et al., 2007). These models differ from each other mainly in the way that they assess if integration happens; The MLE assumes that integration always happens, whereas the BCI model uses the Bayes rule of conditional probability to assess, for a given auditory and visual signal, the likelihood that these signals originated from a common cause or that they originated from separate causes. As the MLE model assumes the forced fusion of all stimuli, which is clearly not correct for stimuli that are sufficiently separated in time or space, the BCI model better represents the current understanding of the integration process. Nevertheless, many studies refer to the MLE model.

Since audio-visual integration relies on the relative reliability of the sensory systems, both with respect to the weighting of the audio and visual cues and with respect to the size of the spatial and temporal integration windows, a shift in the relative reliability of the auditory and visual system would be expected to affect audio-visual localization. Thus, for hearing-impaired listeners, the shift in the relative stimulus weighting, as a result of the decrease in auditory reliability, is hypothesized to result in an increased reliance on visual information and an increased spatial integration window.

1.2 Overview of the thesis

In this thesis, the effect of visual information on auditory localization and localization behavior in normal and hearing-impaired listeners was investigated. By using a head-mounted display (HMD) in combination with a 64-loudspeaker array, a realistic and controllable visual and auditory environment was presented to the participants. With this, audio-visual integration was compared in normal-hearing versus hearing-impaired listeners using the spatial ventriloquist paradigm with realistic stimuli. Furthermore, experiments were conducted using more natural and realistic stimuli than typically employed in traditionally studies to test if audio-visual integration is also effective towards more real-life conditions. In the ventriloquist paradigm, auditory and visual stimuli are presented with a spatial disparity. If they are integrated, this results in a noticeable shift of the auditory stimulus towards the position of the visual stimulus. If the spatial distance between the stimuli is too large, they are not integrated and this shift does not occur. This paradigm can be used to investigate how spatial disparity affects integration, but since the distance over which stimuli are integrated and the strength of the shift are influenced by the participants' localization abilities, it has also been proposed as a way to investigate localization abilities. However, the ventriloquist paradigm can be susceptible to a response bias where participants respond to visual information instead of audio-visual information.

Chapter 2 presents background material on audio-visual integration and the factors that influence it.

In *Chapter 3*, a Gaussian clustering and categorization method to distinguish audio-visual responses (reflecting integration) from visual responses (reflecting a response bias) is investigated. The advantages and limitations of this method are assessed by applying it to data collected across a variety of conditions.

Chapter 4 investigates whether integration differs between more natural versus more synthetic stimuli by comparing the probability of integration of a noise burst and a light flash, i.e., synthetic stimuli versus the probability of integration of an impact sound made by a falling ball, i.e., natural stimuli.

The results of *Chapter 3* and *Chapter 4* suggested that investigating the effects of hearing loss and age on audio-visual integration would require the presentation of stimuli at locations between the positions of speakers in the loudspeakers array. However, previous studies using individual loudspeakers found an interaction between wearing an HMD and the perceived sound location. With ambisonics sound reproduction, each loudspeaker contributes to the perceived location of an auditory stimulus. Therefore, this interaction between the HMD and the sound reproduction might be different for ambisonics reproduced sound versus single loudspeaker playback. *Chapter 5* investigates whether wearing an HMD affects ambisonics sound source localization in normal-hearing listeners.

Based on results of *Chapter 3* and *Chapter 4,* a paradigm to study audio-visual integration in hearing-impaired was developed. In this paradigm, the distance between the auditory stimuli was varied, while the visual stimuli were presented at the same location. *Chapter 6* presents a study using this paradigm to investigate how age and hearing loss affect audio-visual integration. The focus here is on the spatial integration window, that is the distance over which auditory and visual stimuli are integrated despite a spatial disparity between the stimuli.

Chapter 7 uses a different task to investigate congruent spatial localization behavior in more realistic settings. In this task, participants search for a given target (either sound, an icon, or both) and the relative roles of auditory and visual information on this search are explored.

Finally, *Chapter 8* presents an overall discussion of the thesis, summarizing the main findings of the individual chapters and their implications, and provides an outlook of future work.

2

Audio-visual integration

Abstract

Much research has been done to explore multisensory integration and particularly audio-visual integration. Here, an overview of the current understanding of this integration process is discussed, focusing mostly on the parts of integration that are relevant for the rest of this thesis. First, the benefits of integration are discussed. Next, ‘optimal integration’ is presented along with the two most commonly used integration models, which inform about how stimuli are thought to be integrated. After this, the ventriloquist effect, which occurs as a result of this ‘optimal integration’, is discussed. This paradigm is used to not only study how integration occurs, but also what facilitates and limits integration. The most relevant limits and facilitators of integration for this thesis are discussed next. The chapter concludes with a summary of the most relevant background information.

2.1 Benefits of integration

People experience the world in a multisensory way and by combining information from the various sensory systems, more complete, precise, and more accurate knowledge of our surrounding can be obtained. This combination of information is referred to as multisensory integration and it has numerous benefits.

2.1.1 Faster reaction times

An important feature of multisensory integration is faster reaction times. The use of multiple sensory systems would, in general, predict faster times, since signals are processed separately and the fastest process will determine the reaction time. If there is some overlap in the distribution of the unimodal reaction times, then the average reaction time in the redundant presentation (audio-visual) will be faster than in unimodal conditions. This is referred as a race-model (Raab, 1962). There is, however, a limitation to the improvement in reaction time that is possible in this model since the fastest bimodal response could never be faster than the fastest unimodal response. However, in multisensory integration, bimodal reactions times have been found to be faster than what is predicted by the race model in audio-visual conditions (e.g., Diederich and Colonius, 2004; Gondan et al., 2005; Mahoney et al., 2011; Miller, 1982; Molholm et al., 2002; Schröger and Widmann, 1998; Yang et al., 2011), as well as conditions using a combination of other modalities (e.g., Diederich and Colonius, 2004; Forster et al., 2002; Mahoney et al., 2011; Murray et al., 2005). This is the result of signals from multiple sensory systems being processed together, i.e., co-activation (Miller, 1982); Signals from multiple sensory systems contribute to the same criterion which therefore reaches threshold faster than is possible if only one signal was contributing.

2.1.2 Accuracy and precision

By combining the information from the various sensory systems, the accuracy and precision of the fused percept can be improved. This has been shown in many sensory combinations. By combining visual and haptic information, texture perception (Heller, 1982) and object size and shape judgements (Ernst and Banks, 2002; Helbig and Ernst, 2007) can be improved. Ambiguity of speech sounds can be disambiguated by visual information (Plass et al., 2017). Audiovisual integration can also improve localization precision (e.g., Alais and Burr, 2004; Freeman et al., 2018; Odegaard et al., 2015) and Odegaard et al. (2015) showed that since both auditory and visual localization is consistently inaccurate (due to a bias towards the periphery in the case of auditory localization and a bias towards the center in the cases of visual localization) localization accuracy was also improved as a result of integration. Finally, distance estimation was found to be significantly more accurate and less variable when both audio and visual cues were provided (Anderson and Zahorik, 2014).

2.1.3 Stimulus enhancement

Another potential benefit of integration is the enhancement of stimuli where, due to multiple sensory signals contributing to one percept, detection of stimuli and stimuli events is enhanced (Andersen and Mamassian, 2008; Frassinetti et al., 2002; Lovelace et al., 2003; Noesselt et al., 2008). Also perceptually, this enhancement of stimuli has been shown to shift the perceived intensity of stimuli. For example, when noise is presented with light, the noise tends to be rated as louder (Odgaard et al., 2004) and vice versa (Stein et al., 1996). The lower the saliency of the most dominant stimuli, the larger the multimodal enhancement (Bernstein et al., 1973; Diederich and Colonius, 2004; Stein et al., 1996). This is referred to as the inverse effectiveness (Stein and Meredith, 1993). The origins of this effect, however, are not entirely clear with some studies suggesting that it is a result of a shift in the decision criterion, i.e., a response bias, rather than a lowering of the detection threshold (Lippert et al., 2007; Odgaard et al., 2003; Pápai and Soto-Faraco, 2017). However, the ventriloquist effect was found with stimuli below awareness level (Delong et al., 2018), suggesting that stimuli below awareness level can be integrated. Moreover, the enhancement of stimuli can be found in more complex settings; When participants were tasked with finding a specific visual target in a cluttered set of similar visual stimuli, Van der Burg et al. (2008) found that a synchronous auditory sound "pip" made the visual target stand out more. Participants were much faster in locating the target, due to the increased salience of the visual target.

2.2 Optimal integration

While the benefits of multisensory integration are mostly clear, there is still a lot of discussion as to how it is achieved. Since the sensory systems process information differently, some temporal, spatial or even informational misalignment between stimuli can occur from a multimodal source. Thus, it is not surprising that some misalignment is tolerated and does not prevent integration. However, when misaligned stimuli are integrated they are combined into one percept with one location, one point in time etc. How this misalignment between modalities is resolved is still under debate. Currently, the two main models of integration both assume that sensory information integrated in a statistically near-optimal manner, by weighting sensory information relative to their reliability.

A highly cited study by Alais and Burr (2004) demonstrated this weighting of sensory information by showing the inverse of what had regularly been established; While many studies before had shown that visual information could bias the perceived location of an auditory stimulus (e.g., Battaglia et al., 2003; Jack and Thurlow, 1973; Jackson, 1953; Klemm, 1909; Tastevin, 1937), Alais and Burr (2004) showed that auditory information could also bias visual information.

Alais and Burr (2004) did this by blurring the visual stimulus, thus lowering the relative localization reliability of the auditory and visual stimuli. When participants could localize the auditory stimulus more reliably (i.e., a lower auditory localization variance), the perceived location of a visual stimulus was shifted towards the position of the auditory stimulus. This integration model, referred to as maximum likelihood estimation (MLE) (Ernst and Banks, 2002), minimizes the variance of the combined percept (Alais and Burr, 2004).

However, one clear flaw of the MLE model is the forced fusion assumption. The model defines how stimuli are integrated in a statistically optimal fashion, however, it does not define when stimuli are integrated. As such, it implicitly assumes that any audio and visual stimuli, regardless of origin, are integrated, even though both top-down and bottom-up influences have been shown to influence the probability of integration (see section *Factors influencing integration* for a review). As such, Bayesian Causal Inference (BCI) was suggested to address this flawed forced fusion assumption using Bayesian statistics (Ernst and Bülthoff, 2004; Körding et al., 2007).

In the model by Körding et al. (2007), Bayes rule of conditional probability is used to assess, for a given auditory and visual signal, the likelihood that these signals originated from a common cause or that they originated from separate causes. Any characteristic of the signal could be used, e.g., the location, size, timing or weight of the stimuli, however the model was formulated using stimulus location. This assessment also includes a prior, which conveys how likely a person is to integrate information, thus allowing the model to adjust for top-down influences. Next, the auditory and visual signals are combined in the same reliability weighted manner as in the MLE model to calculate the audio-visual signal. Finally, depending on the decision-making strategy and the likelihood for a common versus separate cause, the audio, visual and audio-visual signals are combined in different manners. The statistically optimal decision rule would be the maximum-a-posteriori (MAP) solution, also known as model averaging, which again minimizes the response variance by weighting the integrated and non-integrated percepts relative to the likelihood of the causal structure (Körding et al., 2007). For example, if the task is to indicate the location of an auditory target, the perceived auditory and audio-visual location will be weighted, respectively, by the likelihood of the separate and common cause. If the audio and visual distributions are Gaussian, then so is the posterior distribution. In this case, the MAP represents the mean of the distribution. Moreover, if a uniform prior is assumed, then the BCI model with MAP decision rule is equal to the MLE model. For a more in depth description of their model, see Körding et al. (2007).

Wozny et al. (2010) expanded upon Körding et al. (2007), by formulating different decision rules than model averaging, namely probability matching and model selection. Model selection is a winner-takes-all decision, where the more likely causal structure solely determines the outcome. In probability matching, the decision rule is instead to take a random percept from the posterior distribution, which over many trials will then mimic the posterior distribution. Although this does not minimize the response variance, since it matches the variance of the posterior distribution, this was found to be the most commonly applied decision rule by Wozny et al. (2010). Rohe and Noppeney (2015), however, found evidence for model averaging.

Good fits with both models have been found (e.g., MLE: Ernst and Banks, 2002; Godfroy-Cooper et al., 2015; Heron et al., 2004; Moro et al., 2014, BCI: Acerbi et al., 2018; Magnotti et al., 2013; Roach et al., 2006; Winkel et al., 2017). However, there have been several studies where the data did not fit these models (e.g., Arnold et al., 2019; Battaglia et al., 2003; Meijer et al., 2019; Pick et al., 1969).

A potential explenation that has been posited is that many studies compare unimodal versus bimodal performance rather than compare the model predictions with probability summation, i.e., the benefit that would be expected based on presenting two non-integrated stimuli, rather than unimodal stimuli (Arnold et al., 2019). For a further overview of the current models, including MLE, BCI, race models etc., see Colonius and Diederich (2020).

2.3 The ventriloquist effect

To investigate multisensory integration, it is clearly necessary to present stimuli from multiple modalities. However, to do this, a wide range of experimental paradigms can be used. While some paradigms use congruent stimuli, i.e., where the stimuli from the different modalities convey the same information, many paradigms instead use incongruent stimuli to explore how information is integrated and when this integration process breaks down. One of such incongruent paradigms is the ventriloquist effect (Pick et al., 1969). The ventriloquist effect originally referred to the phenomenon where a puppet would appear to talk, but nowadays more generally refers to shifts in the perceived location (spatial) or timing (temporal) of auditory or visual stimuli as a result of integration. As discussed in a previous section, this shift in the perceived location of sound towards the position of the visual stimulus is the result of optimal integration (Alais and Burr, 2004). In many studies using this paradigm, participants are tasked with localizing the auditory stimulus, which is presented with a temporally congruent, but spatially disparate visual stimulus (e.g. Bosen et al., 2016; Delong et al., 2018; Jackson, 1953). Integration can then be studied by looking at how the visual bias, i.e., the shift in the auditory position, changes as a function of spatial disparity. The spatial ventriloquist effect can be tested both using relative (see *Chapter 6*) and absolute measures (see *Chapter 3* and *Chapter 4*) and the paradigm can be used to explore not only the spatial limits of integration, but also the relative stimulus weighting.

Subjective paradigms involving judgements of congruence (whether stimuli occured at the same location) have also been used (e.g., Godfroy et al., 2003; Lewald and Guski, 2003). However, since differences have been found between the subjective and objective measures (Bosen et al., 2016), the experiments in this thesis have focused on objective measures.

A phenomenon similar to the spatial ventriloquist effect exists in the temporal domain. While the spatial acuity of the visual system is better than that of the auditory system, in the temporal domain it is the auditory system that biases the visual system (van Opstal, 2016). The temporal shift of the visual stimulus can be investigated using relative (Morein-Zamir et al., 2003) or absolute measures (Fendrich and Corballis, 2001), the perceived rate of visual stimuli (Shipley, 1964) and even can result in a shift of the perceived number of visual stimuli (Shams et al., 2000, 2002). When the temporal disparity is sufficiently increased such that the stimuli are no longer integrated, these tasks becomes much easier and performance, e.g., as measured by the ability to judge which stimuli appeared first, improves.

2.4 Factors influencing integration

As mentioned earlier, optimal integration is influenced by the reliability of the information received from each modality. However, there are many other bottom-up and top-down processes and other factors that influence the integration process. Here, a brief review of the factors that are relevant to the experiments conducted are presented.

2.4.1 Temporal alignment

Temporal alignment between stimuli is probably the most important factor for integration and many studies have found no integration effects when stimuli are presented asynchronous (e.g., Bertelson and Aschersleben, 1998; Caclin et al., 2002; Frens et al., 1995; Jack and Thurlow, 1973). But, as illustrated by illusions, such as the temporal ventriloquist effect, integration can still occur despite some temporal misalignment. However, the probability of integration decreases as the temporal disparity increases (e.g., Kuling et al., 2013; Lewald and Guski, 2003; Slutsky and Recanzone, 2001). The range of temporal disparity where integration occurs more than half of the time is referred to as the temporal-integration-window (Colonius and Diederich, 2011). The temporal integration window is asymmetric, as the tolerance for visual leading asynchronies has been found to be much larger than audio leading asynchronies (Bhat et al., 2015; Slutsky and Recanzone, 2001; Wassenhove et al., 2007). Moreover, the range of the integration window is not fixed. Proficiency in temporal tasks such as music (Bidelman, 2016) or video-games (Di Luzio et al., 2021) are associated with a sharpened temporal window of integration. Additionally, with specific training an almost 40% narrowing of the temporal window was found (Powers et al., 2009). Training also reduced the susceptibility of older adults to the sound-induced flash illusion (Setti et al., 2014). However, this training has been found to be task specific (Powers et al., 2016). The temporal window of integration can also be widened by exposure to asynchronous stimuli (Navarra et al., 2005) and recalibrated to meet specific task demands (Mégevand et al., 2013).

Spatial alignment

As with temporal alignment, spatial alignment also facilitates integration. However, the influence of spatial alignment appears to be smaller than that of temporal alignment, as multisensory effects and illusions have been found to occur with spatially disparate stimulus (e.g., Bertelson et al., 1994; Fleming et al., 2020; Van der Burg et al., 2008). As with the temporal disparity, the probability of integration decreases as the spatial disparity increases and in the spatial domain too, a large variation in the width of the spatial integration is found (e.g., Bizley et al., 2012; Bolognini et al., 2005; Jackson, 1953). In general, though, more objective, non-biased, measures appear to result in smaller integration windows (about 4 to 10 degrees Bertelson and Aschersleben, 1998; Stawicki et al., 2019; Stenzel et al., 2019) than subjective or potentially biased measures (6-30 degrees, Godfroy et al., 2003; Jackson, 1953; Lewald and Guski, 2003). However, effects of integration have been found with spatial disparities much larger than suggested by the spatial integration window (e.g., Jackson, 1953; Montagne and Zhou, 2016). As with the temporal integration window, the size of the spatial integration window was found to be dependent on localization reliability, with better localization abilities resulting in a narrowing of the spatial integration window (Rohe and Noppeney, 2015).

2.4.2 Attention

The effect of attention on multisensory integration is still highly debated. Some studies have investigated effects of focusing attention on a specific modality and found conflicting results. DeLoss et al. (2013) found that focusing on the auditory modality increased the occurrence of the sound-induced flash illusion, whereas focusing attention on the visual modality decreased the occurrence. In contrast, other studies found that modality specific attention reduced integration regardless of the modality (Mozolic et al., 2008; Rohe and Noppeney, 2018). Again different results were found by Odegaard and Shams (2016), who found that although selective attention to a modality generally improved the sensory precision of that modality (visual in both temporal and spatial domain, auditory only in the temporal domain), integration tendencies were unaffected. Similar null-effects were found with haptic and auditory localization; Directing attention to the tactile modality did not increase the auditory shift (Caclin et al., 2002).

Other studies looked at spatial attention and generally found no effect. Comparing focusing on a central location versus focusing on a peripheral visual stimulus, no difference in the biasing effect of the peripheral visual stimulus on the perceived location of the auditory stimulus was found (Bertelson et al., 2000). When two competing visual stimuli at two different sides were presented, again, no effect of deliberate visual attention was found (Bertelson et al., 2000). However, when the size of these visual stimuli varied, the larger visual stimulus evoked a stronger bias (Bertelson et al., 2000). To further investigate these results, the authors ran an additional study on the effect of automatic visual attention by using an odd-one out visual stimulus and, again, found no effect of attention. Instead, while the smaller odd-one out visual stimulus successfully attracted attention, the larger visual stimuli evoked a stronger visual bias (Vroomen et al., 2001). In *Chapter 4* we further investigate effects of attention.

2.4.3 Semantic congruence

The idea of an effect of semantic congruence is that because natural stimuli co-occur, an association might be learned throughout life that binds these more naturally co-occurring stimuli more closely together than less natural stimuli that are regularly used in studies (Laurienti et al., 2004). For example, the color blue and the word blue or a dog and a barking noise are associated in a way that noise and blinking lights are not, therefore participants might be more likely to integrate these semantically congruent stimuli. Similar to the effects of attention, conflicting effects of semantic congruence have been found. Laurienti et al. (2004) found that participants could faster identify a color when both the auditory and visual stimuli matched in color, versus when they indicated different colors. Similarly, participants were less sensitive to temporal order, i.e., they were integrating more, when stimuli were semantically matched (Chuen and Schutz, 2016; Vatakis and Spence, 2007). Thomas and Shiffrar (2013) even found an improvement in visual sensitivity as a result of semantically congruent auditory stimuli, regardless of temporal alignment. Additionally, an fMRI study showed that object familiarity and semantic congruence affected the involved neural regions (Hein et al., 2007). Further, in the spatial ventriloquist effect large facilitative effects were found by some studies (Jackson, 1953; Thurlow and Jack, 1973; Warren et al., 1981).

On the other hand, other studies on the spatial ventriloquist effect found no effect of semantic congruence (Jack and Thurlow, 1973; Radeau and Bertelson, 1977) and some other multisensory interactions were found to be unaffected by semantic congruence (Koppen et al., 2008). Instead, the effect attributed to semantic congruence by some studies might be facilitated by other factors. For example, Steinweg and Mast (2017) posited that cautious response behavior in incongruent conditions, rather than semantic congruence in congruent conditions caused the difference in results. In *Chapter 4*, we discuss temporal correlation as another potential factor that can explain part of the effect of semantic congruence.

2.4.4 Age

Neural processing generally slows down and sensory systems decline with age (De Boer-Schellekens and Vroomen, 2014). As the multisensory benefit is strongest when the reliability is low (Stein and Meredith, 1993), it follows that the use of multisensory integration should increase with age, perhaps as a compensation strategy. Indeed, while unimodal performance is usually worse in older then in younger adults, multisensory benefit is higher in older adults; Saccadic reaction times to the onset of a visual target with and without an auditory stimulus was slower in all conditions in older adults, but the benefit of the auditory stimulus was larger in the older adults (Diederich et al., 2008). Multisensory gain was larger in older participants (Zou et al., 2017). Similarly, audio-visual speech gain was enhanced in older adults (Dias et al., 2021) and the multisensory benefit in the reaction times of older adults was significantly larger compared to the benefit that young adults experienced (Laurienti et al., 2006; Mahoney et al., 2011). This age-related enhancement of multisensory gain was still present when accounting for cognitive slowing (Peiffer et al., 2007).

In line with these expectations, several studies have shown indication of increased integration in older adults (e.g., DeLoss et al., 2013; Wang et al., 2018) and an increased susceptibility to multisensory illusions (Chan et al., 2021; Dobreva et al., 2012; Hernández et al., 2019; Hirst et al., 2019; Narinesingh et al., 2015; Setti et al., 2011). Additionally, Hirst et al. (2019) showed that this increased susceptibility is due to age-related effects that are independent of unimodal performance. Besides being more susceptible to audio-visual illusions, older adults also integrate over larger ranges of temporal and spatial disparity, i.e., they have increased integration windows (Bedard and Barnett-Cowan, 2016; De Boer-Schellekens and Vroomen, 2014; Diederich et al., 2008; McGovern et al., 2014). Additionally, adaptation to audio-visual temporal disparities was found to be reduced in older vs younger adults (Chan et al., 2014). Interestingly, while these objective measures of integration have found an effect of age, subjective measures such as perceptual synchrony judgements appear to be unaffected by age (Bedard and Barnett-Cowan, 2016; Scurry et al., 2020).

However, the results from several studies contrast with this general view of increased integration with age. Diederich et al. (2008) found that despite the increased benefit and the wider temporal integration window, the probability of integration was lower in older adults, compared to younger adults, due to slowing of the peripheral sensory processing. Similarly, Scurry et al. (2020) also found decreased integration in older adults and in Huyse et al. (2014), investigating audio-visual speech, no effect of age was found when the visual stimulus was clear. However, when the visual stimulus was degraded and the noise was stationary, the audio-visual gain was reduced in older adults, compared to younger adults (Huyse et al., 2014). Finally, no effect of age on the ventriloquist effect was found by Stawicki et al. (2019), but see *Chapter 7* for our findings on the effects of age on the ventriloquist effect.

2.4.5 Visual impairment

The effects of visual impairments on audio-visual integration are mostly in line with what is expected based on the integration models. Richards et al. (2018) found that unimodal and bimodal localization precision was degraded in visually impaired participants (amblyopia) compared to a control group. This decrease in bimodal localization precision was in line with MLE. Similarly, people with one eye, although slower in response times than normal seeing adults, integrated in line with MLE (Moro et al., 2014). Visual localization was improved in visually impaired patients, when spatially congruent stimuli were presented along with the visual stimuli (Frassinetti et al., 2005, 2002) and, again in line with expectations, a decrease of the McGurk effect was found in visually impaired adults and children (Moro and Steeves, 2018; Narinesingh et al., 2015; Wan et al., 2014). This decreased McGurk effect could reflect the decrease in visual reliability. Finally, in the temporal domain, Narinesingh et al. (2017) investigated the effect of the less dominant sensory system on the integration window and found an increased temporal window in people with visual impairments. This study served as one of the inspirations for the study in *Chapter 6*.

2.4.6 Hearing impairment

In line with expectations, in the spatial domain, listeners with acute moderate and chronic severe unilateral hearing loss showed a stronger visual bias on the impaired side (Venskytis et al., 2019) and cochlear implant users did not show the spatial multisensory benefit in response times (from spatially matched audio cues) that normal hearing listeners do (Pavani et al., 2017).

Surprisingly, in the temporal domain, where a hearing-impairment would be expected to have more influence compared to the spatial domain, no difference between normal and hearing-impaired listeners was found (Başkent and Bazo, 2012), even though the groups not only varied in hearing abilities, but also in age. However, this study used subjective judgements, which have been shown to differ significantly from the objective measures (Bosen et al., 2016; Van Eijk et al., 2008). Then again, a similar null-effect was found in adults with a cochlear implant. In a temporal order judgment task, no difference between the normal hearing and CI users was found and in the synchrony judgement the point of subjective simultaneity was even less-visual leading (Butera et al., 2018).

In contrast, Puschmann et al. (2014) found that hearing-impaired listeners were more distracted by cross-modal distractors. Finally, Altieri and Hudock (2014) found an accuracy-speed trade off in hearing impaired listeners, where some listeners would benefit from the increased accuracy of multisensory integration, but did not show the expected improvement in reaction-time, vice versa others would benefit from the increased reaction time but did not show the expected accuracy improvement.

2.5 Summary

Visual information has been shown to strongly influence auditory localization, especially through integration. This integration process is affected by numerous factors, including sensory reliability and, in connection to sensory reliability, sensory impairments. While in the case of visual impairments results are mainly consistent, in that decreasing visual reliability reduces the influence of visual information over auditory localization and increases susceptibility to auditory dominated illusions, in the case of auditory impairments, the results are more conflicting. In the temporal domain, no effect of hearing impairments were found. However, some studies with participants with a unilateral hearing loss and cochlear implant users showed an increased visual bias in the spatial domain. How mild-to-moderate symmetrical hearing loss affects spatial integration, however, has not been investigated. This and various other aspects of the spatial ventriloquist effect inspired the studies in the following chapters.

3

Using Gaussian mixture model clustering to analyze the ventriloquist effect[a]

Abstract

When auditory and visual stimuli are presented with a small spatial disparity they can be integrated, resulting in a shift of the perceived location of the auditory stimulus. This is generally referred to as the ventriloquist effect. The ventriloquist effect can be used to study how spatial disparity between the auditory and visual stimuli affect integration. However, the traditional paradigm, where participants indicate the perceived location of the auditory stimulus, can be confounded by a response bias, such that it can be difficult to separate audio-visual responses from visual responses. To remove the effect of response biases on the ventriloquist effect paradigm, we used Gaussian mixture model clustering and categorization method to group data and categorize them as either audio, visual or audio-visual. 16 normal hearing participants with (corrected to) normal vision performed the audio-visual localization task. Although the Gaussian mixture model method was able to reduce some of the response bias in some participants, it lacked consistency in both the clustering and categorization potentially resulting in an over-

---

[a] This chapter is based on Huisman et al., (2020).

estimation of the spatial disparity where the audio-visual stimuli are still integrated. Interestingly, the results showed an effect of the relative positioning of the auditory and visual stimuli, where integration was more likely when the visual stimulus occurred further outwards relative to the auditory stimulus. Since auditory stimuli are generally perceived further outwards and visual stimuli further inwards as compared to where they are presented, this could be explained by unimodal biases counteracting each other.

3.1 Introduction

The ventriloquist effect (Howard and Templeton, 1966) is a phenomenon where the perceived location of an auditory stimulus is shifted towards the position of a temporally coincident visual stimulus as a result of near optimal audio-visual integration. This integration is near optimal in the sense that it minimizes the localization variance of the integrated stimuli by weighting the auditory and visual information relative to the inverse of their localization variance (Alais and Burr, 2004; Heron et al., 2004). As the spatial acuity of the visual system is significantly better than that of the auditory system, the perceived location of an integrated, but spatially separate, visual and auditory stimulus is strongly biased towards the visual stimulus (e.g., Alais and Burr, 2004; Bertelson et al., 2000; Bosen et al., 2016; Hairston et al., 2003; Heron et al., 2004; Jackson, 1953; Thurlow and Jack, 1973; Vroomen et al., 2001). Some deviations from this optimal integration have been found. In particular, studies have found a visual bias which was stronger than predicted based on this weighted averaging (Arnold et al., 2019; Meijer et al., 2019). However, this increased visual bias could be accounted for in Meijer et al. (2019) by uncertainty about the causal structure with the Bayesian Causal inference model (Körding et al., 2007).

The ventriloquist effect has regularly been used to investigate how spatial disparity affects audio-visual integration (e.g., Bosen et al., 2016; Heron et al., 2004; Jackson, 1953; Körding et al., 2007; Thurlow and Jack, 1973). In this paradigm, spatially disparate, but temporally coincident, visual and auditory stimuli are presented to the participant, who is then tasked with localizing either the auditory or both stimuli (see Bruns (2019) for a review). As the distance between the visual and auditory stimulus increases, the probability of integration decreases and the shift in the auditory stimuli, i.e., the visual bias, decreases (e.g., Jackson, 1953; Lewald and Guski, 2003; Stenzel et al., 2019; Thurlow and Jack, 1973). The distance at which stimuli are integrated 50% of the times, is referred to as the spatial integration window (e.g., Meredith, 2002; Rohe and Noppeney, 2015). Based on the participants responses the visual bias can be calculated, which indicates how strongly participants are biased towards the visual stimulus. In addition, the integration window can be calculated which shows how far stimuli need to be apart for integration to break down.

However, this ventriloquist paradigm is particularly susceptible to a response bias (Chen and Vroomen, 2013; Radeau and Bertelson, 1977); participants are asked to indicate the perceived location of the auditory stimulus, which, if integrated, will be shifted strongly towards the location of the visual stimulus. However, this is not easily distinguishable from visual localization. For example, as Radeau and Bertelson (1977) noted about Jackson (1953)'s study which used a kettle and whistle, it is unclear if, despite a spatial disparity between the whistle and the steaming-kettle, participants perceived the whistle to come from the kettle, i.e., reflecting integration, or if they assumed the sound came from the steam-emitting kettle because of their previous encounters with steaming kettles, i.e., reflecting a response bias. Such a response bias can result in overestimating the visual bias and the spatial integration window.

Some adjustments to the setup have been proposed in previous studies to avoid such a response bias. For example, Vroomen and Stekelenburg (2014) suggested a two-alternative forced choice (2AFC) procedure where in one trial a static sound is presented and in the other a left-right alternating sound sequence. These audio stimuli are accompanied by visual left-right flashes. The task is to identify the trial with the alternating sound, which is simple if the stimuli are not integrated, but becomes much more difficult if the static sound is integrated with the alternating visual stimulus. Bertelson and Aschersleben (1998) and Stawicki et al. (2019) have used a similar approach with a left-right discrimination task with a staircase procedure (Levitt, 1971). In this setup the spatial disparity between two auditory stimuli is adjusted according to this staircase procedure. Temporally coincident visual stimuli are presented at the center, such that the apparent distance between the auditory stimuli is strongly reduced if the auditory and visual stimuli are integrated. This distance is then increased again until the stimuli are no longer integrated. In this manner a spatial integration window can be estimated. Finally, reaction times, either pooled (Freeman et al., 2018; Miller, 1986) or individual (Stenzel et al., 2019) can be used to confirm integration. For example, Stenzel et al. (2019) used the decrease in response time that occurs when stimuli are spatially congruent as opposed to incongruent, known as the Simon effect (Simon, 1990), to determine the limits of the ventriloquist effect.

Although these paradigms may avoid the response bias, they do result in the loss of some information. By shifting from absolute localization to an AFC task, the paradigms focus on determining the spatial integration window, but the strength of the visual bias is no longer calculated. In the case of reaction times, pooled data do not confirm integration for individual trials, so they are still susceptible to response bias trials. Here we suggest an alternative analysis for the original ventriloquist effect paradigm that could potentially address the response bias, while maintaining the ability to easily calculate a spatial integration window and determine the visual bias, namely Gaussian mixture model clustering (GMM clustering) (Bishop, 2006). GMM clustering is a probabilistic method that clusters data into Gaussian distributed clusters. The parameters of these clusters are generally determined by an expectation-maximization algorithm (Dempster et al., 1977). As localization behaviour is Gaussian distributed, these clusters should be able to describe the underlying (auditory, visual and audio-visual localization) behaviour well. Thus, we can cluster the data, calculate the slopes of the clusters, and categorize the clusters based on the slopes. If a cluster in the auditory localization error data (the difference between where participants perceived the audio and where the audio was presented) is identified that has a slope of approximately 0 (indicating no significant deviation away from auditory localization), we categorize this cluster as an audio-only cluster. If the cluster has a slope that matches with the position of the visual stimulus (a slope of 1), we can assume that this cluster contains the trials where the participant based their judgement solely on visual information. Clusters that deviate significantly from 0 and 1, then correspond to trials where the visual information significantly affected (but did not replace) auditory information, i.e., audio-visual localization. After categorizing the data, further analysis can be performed. For example, the visual bias can be calculated as the slope of the audio-visual cluster and the spatial integration window can be calculated as the range at which half of the responses are audio-visual.

To test this approach, 16 participants participated in the present study using ventriloquist paradigm, where the task was to indicate the perceived location of auditory stimuli, which were presented with spatially disparate visual stimuli.

3.2 Methods

3.2.1 Participants

7 females and 9 males (average age 29.5 ± 13 years) were recruited from the Technical University of Denmark (DTU). All participants reported normal hearing and normal vision. Audiograms confirmed normal thresholds (<20 dB HL) at octave frequencies between 125 Hz and 8 kHz for all participants and all participants scored a 0 or less on a LogMAR visual acuity chart. The experimental procedure was approved by the local Ethics Committee (De Videnskabsetiske Komitéer for Region Hovedstaden; H-16036391) and all participants provided informed consent. The participants were compensated with an hourly rate of 122 DKK.

3.2.2 Apparatus

The experiment took place in the Audio-Visual-Immersion-Lab (AVIL) at DTU. 5 loudspeakers (KEF LS50, KEF, Maidstone, UK) were used to present the auditory cues. These loudspeakers were positioned 2.4 m from the participant in an arc ranging from -30 to 30 degrees azimuth with 15 degrees separation. Participants were seated in a height adjustable chair at the center. This chair was adjusted vertically such that the participants' ears were aligned with the center of the loudspeakers.

The visual cues were presented using an HTC VIVE HMD (HTC Corporation, New Taipei City, Taiwan), which has a horizontal field of view of 100 degrees. Once seated, the HMD was placed on the participants head and the straps and lens distance was adjusted to the participants' preference. The virtual environment was a 1:1 model of the experimental room, created in UNITY3D (Unity Technologies, San Francisco, CA). To ensure spatial alignment between the real and virtual representation of the experimental room, the calibration method as described in Ahrens et al. (2019) was employed. This calibration method uses 3 HTC Vive trackers at known positions in the real world to detect when discrepancies occur and recalibrates the virtual world when either the discrepancy exceeds 2 cm or when tracking is lost on the HMD. In the virtual world, the loudspeaker array was replaced by a gray ring. This ring, which was 5 cm in height, indicated the elevation and distance, but not the exact azimuth, of the loudspeakers. A small white square with a 10-degree visual angle (VA) was placed just below the loudspeaker ring at 0 degrees azimuth. This was the focus point before and during the trials.

Between trials, a small sphere was visible at 2.4 m in front of the HMD at about eye height. This sphere followed the participants' head movements and was an aid for the participants for the visual alignment process, which occurred before every trial. At the start of a trial, this sphere disappeared and reappeared only after the trial was over.

To record their localization judgements, participants used a handheld HTC VIVE controller. In VR, a thin red rod was attached to the end of the controller such that it appeared to have a laser pointer. The participants pointed this "laser" at the location where they perceived the auditory stimuli and pressed a button to record their judgement.

3.2.3 Stimuli

The auditory stimulus was a 20 ms recording of the impact of a handball landing on a carpeted floor, presented at a peak-equivalent (pe) sound pressure level (SPL) of 65 dB. The visual stimuli consisted of an 8-degree VA ball. At the start of a trial, the ball appeared at location above the ring, fell for half a second, bounced once on the ring and then, 20 ms after bouncing, disappeared. Due to a miscorrected latency in the system, the onset of the auditory stimulus was presented, on average 105 ms after the point of collision in the visual stimulus. Some variability in the exact timing was present, because of the frame rate of the HMD and variability in the network connection between the computer. The standard deviation from 200 trials was found to be 15 ms.

3.2.4 Procedure

The experiment consisted of 4 blocks, presented in a fixed order. The experiment started with a unimodal audio condition, followed by a bimodal condition, and then a unimodal visual condition. The experiment ended with a simple pointing task, where participants were asked to point at stationary visual targets. This was used to estimate the motor error in pointing. In the unimodal conditions and in the pointing task, 2 additional loudspeakers, at ± 45, were included. These were not included in the bimodal conditions as these positions were near the limits of the field of view of the HMD.

The unimodal auditory condition consisted of 35 trials. In each trial, sound was presented, randomly, from one of the 7 loudspeakers. Trials at each position were repeated 5 times in this condition. The unimodal visual condition and the pointing task consisted of 21 trials. In each trial, the visual stimulus or the target was randomly presented at one of the 7 loudspeaker positions. These trials were repeated 3 times for each position. The audio-visual block consisted of 322 trials. For each of the 5 loudspeakers used to present the auditory stimuli, visual stimuli were presented at all 7 loudspeaker positions and in a range of 30 degrees centered around that loudspeaker position, using a step size of 3 degrees (see Fig. 3.1). The HMD provides a more limited field of view than normal vision. Thus, to ensure visibility, for the trials where sound was presented from one of the two most eccentric loudspeakers, the maximum eccentricity of the visual location was limited to ± 45 degrees. All combinations were repeated 3 times. In this condition, a maximum audio-visual separation of ±75 degrees was tested, with the densest sampling occurring in the ±30 degrees range of audio-visual separation.

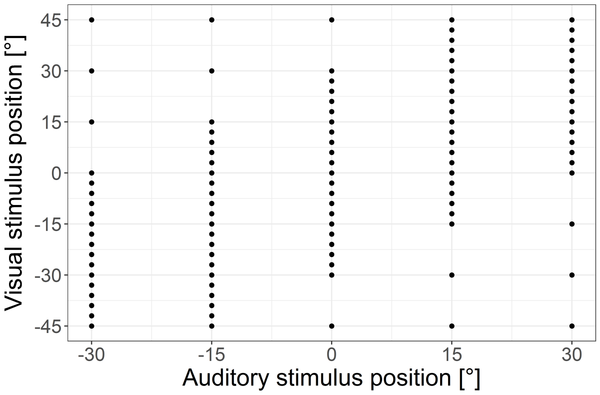


Figure 3.1: All the combinations of auditory and visual stimuli in the bimodal condition are shown. Each combination (indicated by a dot) was repeated 3 times. In later bimodal figures the results from the left hemisphere are mirrored, such that a negative disparity indicates that the visual stimulus occurred in the direction of the center, whereas a positive disparity indicates that the visual stimulus occurred further outwards. The densest sampling occurred from -30 to 15 degrees audio-visual disparity.

3.2.5 Trial

The participants were instructed to hold their head still and look at the focus point while the stimuli were presented and to only move their eyes and head after. To ensure that the participants' heads were facing forward during each trial, the participants had to align a small sphere, which tracked the HMD movement, with the focus point for 1 second. Once aligned for 1 second, the participants could press a button on the controller to start a trial. In the first two blocks (unimodal audio and bimodal stimuli), the participants were asked to point to where they heard the sound came from. In the third block (visual only), the participants were asked instead to point to where they saw the stimulus come from. In all cases, the location responses were restricted to the ring, such that the elevation and distance were fixed. The participants were, however, allowed to use the entire ring, allowing for front-back confusions.

As Pomper and Chait (2017) showed that eye movements can influence audio-visual integration, an additional task was used to ensure that, at the moment of collision, the participants were looking at the focus point straight ahead (rather than at the ball). As the ball collided with the ring, a letter appeared for 200 ms at the focus point. This letter was recognizable only when looking at the focus point. After performing the spatial localization task, a matrix of 16 different letters was presented and the participants were asked to select the letter that had appeared during the trial. If an incorrect letter was selected, the trial was considered invalid and repeated again at a later random position.

The final condition followed a different procedure. Here, the ring was replaced by a recreation of the loudspeaker array. On each trial, the participants were shown a number and were instructed to point at the center of the loudspeaker labelled with that number. As these loudspeakers were continuously visible, this condition was used solely to estimate the motor error in pointing.

3.2.6 Analysis

Trials where the validation was incorrect were removed from the analysis. On average, 11 ± 8 out of 378 trials had to be repeated (2.76%). The localization error was calculated per trial by subtracting the position of the auditory stimulus from the response. The spatial audio-visual disparity at which stimuli were presented was calculated by subtracting the position of the visual stimulus form the position of the auditory stimulus and trials where the auditory stimulus was presented at negative angles were mirrored, such that the disparity was positive when the visual stimuli occurred closer to the center than the auditory stimulus. Due to a logging error specific to the final condition (where very fast responses would be logged at the former rather than the current angle), results with an error over 15 degrees azimuth in this condition were considered invalid. These trials were removed from further analysis.

For the unimodal conditions, the average localization error and variance were calculated per angle. For each condition, at each non-zero localization, a paired t-test was run to examine biases towards centralization or externalization. Additionally, a Brown-Forsythe test between adjacent angles was run to examine the effect of eccentricity on the accuracy. For the bimodal data the clustering and categorization method was used to analyze trends in the data. The bimodal data was clustered per participant using a Gaussian mixture models clustering algorithm (MATLAB, 2017b). The algorithm was run with a maximum of 3 clusters to account for audio-only clusters, visual-only clusters and audio-visual clusters. The optimal number of clusters was then selected using the BIC criteria (Schwarz, 1978). The clustering algorithm was run 20 times, after which the most prevalent clustering result was used in the subsequent analyses.

A linear regression curve was fitted through each cluster. Each cluster was then categorized based on the slope of the regression curve. If pure auditory localization was used to localize the stimuli, the datapoints should (after correcting for auditory localization biases) cluster around 0 degrees localization error. Similarly, if pure visual localization is used, then after correcting for visual localization errors, a slope of 1 would be expected, matching exactly the position of the visual stimuli. However, if the slope of the regression line was significantly different from both 0 and 1, the responses in that cluster were categorized as AV responses. If the slope was not significantly different from 0, responses were categorized as auditory-only responses. Finally, if the slope was not significantly different from 1, then responses were categorized as visual-only responses. To be on the conservative side with the AV categorization, these comparisons were run both with and without a correction for the auditory and visual biases estimated from the responses in the unimodal conditions. Only if there was a significant deviation in both cases, a cluster was classified as AV.

After categorizing the clusters, all data were pooled and a multinomial logistic regression was run to test for temporal order effects, disparity effects and eccentricity effects, using the statistical software R (R Core Team, 2020). The predictors that were included in the analysis were the absolute disparity at the current trial, the absolute disparity at the previous trial, the temporal order (trial number) and the eccentricity at which the sound was presented. For this last factor, we ran the multinomial logistic regression again, but only with disparities from -30 to 15 degrees. This was the range of disparity that was presented at all locations. While larger disparities were presented at more eccentric angles, including these in the analyses might bias the results, hence they were excluded.

Finally, per angle, the probability of integration (i.e. the probability of being sorted into cluster AV in the present study) was calculated as a function of disparity and an averaged 50% spatial window was calculated.

3.3 Results

3.3.1 Unimodal

Fig. 3.2 shows the average localization error of the unimodal conditions. The auditory localization error (max. 13.0° ± 17.6) and the variance, indicated in blue, are greatly increased compared to the visual localization error (max. 4.5° ± 2.9), indicated in red. There is a trend that the accuracy of the visual localization decreases as the eccentricity increases (significant from -30 to 15 degrees azimuth, $p < 0.05$ Brown-Forsythe test between adjacent angles). Additionally, we found a centralized bias where visual stimuli are perceived more towards the center (t-test, $p<0.01$ for all non-zero locations). For auditory localization, instead a bias away from the center can be seen for most, but not all, locations (t-test, $p < 0.01$, for all but 15- and 30-degrees azimuth). The auditory localization bias did not vary significantly with eccentricity. At 0 degrees azimuth, an offset to the left was found for auditory localization (t-test, $p < 0.01$). As most of our participants were right-handed, this bias in contralateral direction could be expected. Lastly, the variance was smallest at 0 degrees azimuth, but did not increase with eccentricity.

The estimated motor pointing error was very small. Across participants, the largest pointing error was approximately a single degree (max. $0.97° \pm 0.4$). The precision of this pointing error did not change with eccentricity, but a small externalizing bias was found ($p < 0.05$ for all paired Welch tests between adjacent angles).

3.3.2 Bimodal

Fig. 3.3 shows the clustering results for four representative participants. As the task was to localize where the sound came from, one would expect that the localization error in Fig. 3.3 would be around 0 (with some deviation due to the localization bias). Indeed, this particular clustering around the auditory position appeared for all participants. However, most participants (13 out of 16) also showed additional clustering around the visual position. A large variability of the responses that occurred within a participant was found, when the same information is presented. Fig. 3.3 shows four examples of different clusters combinations. Participant 7 (top left) had full auditory localization (a single blue cluster). A small negative trend was found, but this was in line with the auditory bias found in the unimodal condition. Participant 15 (top right), had both auditory and audio-visual localization. The auditory localization

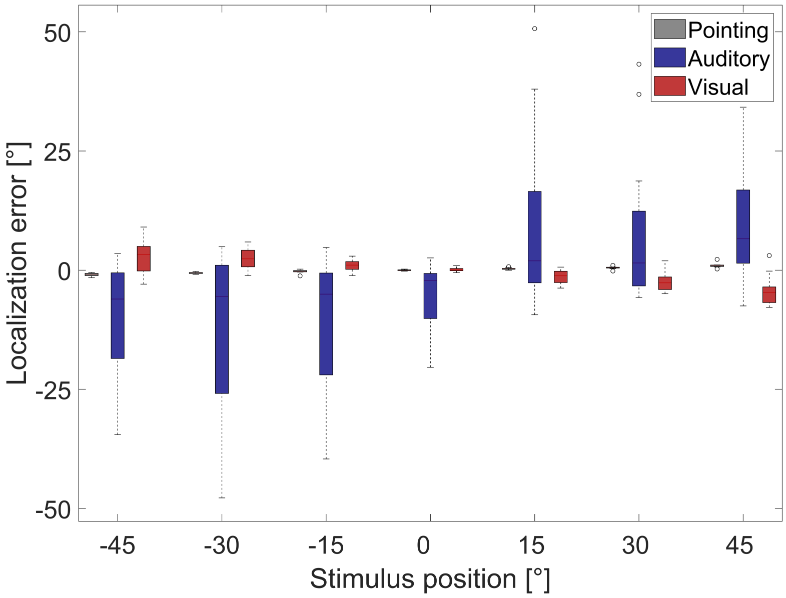


Figure 3.2: The localization error per modality for seven stimulus positions. A negative localization error indicates that the stimulus was perceived leftwards of the stimulus positions, whereas a positive localization error indicates that the stimulus was perceived to the right of the stimulus position. The box extends from the first to third quartile, with the line indicating the median. The whiskers then extend to 1.5 times the interquartile range. Outliers are indicated separately.

cluster extends across the entire range of spatial disparity, whereas the audio-visual cluster is much more dense at smaller disparities. Participant 13, in the bottom left, had mostly responses that were categorized as visual localization. Comparing between the data from participant 13 and 15, it can be seen that the audio-visual cluster found in participant 15's data deviated significantly from visual localization (indicated by the red line), whereas participant 13's data aligned perfectly with visual localization. Finally, the responses from participant 4 (bottom right) at -30 degrees disparity included both audio-only, visual-only as well as audio-visual responses.

The clustering and categorization results are summarized in table 3.1. The three categories, audio, visual and audio-visual describe the categorization of the clusters. Audio-visual clusters, which were defined as anything other than

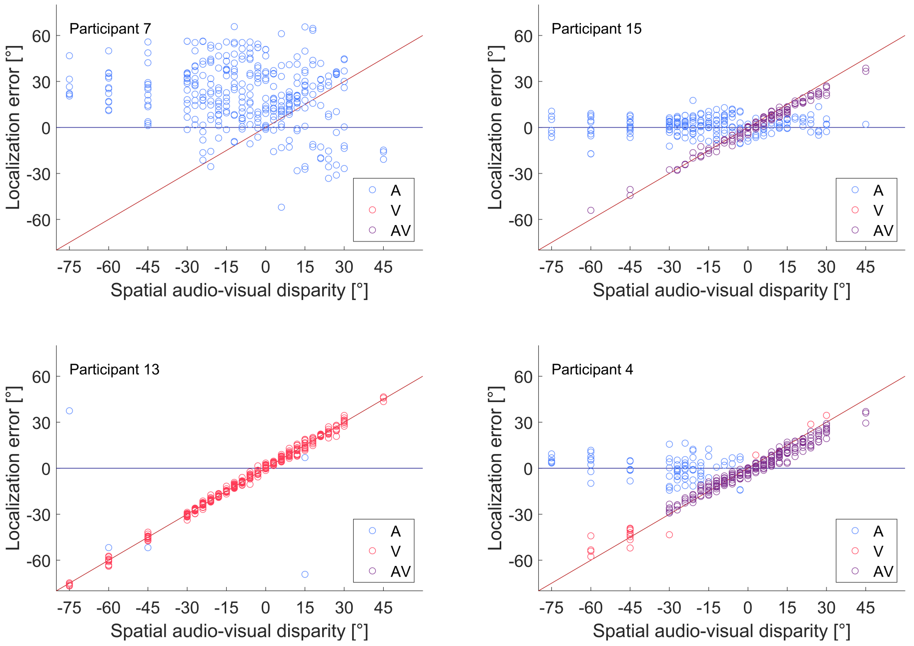


Figure 3.3: The panels show examples of clustering where audio (top left), audio and audio-visual (top right), audio and visual (bottom left) and, finally, audio, audio-visual and visual (bottom right). The color of each data point indicates the category of cluster it belongs to. Each data point corresponds to a single measurement, where the localization error is shown as a function of the disparity between the position of the auditory and visual stimuli. The results for the left hemisphere are mirrored such that a negative disparity indicates that the visual stimulus occurred closer to the midline. The solid blue and red line indicate the hypothetical trend for perfect auditory (blue) and visual (red) localization.

audio or visual, were sub-categorized based on how the visual bias, i.e., slope of the cluster, compared to the predicted visual bias, which was calculated based on the unimodal data. The average range of the clusters is shown in the last two columns.

Audio clusters were found in all but one participant and the data points that were categorized as audio-only, occurred over the entire tested range of tested spatial disparities. Some visual localization behavior was found in four participants. As with the audio-only clusters, the average minimum and maximum spatial disparity at which visual localization behavior was found was close to the minimum and maximum of the tested range. The audio-visual clusters were found in all of the participants, but the average minimum and maximum was, unlike the audio and visual clusters, more limited. The exception to this is the audio-visual clusters with a slope less than 1, which in both occurrences extended over the full range. Only two audio-visual clusters matched the predicted visual bias and most (11/16) were significantly smaller than predicted.

	Sub-categories	**Number of clusters**	**Predicted visual bias**	**Visual bias**	**Range**	
					Min	*Max*
Audio	*Total*	15	0	-0.02	-74.0	36.6
Visual	*Total*	4	1	0.96	-67.5	41.3
Audio-visual	*Total*	16	0.92	0.57	-48.2	39.4
	Larger than 1	1/16	0.99	1.37	-27.0	30.0
	As predicted	2/16	0.87	0.86	-42.0	36.0
	Smaller than predicted	11/16	0.92	0.57	-43.9	39.3
	Smaller than 0	2/16	0.98	-0.11	-75.0	45.0

Table 3.1: 35 clusters were categorized as audio, visual or AV based on the slope of the fitted linear regression curve. Audio clusters had a slope that was consistent with auditory localization, whereas visual clusters had slopes consistent with visual localization. All other clusters, where visual cues influenced, but did not dominate, auditory localization, were considered AV. The columns show, respectively, the number of clusters per category, the average predicted visual bias for these clusters, the average measured visual bias and the average range in degrees (min, max) over which these clusters occurred.

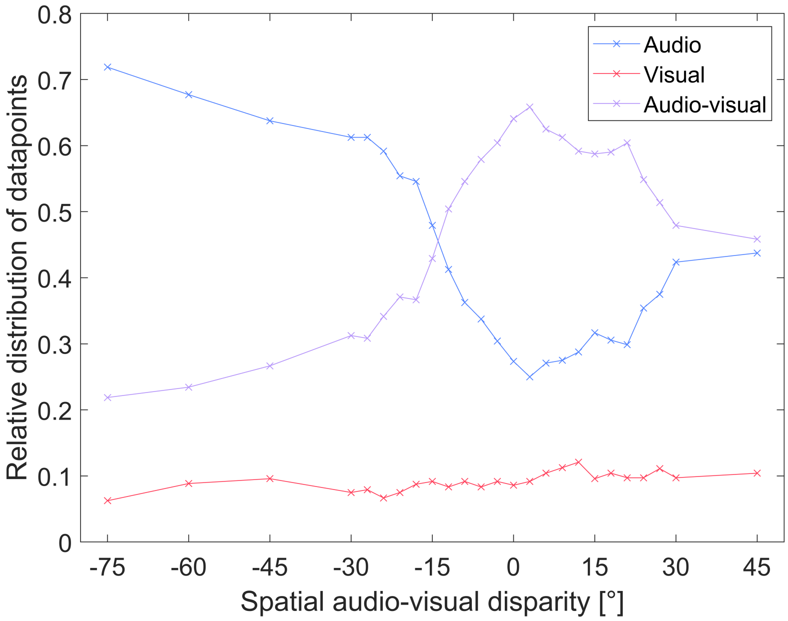


Figure 3.4: At all measured disparities (indicated with the cross marks) the proportion of datapoints categorized as audio (blue), visual (red) and audio-visual (purple) was calculated. The marks were connected to obtain the relative distribution of all combined data points into the clustering categories shown above.

Fig. 3.4 shows that the probability of data being categorized as audio (blue), visual (red) or audio-visual (purple) is strongly influenced by the spatial disparity of the audio-visual stimuli. Although audio-visual responses occurred over the entire range of disparities, a strong increase in audio-visual responses occurred from -15 to 30 degrees of disparity. As can be seen from the audio-visual curve, the highest probability of audio-visual responses did not occur when the stimuli were presented congruently, but for a spatial disparity of 3 degrees, i.e. when the visual stimulus was presented at 3 degrees more eccentric angles compared to the auditory stimulus. The function was also not symmetric; the probability of integration decreased more rapidly at negative disparities. The spatial window of integration (defined here as the spatial disparity where 50% of the responses were AV) was found to occur between -11.9 degrees for negative disparities and at 28.2 degrees for positive disparities.

Since results were mirrored, a negative spatial disparity indicates that the visual stimuli occurred inwards compared to the auditory stimulus, whereas a positive disparity indicates that the visual stimuli occurred further outwards. Visual localization is generally biased towards the center and auditory localization is generally biased away from the center (Odegaard et al., 2015). This was also shown in Fig. 3.2, where visual stimuli were generally perceived closer to the center, whereas auditory stimuli were generally perceived further away from the center. The asymmetry in the probability of integration could be due to these localization biases increasing the perceived disparity when stimuli are presented with a negative disparity and decreasing the perceived disparity when stimuli are presented with a positive disparity. To investigate if the asymmetry occurred due to inherent biases in the sensory systems we applied the same analysis as provided in Godfroy et al. (2003). For this analysis, we only included the data where the visual stimuli were presented at an exact speaker location (see also Fig. 3.5). With this analysis, we found the direction of disparity to be a significant predictor ($p < 0.05$) of integration.

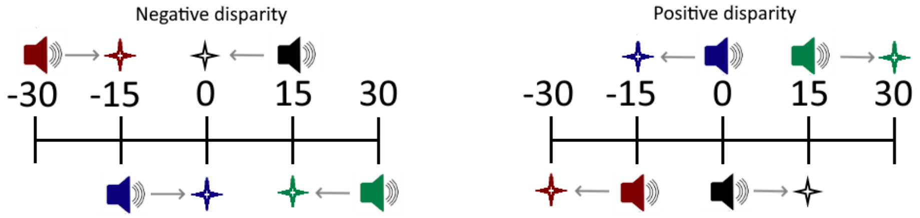


Figure 3.5: As in Godfroy et al. (2003), data were grouped depending on whether the visual stimulus occurred more centric (negative disparity, left panel) or more eccentric (positive disparity, right panel). To consider the effect of whether the visual stimulus occurred more towards centric or eccentric angles, we used only those combinations where the visual presentation angle overlapped with a loudspeaker position and where the disparity was -15 or 15 degrees. The data for the negative disparity group (left) represent the four combinations where the visual stimulus occurred 15 degrees towards the center and data for the positive disparity group (right) represent the four combinations where visual occurred 15 degrees outwards.

3.3.3 Predictors of integration

The multinomial logistic regression analysis indicated that there was a significant effect of temporal order ($p < 0.01$) and disparity ($p < 0.001$), but no effect of eccentricity or disparity at a previous trial. The effect of temporal order (less than 1% change in probability over all trials) was that audio-visual responses became more likely over time, while visual-only responses became slightly less likely. The audio-only categorized responses were not significantly affected by the temporal order.

The disparity between the auditory and the visual stimuli was the largest factor influencing integration ($p < 0.001$). An increase of 1 degree was associated with a 3.2% decrease in the probability of being categorized as an audio-visual cluster. However, there was no significant effect of disparity on the probability of visual-only responses.

Compared to when sound was presented from straight ahead, the probability of integration was significantly lower when sound was presented at ±30 degrees.

Separating left and right hemisphere responses, we found that this was mostly due to responses in the left hemisphere being significantly less likely to be AV responses ($p < 0.001$ at -30 and $p < 0.05$ at -15 degrees).

3.4 Discussion

To deal with response biases in the ventriloquist effect paradigm, we investigated using a Gaussian mixture model clustering and categorization analysis strategy to separate audio, visual and audio-visual responses. The clustering method was able to cluster and categorize data into different groups and differentiate between biased responses versus integrated audio-visual responses consistently in some of the individual participants. In the case of, for example, participant 13 (data shown in fig. 3.3) whose data matched visual localization over the entire tested range, data were categorized as visual localization and therefore did not influence the audio-visual results. However, there were some issues with using a clustering approach with regards to consistency on a group basis that limit the applicability of this analysis method for the ventriloquist effect.

Firstly, the clustering process is not deterministic, and, depending on the stability of the clusters, the resulting clusters can change every time the algorithm is run. To find the ‘best’ clustering, the clustering algorithm was run many times and the most prevalent clustering was used for the analysis. However, reproducibility was lost with this approach. Secondly, although we tried to take deviations due to inherent biases in localization into account, by testing for significant deviations both with and without corrections for these biases, some clusters with a small but statistically significant deviation from 0 were still classified as audio-visual although the influence of the visual cue led to only very small changes in localization compared to audio cues alone. Such potentially incorrect classifications affect the relative number of audio- versus audio-visual-clusters and therefore could have increased the size of spatial window we found. Indeed, only a small subsection of the audio-visual clusters (2/16) matched the predicted bias and most (11/16) had a smaller than predicted visual bias. However, the large variation in the audio-only results might have also led to an overestimation of the visual bias in several cases.

Most problematic for this analysis method is that, while the clusters were very stable for some participants, the results varied considerably for others. Such instability in the results can also occur with other analysis methods, such as Bayesian Causal Inference modelling (see for example Bosen et al. (2016)). However, with more simple methods, such as direct calculations of the visual bias, these issues do not occur. Especially problematic is that the unstable clusters mostly occurred in the places where the separation between audio-visual and visual localization data was unclear, as data points varied each run between being categorized as audio-visual versus visual. As shown in table 1, the predicted bias, based on the unimodal data, was on average 0.92, i.e., quite close to pure visual localization (reflecting a visual bias of 1). With a small variation in results, audio-visual and visual clusters can quickly overlap, reducing the effectiveness of the clustering approach. However, it is particularly in these areas that the method should be best. Nevertheless, as mentioned, the clustering method worked well in data where the audio-visual or visual behavior was consistent, such as for participant 13 and 15 (see Fig. 3.3).

In the bimodal condition, we observed some correspondence between the results we found here with this Gaussian clustering analysis and results found by previous studies. A similar pattern of integration was found, where integration occurred over large ranges of audio-visual spatial separation, but was most likely to occur at small ranges of spatial separation (e.g., Hairston et al., 2003;

Jack and Thurlow, 1973; Jackson, 1953; Slutsky and Recanzone, 2001). However, there were also clear deviations from previous studies. In the current study, we found the spatial window of integration to range from -11.9 to 28.2 degrees, whereas previous studies found spatial windows to be about 5-20 degrees absolute azimuth (Bertelson and Aschersleben, 1998; Godfroy et al., 2003; Lewald and Guski, 2003; Stenzel et al., 2019). For example, Godfroy et al. (2003) found a spatial window of perceived congruence of 6 degrees. The increased spatial window that was found here could be the results of miscategorized audio-visual clusters.

Another factor that warrants discussion is the 105 ms temporal delay, resulting from an incorrect correction for the communication delay between the audio system and the HMD. Studies on the optimal temporal window for audio-visual integration have found varying results. For example, Noel et al. (2018) found a temporal window of 150 ms (audio lagging), while Lewald and Guski (2003) found integration to occur with delays up to 200 ms. It is possible that for some participants, the temporal asynchrony disrupted integration. It is also possible that this also reduced integration even for participants with temporal integration windows that are larger than the asynchrony that was present. However, these results would only decrease the size of the integration window.

Finally, as we found that the optimal point of integration was shifted towards slightly positive disparities (just as in Godfroy et al. (2003)) and that the spatial window was asymmetric, we repeated their analysis using a subset of the data and found, in contrast to their study, a significant effect of the direction of disparity. This is unlikely to be due to the increase in the eccentricity of the visual stimulus. Although localization abilities have been found to limit the window of integration (Rohe and Noppeney, 2015), the increase in variance in visual localization that comes with the increase in eccentricity (e.g., Charbonneau et al., 2013; Freeman et al., 2018; Hairston et al., 2003), is associated with a decrease in the ventriloquist effect (Charbonneau et al., 2013; Hairston et al., 2003). Thus, the current data suggest that biases in the unimodal modalities might affect integration. At negative disparities, when the visual stimulus occurs more centrally, the biases in the modalities increase the perceived disparity. However, at positive disparities these biases would counteract each other and reduce the perceived disparity. Hence, when considering the perceived disparity at non-zero angles, the disparity is actually smallest at slightly positive disparities. This line of argumentation could also explain why the optimal point of integration did not occur when stimuli were congruent, but rather when visuals occurred more eccentric compared to the auditory stimuli. It should be noted, though, that the difference in responses observed between the left and right hemisphere, could affect these results as well.

3.5 Conclusion

To remove the effect of response biases on the ventriloquist effect paradigm, we investigated using a Gaussian mixture model clustering and categorization method to separate visual and audio-visual results. Although the method was capable of filtering out consistent visual responses, in cases where data could be explained by both visual and audio-visual localization behaviour, it lacked consistency. Additionally, despite the broad criteria used to categorize the clusters, some audio-visual were more consistent with audio-only localization behaviour. The inclusion of such clusters in the calculation of the spatial window, could result in an overestimation of the spatial window. As such, although some response bias was reduced, it is likely better to just calculate the visual bias or use different methods to avoid effects of response biases. Interestingly, in contrast to Godfroy et al. (2003), we found a significant effect of the direction of the spatial disparity, which could be explained by biases in the unimodal systems. However, it is difficult to disentangle the degree to which these differences are due localization biases rather than other factors. Follow up studies could investigate these potential effects further.

Acknowledgments

The model of the room and loudspeaker array used in the experiment was created by Kasper Duemose Lund. Audiograms were measured by audiologist Rikke Skovhøj Sørensen.

4

The ventriloquist effect is not consistently affected by stimulus realism[a]

Abstract

Despite more than 60 years of research, it has remained uncertain if and how realism affects the ventriloquist effect. Here, a sound localization experiment was run using spatially disparate audio-visual stimuli. The visual stimuli were presented using virtual reality, allowing for easy manipulation of the degree of realism of the stimuli. Starting from stimuli commonly used in ventriloquist experiments, i.e. a light flash and noise burst, a new factor was added or changed in each condition to investigate the effect of movement and realism without confounding the effects of an increased temporal correlation of the audio-visual stimuli. First, a distractor task was introduced to ensure that participants fixated their eye-gaze during the experiment. Next, movement was added to the visual stimuli while maintaining a similar temporal correlation between the stimuli. Finally, by changing the stimuli from the flash and noise stimuli to the visuals of a bouncing ball that made a matching impact sound, the effect of realism was assessed. No evidence for an effect of realism and movement of the stimuli was found, suggesting that, in simple

---

[a] This chapter is based on Huisman et al. (2021).

scenarios, the ventriloquist effect might not be affected by stimulus realism.

4.1 Introduction

In our everyday lives, our senses are continuously stimulated. While our sensory systems (e.g. auditory, visual or tactile) receive their input separately, it is known that the speed of processing through the neural pathways and the precision and accuracy of our sensory perception is enhanced through the integration of information across the sensory modalities (Diederich and Colonius, 2004; Freeman et al., 2018; Odegaard et al., 2015; Schröger and Widmann, 1998; Stein et al., 1989). As the brain cannot 'know' with certainty which sensory inputs belong together, since processing times and neural representations vary across modalities, it must estimate which sensory inputs should be integrated. This means that it is possible that sensory inputs from stimuli that originated from a different location (and potentially a different source) are integrated into a common percept. In such a situation, the location of the combined percept is determined through statistically optimal integration of, for example, an auditory and a visual percept (Alais and Burr, 2004). By weighting the auditory and visual percept relative to their reliability (i.e., the inverse of the localization variance), the variance of the combined percept is minimized. As the spatial resolution of the visual system is higher than that of the auditory system, the auditory percept is generally strongly biased towards the visual percept, although studies have also shown that the bias can be shifted towards the auditory percept by reducing the reliability of the visual percept (Alais and Burr, 2004). This effect, where spatially disparate audio-visual stimuli are integrated, resulting in a shift of the perceived location, is called the (spatial) ventriloquist effect (Howard and

Templeton, 1966).

Many studies have investigated aspects of audio-visual integration through this ventriloquist effect. However, most studies have used relatively simple stimuli, such as noise bursts and light flashes or white circles (e.g. Alais and Burr, 2004; Bosen et al., 2016; Wozny et al., 2010). While these studies provide insights into fundamental features of audio-visual integration, it is unclear to what extent the results obtained with these laboratory stimuli generalize towards real-world scenarios, as natural audio-visual stimuli share, besides temporal and spatial alignment, also contextual and semantic features, which are associated with those stimuli based on prior experience (Laurienti et al., 2004). As the shift in the perceived location of the auditory stimulus has been shown to arise relatively late in the neural processing (Bonath et al., 2007), top-down processes likely can influence the biasing effect of visual information. Indeed, top-down influences, like semantic congruence, attention and motivation, have recently been shown to be able to influence audio-visual integration (Bruns et al., 2014; Chuen and Schutz, 2016; Kramer et al., 2020; Laurienti et al., 2004; Taylor et al., 2006; Thomas and Shiffrar, 2013; Vatakis and Spence, 2007; Wassenhove et al., 2007; Zierul et al., 2019), see Bruns (2019) for an overview. However, not in all scenarios (Bertelson et al., 2000; Koppen et al., 2008; Radeau and Bertelson, 1977; Talsma et al., 2010; Thurlow and Jack, 1973; Vatakis and Spence, 2008; Vroomen et al., 2001). Specifically, in the case of the ventriloquist effect, the influence of stimulus realism remains unclear.

Jackson (1953) found that participants responded to (audio-) visual information over far larger ranges of spatial separation for realistic stimuli (kettle blowing steam with a whistling noise) than for artificially matched stimuli (light and bell). However, it is unclear whether the observed visual bias was due to audio-visual integration or due to a response bias (Vatakis and Spence, 2007, 2008), i.e., people might have adjusted their response to match their expectation as they assumed that audio and visual information would belong together, or their decision might have been based on the increased temporal correlation between the steam and the whistle (Chen and Spence, 2017). Similar effects of realism on the probability of audio-visual integration were found by Warren et al. (1981), who used synchronized and desynchronized audio-visual speech stimuli. While they attributed the difference in the visual bias found between these stimuli to the difference in realism, the temporal correlation could also have accounted for this effect. Thurlow and Jack (1973) investigated various facilitators of the ventriloquist effect. In various experiments, using both speech and non-speech signals, they found more audio-visual integration when the stimuli were more realistic. However, they did not differentiate between movement (i.e. facial movements of a puppet) and realism (facial features). Indeed, using a similar experimental setup, the same authors found a significant effect of the movement, but no effects of realism (Jack and Thurlow, 1973). However, again, the stimuli varied not only in movement and realism, but also with respect to their temporal correlation. Hence, it remains unclear if the increased realism of the moving stimuli or the increased temporal correlation between the visual and auditory stimuli was the facilitative factor.

Radeau and Bertelson (1977), using a voice in combination with a modulated light or an image of the speaker, found that audio-visual adaptation (an after effect of the ventriloquist effect) was unaffected by semantic congruence and was only due to the temporal synchronization and Parise et al. (2012) demonstrated that temporal correlation, rather than temporal alignment, facilitated integration. These results further support the hypothesis that effects of realism and movement were mainly driven by an increased temporal correlation of the auditory and visual stimuli. Thus, the ventriloquist effect could be dominated, in some scenarios, solely by 'low-level' factors, such as the spatial and temporal aligment and the temporal correlation.

With the recent rise of virtual reality, it has become easier to create and manipulate the realism of the stimuli. In the present study, a ventriloquist experiment was designed where the realism of the stimuli was varied stepwise to investigate the effect of realism, while maintaining a similar temporal correlation between the stimuli. Starting from the baseline condition using noise burst and light flashes, three factors were introduced: attention (through a distractor task), movement (through movement of the visual stimulus), and realism (through a change of the stimuli). To maintain the similar temporal correlation, movement was added only to the visual stimulus. The distractor task was introduced to ensure that participants were focused on the intended location, as Pomper and Chait (2017) showed that eye movements can influence audio-visual integration. As the ventriloquist effect has been shown to be unaffected by attention in similar conditions (Bertelson et al., 2000; Vroomen et al., 2001), no effect of attention was expected. However, based on most previous findings, increased stimulus realism was hypothesized to facilitate audio-visual integration over longer ranges of spatial disparity between the auditory and visual stimuli.

4.2 Methods

4.2.1 Participants

21 participants (11 female, 10 male; age 29 ± 10 years) were recruited from the Hearing Systems Section's volunteers' database and from the DTU student community for this experiment. All participants reported normal vision and normal hearing. This was confirmed with standard clinical tests. All partici-

pants had normal hearing thresholds at octave frequencies between 125 Hz and 8 kHz and all scored a visual acuity rating of at least 0 on a LogMAR visual acuity chart (Elliott, 2016). Data from participant 15 were excluded from the analysis based on extreme outliers in the unimodal visual and pointing conditions (20 datasets remained). The procedure was approved by the local ethical committee “Videnskabsetiske Komitéer for Region Hovedstaden” (H-16036391) and all participants provided written informed consent. The participants were compensated with an hourly rate of 122 DKK.

4.2.2 Apparatus

The experiment took place in the Audio-Visual-Immersion-Lab (AVIL) of the Technical University of Denmark. Auditory stimuli were presented using seven loudspeakers (KEF LS50, KEF, Maidstone, UK) that were part of a 64-loudspeaker array. The loudspeakers used were evenly positioned between ± 45 degrees azimuth at a distance of 2.4 m. In the center of the loudspeaker array was a height adjustable chair. The chair was adjusted such that the height of the participants' ears were aligned with the centers of the loudspeakers.

For the presentation of the visual stimuli, an HTC Vive HMD (Head Mounted Display; HTC Corporation) was used. This HMD was run with a separate computer, which was controlled by the computer that ran both the experiment and the loudspeaker array. A 1:1 model of the experimental room, created in UNITY3D (Unity Technologies), was used for the virtual environment. Calibration was done as in Ahrens et al. (2019), ensuring spatial alignment between the real and the virtual environment. For the calibration, three HTC Vive Trackers were placed at known positions and were tracked during the experiment. A shift of more than 1 cm in the position of one of the trackers, or the HMD losing tracking, resulted in a recalibration of the virtual world.

In the virtual environment, a virtual loudspeaker array was not included until the last part of the experiment. Instead, a gray ring (10 cm in height) was used to indicate the height of the loudspeaker array. At 0 degrees azimuth, just below this ring, a white square was placed that served as a focus point during the experiment (see Fig. 4.1). The virtual environment was continuously visible during the experiment and did not change, except in the last task where the ring was replaced by the loudspeaker array. Between trials, a small sphere was used to help participants with the visual alignment process. This sphere was positioned at about eye height at a distance of 2.4 meters straight ahead of the participants and moved synchronously with their head movements. At the start of each trial, this sphere disappeared. Only after the trial was finished did it reappear. To proceed through the experiment and record their localization judgements, the participants used a handheld HTC VIVE controller. In the virtual environment, a thin red rod was attached to simulate a laser pointer to help the participants point towards the perceived auditory stimuli. This “laser” disappeared at the start of a new trial and reappeared when a response from the participant was requested.

4.2.3 Stimuli

Three different visual stimuli and two different auditory stimuli were used in this experiment. However, not all combinations were tested. The experiment was designed such that each bimodal condition, containing a single set of stimuli, added one new factor. The baseline condition represented the commonly used laboratory conditions in audio-visual experiments, i.e. flashes and noise bursts. For these baseline stimuli, the magnitude spectrum of the realistic sound was combined with a randomized phase to obtain a noise with the same loudness as that of the original recordings of the real handball impact stimulus. The visual baseline stimulus was a 20-ms light blur that appeared synchronously with the auditory stimulus above the loudspeaker ring. The light blur was 33.56 cm in diameter, corresponding to an 8-degree visual angle, as indicated in Fig. 4.1. The Gaussian blur had a standard deviation of approximately 5.5 cm (standard Gaussian blur scaled to the size of the visual stimulus).

The second bimodal set consisted of the same baseline stimuli, but it introduced a distractor task where a letter was shown on the white screen in the center at the same time as the other stimuli, as indicated in the middle panel of Fig. 4.2 for the flash and distractor stimulus. The purpose of the distractor task was to ensure that participants were fixating straight ahead during each trial (which is particularly important for later conditions involving moving visual stimuli). As no effect of attention was expected, this condition was included as a control. The letter remained visible for only 200 ms. After the participants had finished the localization task of the auditory stimuli, they were shown a matrix of 16 letters and had to select the letter that had appeared during the collision. If the participant was incorrect, the trial was repeated at a later time chosen at random. This process was repeated if the participant continued to indicate an incorrect letter.

The third set of stimuli again used the baseline stimuli (with the distractor task) but introduced movement. The visual stimulus appeared above the ring at the start of the trial, fell for half a second, bounced once on the ring and then disappeard 20 ms after bouncing. The bouncing on the ring was the trigger for the audio stimulus.

The realistic stimuli consisted of the sound and visuals of a dropping ball. The auditory stimulus was a 20-ms recording of the impact of a handball landing on a carpeted floor, presented at a peak equivalent (pe) sound pressure level (SPL) of 65 dB. As illustrated in the right panel of Fig. 4.1, the visual stimulus was a blue ball, in the same size as the flash stimulus (33.56 cm, or 8-degrees visual angle in diameter). As with the moving flash stimuli, the ball appeared above the ring, fell down, bounced once on the ring, triggering the auditory stimulus, and disappeared.

Due to a miscorrected latency in the system, the audio was played, on average, 105 ms after the visual stimulus. There was a variation of ±13 ms due to the frame rate of the HMD and a variation in the communication speed between the computers running the virtual environment and the audio system. This asynchrony was the same across all conditions. Furthermore, as the visual stimuli appeared slightly above the ring, there was a slight elevation difference of 3 degrees between the auditory and the center of the visual stimulus. However, due to the low sensitivity to incongruities in elevation (Godfroy et al., 2003), this should not affect the integration.

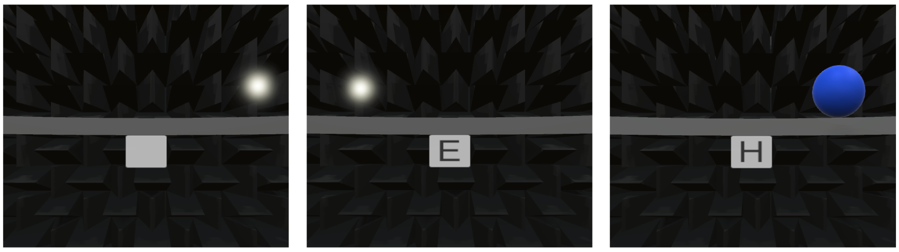


Figure 4.1: The different stimulus conditions and the experimental setup. The ring is visible in gray with the focus point below it. The stimuli shown here, from the left to the right, represent the baseline visual stimulus, the attention visual stimulus and the congruence stimulus. The middle and the right panels also also illustrate the distractor task stimulus.

4.2.4 Conditions

The main task of the experiment consisted of a localization task, using only auditory or only visual (i.e. unimodal) stimulation, or a combination of both (i.e. bimodal stimulation). In total, the experiment consisted of ten conditions which were divided into four blocks (see table 4.1). The block order was fixed:

unimodal audio, bimodal, unimodal visual, pointing. However, within each block, the conditions were presented in a counterbalanced manner across participants.

The first two conditions were unimodal audio-conditions. Here, the sounds were presented randomly from one of the seven loudspeakers and each position was repeated five times resulting in 35 trials each. As the HMD has a limited field of view (110 degrees), the visual stimuli were limited to a maximum eccentricity of ± 45 degrees. Because of this, the two outer loudspeakers were not used in the bimodal conditions. For each of the five loudspeakers used to present sound in the bimodal conditions, visual stimuli were presented in a 30-degree range around that loudspeaker in 3-degree steps, and also at the other six loudspeaker locations. As a result, the densest sampling occurred between 15 and -30 degrees audio-visual disparity, and the maximum disparity was up to ± 75 degrees. The sampling is also shown in Fig. 4.2, which shows for each auditory stimulus position, all tested visual positions. Each combination was presented three times, leading to a total of 322 trials per bimodal condition.

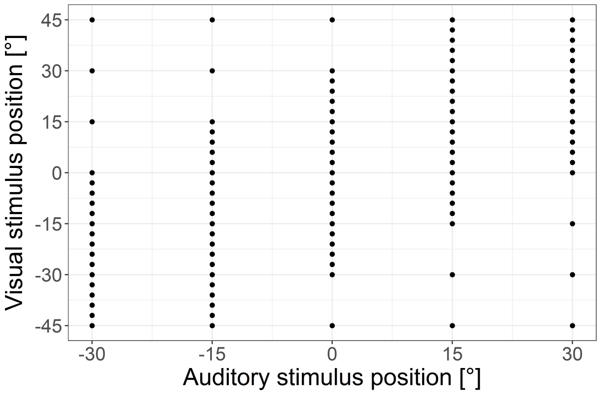


Figure 4.2: The left panel shows all unique combinations of auditory (abscissa) and visual (ordinate) positions in the bimodal conditions. Each combination (indicated by a dot) was repeated three times.

The next block consisted of the three unimodal visual conditions. At this point in the experiment, the task changed from localizing sound to localizing visual stimuli. This condition included positions at all seven loudspeakers and each position was tested three times, resulting in 21 trials per condition.

To account for potential biases in the pointing response (Ahrens et al., 2019), a 'pointing' condition was included where the participants had to point at a continuously present static visual stimulus. In this task, no distractor ring was used and the gray ring was replaced by a model of the loudspeaker array. Participants were then shown a number and had to point with the 'laserpointer' at the center of the loudspeaker with that number. This was the final task of the experiment. As in the unimodal visual conditions, this task used all seven loudspeaker positions and three repetitions were conducted, resulting in 21 trials. The conditions and stimuli are summarized in table 4.1.

Block	Stimuli	Distractor task	Movement	Realism
1. Audio	Noise burst	Yes	No	No
	Ball impact sound	Yes	No	Yes
2. Audio-visual	Baseline (noise + flash)	No	No	No
	Attention (noise + flash)	Yes	No	No
	Moving (noise + moving flash)	Yes	Yes	No
	Realism (ball + moving ball)	Yes	Yes	Yes
3. Visual	Flash	Yes	No	No
	Moving flash	Yes	Yes	No
	Moving ball	Yes	Yes	Yes
4. "Pointing"	Loudspeakers target	No	No	Yes

Table 4.1: The experiment consisted of ten conditions presented in four blocks. Blocks were presented in a fixed order, but within a block, conditions were counterbalanced across participants. Each bimodal condition added a new factor.

4.2.5 Analysis

Data were analyzed using the statistical software R (R Core Team, 2020). The unimodal data were analyzed using Levene's test to evaluate differences in the variance and an ANOVA was applied to the localization data. To analyze the bimodal results, the localization error was calculated per participant, condition, auditory stimulus position and angle by subtracting the position of the auditory stimulus from the response. This localization error was corrected by subtracting the mean localization error in the congruent trials at each loudspeaker location to account for angle-dependant localization biases. This corrected error was then divided by the spatial audio-visual disparity to calculate the visual bias. The spatial audio-visual disparity itself was calculated as the position of the visual stimulus minus the position of the auditory stimulus, with positive disparities indicating that the visual stimulus occurred more to the right compared to the auditory stimulus. An ANOVA analysis compared the visual bias with the absolute spatial audio-visual disparity, condition, absolute auditory stimulus position and the relative stimuli positioning as potential predictors. The relative stimuli positioning refers here to if the visual stimulus occurred outwards compared to the auditory stimulus, or if it occurred closer to the center. To investigate how the various factors affected the results, a Bonferroni-corrected post-hoc within factor comparison analysis was used. For this analysis the disparities larger than 30 degrees were not included (i.e., only the densely tested area was included), as the initial analysis revealed interactions which could not be explored when these data points were included. Dropping these specific points from the analysis did not affect the results.

4.3 Results

4.3.1 Pointing bias

Fig. 4.3 shows the localization error as a function of the stimulus position when pointing at a continuously present visual target. This task was included to measure the error in pointing. As a 'laser-pointer' was included, the accuracy and precision of pointing is very high. As can be seen in Fig. 4.3, the maximum me-

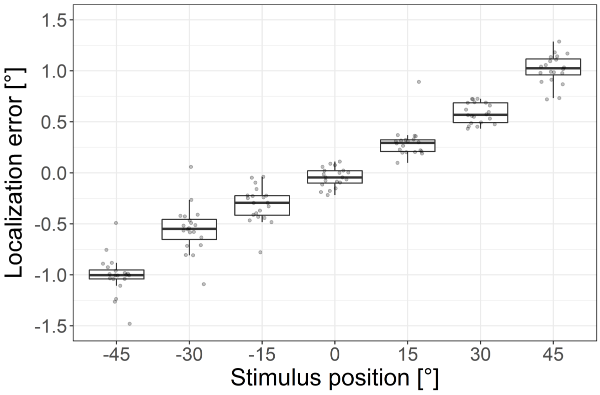


Figure 4.3: The median error in pointing per participant and angle. Each datapoint corresponds to the median error for one participant at a stimulus position. The boxes extend from the first to the third quartile, the line shows the median perceived response. Outliers are indicated separately.

dian localization error was around one degree and the variance of the error was below one degree. There was a small dependency, i.e., bias, of the localization on the stimulus position, with slightly increased errors at higher eccentricities (± 1 degree at ± 45 degrees azimuth, $[F_{1,6} = 3.234, p < 0.01]$). The variance did not vary significantly with stimulus position $[F_{1,6} = 3.234, p = 0.631]$.

4.3.2 Unimodal conditions

The unimodal conditions were used to test if there were differences in localization between the stimuli that were used. Fig. 4.4shows the localization error for the auditory (left panel) and the visual stimuli (right panel) as a function of the presentation angle. For the auditory stimuli, the baseline stimulus (noise burst) is indicated in light blue and the congruent stimulus (ball impact audio) is indicated in dark blue. Levene's tests showed that, for the auditory stimuli, the variance varied significantly only with angle $[F_{1,6} = 2.662, p < 0.05]$, but not with condition $[F_{1,6} = 1.341, p = 0.2470]$. Similarly, the localization error also varied with angle $[F_{1,6} = 51.035, p < 0.001]$ and not with condition $[\mathrm{F}_{1,1} = 0.251, p = 0.6162]$. However, an interaction between the stimulus angle and condition was found $[F_{1,6} = 2.645, p < 0.05]$.

The right panel of Fig. 4.4 shows the localization data for the visual stimuli, with the static flash data shown in light red, moving flash data shown in red and the ball stimulus data shown in dark red. For the visual stimuli, the localization error and variance were much smaller than for the auditory stimuli. Moreover, a clear trend can be seen in the visual responses, where responses were closer to the center (positive errors at negative angles and vise versa) when the stimulus was presented more laterally. Additionally, the variance also increased with presentation angle. This was confirmed by the statistical analysis, which revealed an effect of stimulus position on both the variance $[F_{1,6} = 4.6560, p < 0.001]$ and the localization error $[F_{1,6} = 55.935, p < 0.0001]$, but no effect of the stimulus used on either the localization error $[F_{1,2} = 0.305, p = 0.7375]$ or the variance $[F_{1,2} = 0.721, p = 0.1520]$ respectively.

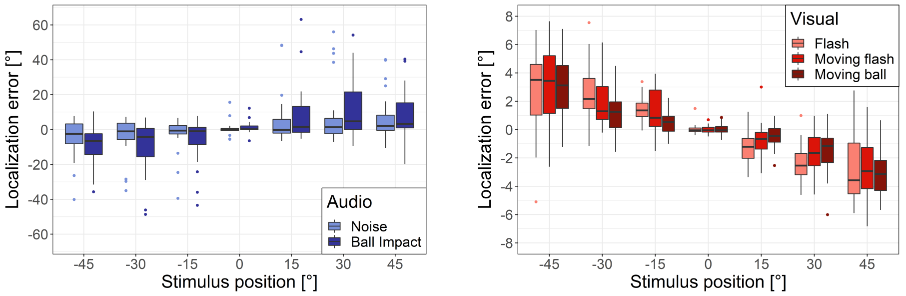


Figure 4.4: The left panel shows the localization error for the two different auditory stimuli. The right panel shows the localization error for the three different visual stimuli. The boxplot extends from the first to the third quartile, with the median across all participants shown in black. The whiskers extend to 1.5 times the interquartile range. Outiers beyond this range are indicated separately. Note the different ordinate scales for each figure.

4.3.3 Bimodal condition

Fig. 4.5 shows the localization error as a function of the spatial disparity between the auditory and the visual stimuli in the four bimodal conditions for participant 4. A bias, where responses are shifted towards the position of the visual stimulus,

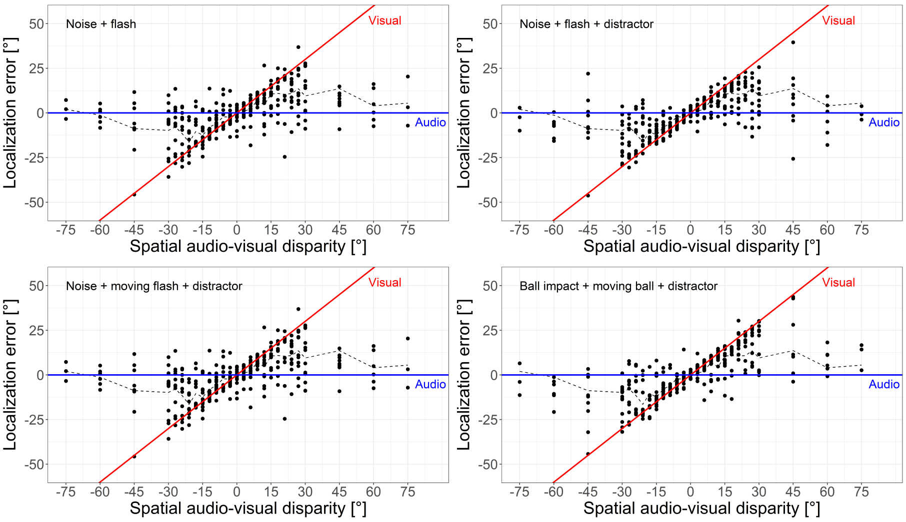


Figure 4.5: Bimodal responses of a representative participant, participant 4. The four panels show the results for each of the bimodal conditions. The conditions are indicated in the top left corner of the panel. Perfect visual localization is indicated by a red line and perfect auditory localization is indicated with a blue line. The dashed curve (black) shows the mean response as a function of audio-visual disparity. When the spatial disparity was small, participant 4 showed a visual bias on most trials i.e., most responses shifted away from auditory localization towards the visual localization line. No clear difference in either the range or the strength of the visual bias was found between the four conditions.

can be seen in all conditions. However, comparing the average (dashed line) responses, no clear effect of condition is visible for this participant.

The visual bias, averaged across participants, per condition is shown in the Fig. 4.6. The left and right panels show the relative stimuli position. In the left panels, A-V-A, the visual stimulus occurred inwards relative to the auditory stimuli, whereas in the right panels it is instead the auditory stimulus that occurs inwards relative to the visual stimuli. The upper panels shows results for when the auditory stimulus was positioned at 0 degrees azimuth, the middle panels show the results for when the auditory stimulus was presented at ±15 degrees and the bottom panels show the results for when auditory stimuli were presented at ±30 degrees. In the upper panels the auditory stimulus is presented at 0 degrees, as such the left panel show reponses where the visual stimulus was presented in the left hemisphere and the right panel shows responses for visual

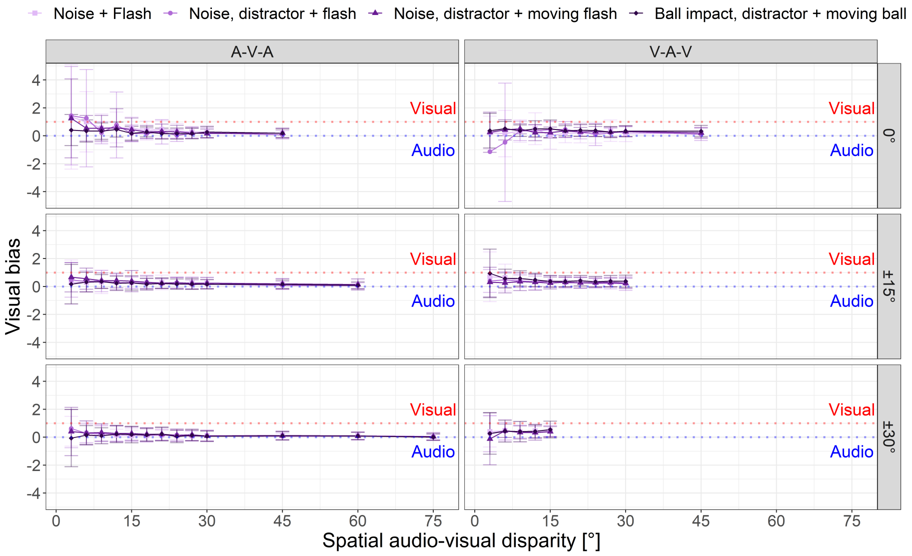


Figure 4.6: Average visual bias as a function of spatial audio-visual disparity per condition. The left and right panels show the relative positioning of the stimuli. The A-V-A panels show when the visual stimulus occurred left (0 degrees) or more towards the center (±15 and ±30 degrees) compared to the auditory stimulus. For example, for 30 degrees disparity, visual stimuli occurred towards the left in this setup, whereas for -30 degrees the stimuli occurred towards the right. The V-A-V panels show the visual bias for when the stimuli occurred right (0 degrees) or further outwards (±15 and ±30 degrees) compared to the auditory stimulus. The horizontal panels show the results per (absolute) auditory stimulus position. The red dotted line indicates a complete visual bias, where localization responses are completely shifted towards the visual stimulus, whereas the blue dotted line indicates pure auditory localization, without any visual bias. The various conditions are indicated by different shapes and purple shades. Due to limitations of the field of view of the HMD, not all disparities could be tested for all angles, hence the difference in start and end points. For one point in the topright panel the standard deviation is not included as the difference in conditions cannot be assessed on the required scale. This point is the noise + flash with distractor condition at 3 degrees spatial disparity (-1.14 ± 7.90).

stimuli presented in the right hemisphere. As no visual bias can be calculated at 0 degrees spatial disparity, where the auditory and visual position overlap, the curve is interrupted at this position. Since the visual bias is calculated by dividing the localization error by the spatial audio-visual disparity, similar errors in localization cause much larger changes in the bias at small disparities. This results in a steep increase of the standard deviation as the disparity decreases.

As visible in Fig. 4.6, the visual bias was found to decrease, in most cases, significantly with increasing absolute spatial audio-visual disparity [$F_{9,6980} =$

$5.513, p < 0.0001$] and varied depending on both the relative positioning of the stimuli [$F_{1,6980} = 5.034, p < 0.05$] and the absolute position of the auditory stimulus [$F_{2,6980} = 10.317, p < 0.0001$]. However, significant interactions between these factors were found, namely an interaction between the effect of the relative and auditory stimulus positioning [$F_{2,6980} = 18.373, p < 0.0001$], an interaction between the spatial disparity and the relative stimulus positioning [$F_{9,6980} = 2.187, p < 0.05$] and a three-way interaction [$F_{13,6980} = 4.427, p < 0.0001$]. In the A-V-A stimulus setup, at small disparities (<15 degrees) the visual bias was larger when the auditory stimulus was presented at 0 degrees compared to ±15 and ±30 degrees [3 degrees, 0-15 : $t_{6980} = 5.657, p < 0.0001$; 3 degrees, 0-30: $t_{6980} = 6.105, p < 0.0001$]. On the contrary, in the V-A-V setup, the visual bias was lower when the auditory stimuli were presented at 0 degrees [0-15 : $t_{6980} = -5.278, p < 0.0001$].

The results for the various conditions (see Fig. 4.6, upper panels) were very similar at larger audio-visual disparities. However, when the stimuli were close together, the visual bias increased and some differences appeared between the conditions. Although no main effect of condition was found [$F_{3,6980} = 0.4372, p = 0.4372$], there was a significant interaction between the relative stimuli positioning and the condition [$F_{3,6980} = 10.108, p < 0.001$] and a three-way interaction between the auditory stimulus position, the relative stimuli positioning and the conditions [$F_{13,6980} = 4.427, p < 0.001$]. As can be seen in Fig. 4.6, upper left panel, when the visual stimulus occurred in the left hemisphere with the auditory stimulus at 0 degrees azimuth (denoted, A-V-A, but since the auditory stimulus occurred at the center, this corresponds to stimulus occuring left), the realistic stimuli produced a significantly smaller visual bias than the noise and flash (baseline condition) stimuli [$t_{6980} = 3.354, p < 0.01$]. The other combinations did not reach statistical significance.

In contrast when the visual stimulus was presented in the V-A-V setup, these realistic stimuli evoked a much more similar visual bias and it was instead the second set of stimuli (noise and flash with a distractor) that produced a lower visual bias. Both at 0 [$t_{6980} = -3.372, p < 0.01$] and ±15 degrees [$t_{6980} = 3.034, p < 0.05$] the difference between the second and fourth (realistic stimuli) condition was significant. Curiously, a negative visual bias can be seen for the noise and flash with distractor stimuli in the upper right panel of Fig. 4.6 indicating that participants perceived the auditory stimulus to be futher away from the visual stimulus. Again, the stepwise comparison between the first and second, second and third and third and fourth condition was not significant.

To see how if introducing the additional factor (attention, movement, realism) improved the model, equality constraints were tested using a Bayes factor test [38]. The fully unconstrained model performed worse than the combined noise and flash with and without distractor model, indicating that this factor indeed did not improve the model ($BF = 2.3501e^{-11}$). Similary, the fully unconstrained model performed worse than the model with the combined noise and flash with distractor and moving noise and flash with distractor stimuli ($BF = 9.6465e^{-7}$) and the combined moving noise and flash with distractor and the realistic stimuli ($BF = 1.7162e^{-7}$).

4.4 Discussion

The present study investigated if movement and realism of the stimuli influence the spatial ventriloquist effect. Starting from stimuli that are commonly used in experiments (noise burst and light flash stimuli), new factors were added to the stimuli in a stepwise fasion to be able to differentiate between the effects of movement and realism, while maintaining a similar temporal correlation between stimuli. The results of this study showed no consistent effects of the studied factor, however some differences, in specific stimulus combinations, were found.

In the V-A-V stimulus setup, where the visual stimulus occurred at increased eccentricities, or, when the auditory stimulus was presented exactly in the center, right compared to the auditory stimulus, the realistic stimuli evoked a significantly larger visual bias compared to the flash and noise stimuli with distractors. However, this occurred only at small spatial audio-visual disparities and the difference between the other stimuli was not significant, i.e., the stepwise comparison between the first and second, second and third condtion etc., was not significant. Moreover, in the A-V-A stimulus setup, where it was instead the visual stimulus that occurred more towards the center (or to the left), the realistic stimuli evoked the smallest visual bias. However, again no stepwise comparison was significant. Thus, no consistant effect of any of the factors by itself was found and at most audio-visual disparities no effect was found at all.

Although the results are inconclusive with regards to the effect of realism, the similar results that were found with the various stimuli do call into question the size of the effect that realism could have. The bayesian model comparison showed no improvement with any of the factors that were included, although realism was the closest to improving the model. Studies such as Jackson (1953) found large facilitative effect of stimulus realism. The lesser to no effect found in the present study could indicate, as hypothesized, that the temporal correlation between stimuli in other studies (Jack and Thurlow, 1973; Jackson, 1953; Thurlow and Jack, 1973; Warren et al., 1981) facilitated at least part the effect of realism. These results are in line with Radeau and Bertelson (1977) who compared continuous speech with either a face or a modulated light and found a significant effect of synchronisation, but not realism. However, besides the same temporal correlation in the various conditions, there are some alternative explanations for the smaller/lack of results found in the present study and there are limitations to the present study that warrant discussion.

Firstly, as effects of top-down influences have been shown in some, but not all, cases of audio-visual integration, it is possible that there are specific experimental setups where these effects become relevant. For example, it has been suggested that attention only affects audio-visual integration when the stimulus salience is low (Talsma et al., 2010). As the stimuli were presented well above threshold levels in the present study, the stimulus salience was high. As such, the lack of an strong influence of high level factors on the ventriloquist effect could be the result of the high stimulus salience.

Similarly, it is possible that contextual factors contribute to deciding which stimuli to integrate when there are several competing stimuli. This has been supported by a study by Bailey et al. (2018) where it was shown that realism of the stimuli facilitated integration, but only in a cue-rich environment. As such, the simple setup used in the experiment could contribute to the lack of a consistent effect of realism. However, this explanation cannot fully account for the discrepanzcy between the results from the present and previous studies. For example, Jackson (1953) used a similarly simple setup, but still found a large facilitative effect of realism. Thirdly, as mentioned also in the introduction, a common problem with the ventriloquist paradigm is a response bias (Vatakis and Spence, 2007, 2008). Since audio-visual and visual responses are very similar, it can be difficult to differentiate between true integrative responses and a response bias. Since audio-visual integration decreases the response times, response times are generally used to confirm integration, through a violation of the race model (Miller, 1982). However, both the response method and the delay in the auditory stimuli added substantial variation to reaction times (and the localization responses). Thus, in the present study, it was not possible to test for a violation of the race model to confirm that integration occurred. While, the biases in the localization results for bimodal stimuli compared to unimodal stimuli indicate that integration occurred, it cannot be fully ascertained that the visual bias is not, at least partially, due to response biases towards the visual stimulus.

Finally, the visual bias at small disparities was smaller than anticipated and substantial variation in the visual bias was found. The smaller visual bias is likely due to temporal delay. Studies on the optimal temporal window for audio-visual integration have found varying results. While some studies found integration windows that would still support integration in the present study (Donohue et al., 2010; Lewald and Guski, 2003; Noel et al., 2018) it is possible that for some participants the temporal asynchrony disrupted integration. However, as the temporal disparity was present in all conditions, this disruption should lower the visual bias equally in all conditions. The large variation can be attributed to the response method. As shown in Fig. 4.4, the variance in localization of unimodal stimuli was quite large. Especially at small audio-visual disparities such variation can strongly influence the calculation of the visual bias. The use of discrete reponse options could largely reduce such variance.

Much stronger than the effect of condition, was the effect of the relative positioning of the stimuli, which was dependent on the angle of the auditory stimulus. At ±15 degrees and ±30 degrees the visual bias was larger in the V-A-V setup. This is similar to the results from Hairston et al. (2003) and Charbonneau et al. (2013), where centrally positioned (visual) stimuli evoked a greated bias than more peripheral (visual) stimuli did. Alternatively the increased bias could also be as result of perceptually closer stimuli. As visual localization shows a bias towards the center and auditory localization tends to show a bias away from the center, these biases might counteract each other and reduce the perceived disparity when the visual stimulus is positioned further outwards compared to the auditory stimulus. This hypothesis was tested, but not supported in the study of Godfroy et al. (2003). Either way, the large difference that occurred already at small angles of spatial disparity, could warrant a study that further investigates the effect of relative positioning on the ventriloquist effect, as it could provide further insight in how the biases in unimodal localization affect integration.

When the auditory stimulus was presented at 0 degrees azimuth, there was still an effect of relative stimuli positioning. In this case, the A-V-A and V-A-V setup correspond to whether the visual stimulus occurred left or right to the auditory stimulus, respectively. Curiously, a much stronger visual bias was found when the visual stimuli were presented to the left. It is possible that the response method contributed to this, however the results from Fig. 4.3 make this less likely as pointing responses were similar both in the left and right hemisphere. Thus, an increased variability in pointing results increasing the visual bias by chance is not a likely explanation. Alternatively it could indicate a mismatch in the virtual and real world. Although care was taken to calibrate these, there could still be small differences. Such a hypothetical mismatch, if consistant across participants, could favour integration in one direction as the stimuli line up better similar to the effect of relative stimulus positioning at ±15 degrees and ±30 degrees. In this case results could indicate a shift of the VR world to the left, however the calibration did not indicate the existence of such a shift.

Overall the results of the present study could be valuable for studies using the less natural noise burst and light flash stimuli. These stimuli are much easier to create in laboratory settings but, based on the results here, should still generalize well to more more ecologically valid stimuli. At the same time, the discrepancy in the various experiments investigating the influence of realism on audio-visual integration, suggests that top-down factors influence integration only in more complex experimental settings. To further investigate how well these studies generalize also to real world settings, future studies could explore in which enviroments high-level features become impactful.

4.5 Conclusion

The present study investigated the influence of realism on the ventriloquist effect. No consistant evidence for an effect of movement or realism on the visual bias was found, as in one particular stimulus setup realistic stimuli evoked a slightly stronger visual bias, whereas in another setup they evoked a slightly weaker visual bias. Either way, the results indicate that the effect of realism, if present, is minor at best. While other studies have observed a more noticeable effect of realism, the more realistic conditions involved audio-visual stimuli with higher temporal correlation than the less realistic conditions. The present study suggests that it was the temporal correlation between the auditory and visual stimuli, rather than realism per se, that more strongly facilited integration in previous studies. As such, previous studies on the ventriloquist effect, which used the less natural noise burst and light flash stimuli, should generalize to more realistic stimuli. However, the present study presented only a simple environment. It is possible that high-level factors such as attention and realism influence integration more strongly only in complex settings with competetive stimuli. To differentiate between these factors, future studies might investigate the influence of realism in more complex settings. In simple settings the effect of stimulus realism facilitates integration only to a minor or no extent.

Acknowledgments

This work was carried out as part of the research activities at the Centre for Applied Hearing Research (CAHR) at DTU, supported by GN Hearing, Oticon and WSA. Audiograms were measured by audiologist Rikke Skovhøj Sørensen. The model of the room and loudspeaker array used in the experiment was created by Kasper Duemose Lund.

5

Ambisonics sound source localization with varying amount of visual information in virtual reality[a]

Abstract

To reproduce realistic audio-visual scenarios in the laboratory, Ambisonics is often used to reproduce a sound field over loudspeakers and virtual reality (VR) glasses are used to present visual information. Both technologies have been shown to be suitable for research. However, the combination of both technologies, Ambisonics and VR glasses, might affect the spatial cues for auditory localization and thus, the localization percept. Here, we investigated how VR glasses affect the localization of virtual sound sources on the horizontal plane produced using either 1st-, 3rd-, 5th- or 11th-order Ambisonics with and without visual information. Results showed that with 1st order Ambisonics the localization error is larger than with the higher orders, while the differences across the higher orders were small.

The physical presence of the VR glasses without visual information increased the perceived lateralization of the auditory stimuli by on average about 2°, especially in the right hemisphere. Presenting

---

[a] This chapter is based on Huisman et al., (2021), under revision.

> visual information about the environment and potential sound sources did reduce this HMD-induced shift, however it could not fully compensate for it. While the localization performance itself was affected by the Ambisonics order, there was no interaction between the Ambisonics order and the effect of the HMD. Thus, the presence of VR glasses can alter acoustic localization when using Ambisonics sound reproduction, but visual information can compensate for most of the effects. As such, most use cases for VR will be unaffected by these shifts in the perceived location of the auditory stimuli.

5.1 Introduction

With the recent increase in quality and availability, head mounted virtual reality displays (HMDs) are now regularly used in combination with virtual sound environments to create more realistic and immersive audio-visual experiments (e.g., Echevarria Sanchez et al., 2017; Kessling and Görne, 2018; Suárez et al., 2019). Although for many studies, headphones might suffice as the playback method for this acoustic environment, there are also many cases where loudspeaker playback might be preferred to preserve the participants own head-related transfer function or to be able to wear hearing aids or other ear-worn devices. This is where potential problems can arise, as recent studies have shown that HMDs affect the acoustic signals (Ahrens et al., 2019; Genovese et al., 2018; Gupta et al., 2018).

The shape of the ear, head and body modify sound as it reaches the ear, resulting in interaural level differences, interaural time differences and spectral changes which are used for the localization of sound sources (for an overview, see Blauert (1997) or Hartmann (1999)). The added volume of the HMD modifies these cues, increasing the lateralization of the perceived location of stimuli (Ahrens et al., 2019; Gupta et al., 2018). Such changes in the perceived location of the sound could not only affect the perceived spatial location of a sound, but also the integration of audio-visual stimuli.

Current experiments have only assessed the effect of the HMD when presenting sound from a single loudspeaker. However, VR will regularly require more complex playback methods to be able to present spatial audio from any location, independent of the loudspeaker setup. Ambisonics is a commonly used playback method for such a purpose. It encodes audio by decomposing a sound field into spherical harmonics and can provide full-sphere surround sound (Gerzon, 1973). In its basic form (1st-order Ambisonics), four channels corresponding to the first four spherical harmonics are used to encode the sound field. However, additional spherical harmonics can be included to improve the directional resolution of the reproduction (Ahrens et al., 2020; Bertet et al., 2013; Gerzon, 1973), referred to as higher-order Ambisonics (HOA). To accurately reproduce the encoded sound field, the number of loudspeakers should match the number of spherical harmonics used for the encoding. Thus, at least $(n+1)^2$ loudspeakers are needed for a full spherical representation and $2n+1$ loudspeakers for horizontal-only, where n is the Ambisonics order. An increase in Ambisonics order has been shown to result in an increased localization accuracy (Bates et al., 2007; Bertet et al., 2013; Pulkki and Hirvonen, 2005; Thresh et al., 2017).

Ambisonics reproduces the sound field through interactions of the audio signals from multiple loudspeakers simultaneously, independent of the direction of the source. Thus, the effect of the HMD on sound localization that has been shown with single loudspeaker playback, might be different when employing

Ambisonics reproduction.

In virtual audio-visual scenes the audio is not presented in isolation, but in combination with visual information, which is known to strongly influence sound localization (e.g., Dufour et al., 2002; Gori et al., 2014; Tabry et al., 2013). When audio and visual stimuli are presented in close temporal and spatial proximity, they are integrated into one common percept, increasing the accuracy and precision of localization (e.g., Alais and Burr, 2004; Freeman et al., 2018; Odegaard et al., 2015). As a result of this process, when the audio and visual stimuli are not exactly at the same position, but still integrated, the perceived location of the auditory stimuli is shifted strongly towards that of the visual stimulus (the so called 'ventriloquist effect' (e.g., Alais and Burr, 2004; Jackson, 1953; Lewald and Guski, 2003; Thurlow and Jack, 1973). Therefore, it is possible that potential shifts caused by the HMD can be compensated for with visual information. Indeed, when presenting visual information about the environment and potential sources, such as loudspeakers, (Ahrens et al., 2019) saw a decrease in the effect of the HMD on sound source localization, compared to when no visual information was presented.

The aim of the current study was, to investigate the effects and the interactions between the HMD, Ambisonics and visual information on the perceived sound location. Therefore, a sound localization experiment with hand-pointing was performed. Participants located sounds sources which were simulated from angles between -90° to 90° azimuth, using 1st-, 3rd-, 5th- and 11th-order Ambisonics, with and without an HMD. Participants were first tested blindfolded, to avoid biasing effects of any visual information. Next, they performed the same localization task with visual information to test if visual information can compensate for potential effects of the HMD.

5.2 Methods

5.2.1 Participants

21 participants (9 females and 12 males, average 25 ± 3 years) were recruited to participate in the experiment. To ensure normal hearing, audiometric thresholds were measured at octave frequencies between 125 and 8 kHz. Data from participant 7 were excluded due to audiometric thresholds above 20 dB HL. Data from the remaining 20 participants were used in the analysis. The participants were compensated with an hourly rate of 122 DKK. The experimental procedure was approved by the Science-Ethics Committee for the Capital Region of Denmark (H-16036391) and all participants provided written informed consent.

5.2.2 Acoustic reproduction

The experiment took place in the Audio-Visual-Immersion Lab (AVIL) shown in Fig. 5.1, left panel. AVIL is an anechoic chamber containing 64 KEF LS50 loudspeakers, placed in a 4.8 m diameter sphere around a height adjustable chair. For this experiment, only the horizontal ring, containing 24 loudspeakers spaced equidistantly (15° separation), was used for sound reproduction. Audio signals were generated in MATLAB (The Mathworks, Natick, MA) and sent, via 2 TESIRA biamp DSPs with TESIRA SOC-4 Audio DSP cards (biamp Systems, Beaverton, OR), to the amplifiers (Sonible GmbH, Graz, Austria) that drive the loudspeakers.

5.2.3 Visual reproduction

The virtual environment, shown in Fig. 5.1, was a 1:1 reproduction of AVIL. This environment was created in UNITY3D (Unity Technologies, San Francisco, CA) and presented via an HTC VIVE PRO (HTC Corporation, New Taipei City,

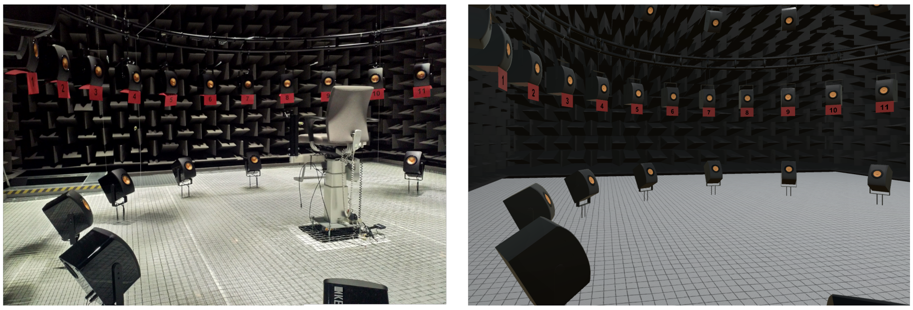


Figure 5.1: Here all the combinations of auditory and visual stimuli in the bimodal condition are shown. Each combination (indicated by a dot) was repeated 3 times. In later bimodal figures the results from the left hemisphere are mirrored, such that a negative disparity indicates that the visual stimulus occurred in the direction of the center, whereas positive disparities indicate that the visual stimulus occurred further outwards. The densest sampling occurred from -30 to 15 degrees audio-visual disparity.

Taiwan) VR setup. Three HTC VIVE trackers at known positions were used to ensure the spatial alignment between the real and virtual world, by recalibrating the virtual world if discrepancies larger than 1 cm occurred (see Ahrens et al. (2019) for details). When the HMD was not in use, it was placed in front of the participant in sight of the HTC lighthouses that track the position of the HMD and the handheld controllers, to ensure proper calibration of the virtual world also during the real-world conditions.

5.2.4 Pointing Apparatus

A handheld HTC VIVE controller was used to record the localization judgements of the participants in all conditions. By pressing the trigger button on the back of the controller, their judgement was recorded. A model of this controller was rendered in the virtual environment, however there was no physical representation of the participants themselves in the virtual environment. As it was hypothesized that visual information of the body could affect pointing, a condition was included to measure the difference in pointing at visual targets in the real and virtual environment. In the conditions where this pointing bias might have affected data (i.e., when participants had access to visual information), a correction for this pointing bias was applied.

5.2.5 Stimuli and spatialization

The stimuli were created in MATLAB. The auditory stimulus consisted of a 240 ms pink noise burst with a 20 ms ramp, raised cosine window. The stimuli were presented, on average, at 65 dB sound pressure level (dB SPL). To reduce directional loudness cues (Makous and Middlebrooks, 1990; Musicant and Butler, 1984) the sound level was roved by values drawn from a uniform distribution between ± 3 dB. The stimuli were spatialized using Ambisonics panning (Gerzon, 1973). The highest order Ambisonics that can be reproduced with a 24-loudspeaker array in the horizontal plane is 11th-order. To decrease the Ambisonics order, the number of loudspeakers used to produce the stimulus was reduced and the decoder was adjusted accordingly. In this manner stimuli were presented in 1st-, 3rd- and 5th-order Ambisonics, using, respectively, 4, 8 and 12 loudspeakers, spaced equidistantly. The individual loudspeakers used to reproduce each Ambisonics order are indicated in Fig. 5.2. An Ambisonics decoder with dual-band energy normalization was used as in Favrot and Buchholz (2010) and Ahrens et al. (2020)). The low-frequency region received no weighting (basic decoding) and in the high-frequency region 'max-re' decoding was applied. The transition frequency between the weighting methods was set to the Ambisonics order multiplied by 800 Hz. Stimuli were presented from -90° to 90° azimuth in 7.5-degree steps, i.e., at each loudspeaker and halfway in between each loudspeaker. Each position was repeated five times for each Ambisonics order in each condition, resulting in 500 trials per auditory condition and 2000 auditory trials in total.

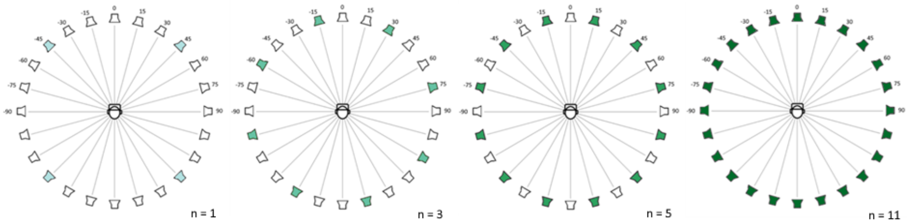


Figure 5.2: The loudspeakers used for the reproduction of 1st-, 3rd-, 5th- and 11th-order Ambisonics, respectively. The loudspeaker pictogram indicates the loudspeaker positions and the colouring the loudspeakers that were used for each of the Ambisonics order conditions.

5.2.6 Pointing bias

To measure the potential pointing bias, as was found in Ahrens et al. (2019) perhaps as a result of a lack of Avatar in VR Schwind et al. (2018), participants were asked to point, both in the real environment and in the virtual environment, at static visual targets, namely the loudspeakers. As shown in Fig. 5.1, the loudspeakers, that were positioned between -90° to 90° azimuth, were numbered from 1 to 13. In this task, participants were shown a number, either on a virtual screen in VR or on an iPad that was placed in front of the participant, and then pointed, in the same manner as in the auditory localization task, at the center of the loudspeaker with that number. The iPad and the virtual screen were only present during this task. Note that participants only pointed at visual, instead of auditory, sources in this last condition. Again, each position was repeated 5 times for each position, resulting in 65 trials per condition for a total of 130 trials in block 3.

5.2.7 Experimental conditions

The experiment consisted of six conditions, presented in three blocks (see table 5.1). The blocks were presented in a fixed order, but within a block, the order of the conditions was counterbalanced across participants. The various Ambisonics orders were tested interleaved, i.e., within each condition the stimuli were presented in all Ambisonics orders. To investigate the effect of the HMD, without any visual biases, participants performed the localization experiment blindfolded in the first block. For the condition with the HMD, the HMD was placed over the blindfold. The second block then investigated if visual information could compensate for the potential effects of the HMD found in the first block. Finally, in the third block, the potential pointing bias, due to the lack of a physical representation of the participants in VR, was measured. Since participants did not have access to any potential biasing visual information in the blindfolded condition, no corrections were applied here. Each acoustic condition started with 20 training trials to ensure participants understood and followed the instructions with regards to pointing and how to proceed through the block. The different conditions are summarized in table 5.1 below.

Block	*Condition*	*Visual information*	*HMD*	*Stimulus*
1	1	Blind-folded	No	Acoustic
	2	Blind-folded	Yes	Acoustic
2	3	Real environment	No	Acoustic
	4	Virtual environment	Yes	Acoustic
3	5	Real environment	No	Visual
	6	Virtual environment	Yes	Visual

Table 5.1: The experiment was divided into six conditions presented in three blocks. Blocks were presented in a fixed order, but within a block, conditions were counterbalanced across participants.

5.2.8 Procedure

The experiment was conducted in two sessions with a maximum of 2.5 hours, with at least three enforced breaks per session (halfway through and in-between blocks). At the start of the experiment, participants were told that sounds would be presented using an Ambisonics sound system which could simulate sounds from anywhere in the room. They were instructed to face forward from their seated position before and during each stimulus, and to point with the controller at the direction where they perceived the sound originated from. After stimulus presentation, the participants were allowed to freely turn around while pointing. After indicating the perceived origin of the stimulus, participants faced forward again and 1.5 seconds after recording the response, the next stimulus played automatically. Participants were instructed to use their entire arm and fixate their wrist whilst pointing and to maintain the same pointing method throughout the experiment. Finally, participants were encouraged to take additional breaks during the experiment (besides the aforementioned three breaks per session) if they needed them.

Participants were then guided into the experimental room, seated at the center of the array and were shown the VR headset. In the first block they were then blindfolded and, depending on the condition, were either fitted with headset or the headset was placed in front of them. In the conditions with visual information participants were fitted with the headset and given the option to adjust the settings to their preferences. The non-rotating chair was then raised to ensure their ears were positioned at height of the loudspeakers.

5.2.9 Analysis

Pointing bias

For the calculation of the pointing bias, responses with an error larger than 15° were treated as outliers and removed from the analysis. 0.011% of the visual trials were rejected based on this criterion. For every participant, the mean pointing bias was then calculated per visual stimulus location in both the real and virtual environment. For the stimuli presented in between loudspeakers, where no pointing bias was measured, the subtracted pointing bias was calculated by linearly interpolating the nearest measured pointing biases. Each response in the second block (with visual information) was then corrected by subtracting the individual pointing bias, i.e., same participant, environment, and location. The visual localization data itself (without the interpolated data points) was also analyzed. For this, a mixed linear model was fitted to the responses with the stimulus location and experimental condition as fixed effects, while the participants and repetitions were considered as random effects. For the computational analysis, the statistical software R (R Core Team, 2020) was used together with the “lmerTest” package (Kuznetsova et al., 2017).

Auditory localization

No outlier removal was conducted on the acoustic localization data. As participant 20 was left-handed, which has been shown to affect auditory space perception (Ocklenburg et al., 2010), data from the left and right hemisphere was flipped for the analysis. For the statistical analysis of the auditory localization responses, a mixed linear model was fitted to the (corrected) azimuth error using the same computational methods as for the pointing bias. As mentioned previously, in the blindfolded conditions no correction for the pointing bias was applied, as no visual information was available in both the real and virtual world. However, as described in section 2.8.1. in the visual condition, the localization error was corrected by subtracting the pointing bias. The stimulus location, Ambisonics order and condition were considered fixed effects, while the participants and repetitions were considered as random effects. To investigate how the different factors affected the localization performance, post-hoc analyses of within factor comparisons were performed. To determine the effect of the HMD, the blindfolded conditions with and without the HMD were compared. To find the effect of visual information on this effect, results between blocks one and two were compared.

5.3 Results

5.3.1 Pointing bias

Fig. 5.3 shows the signed pointing error to visual objects for the real (blue) and virtual (orange) environment. The pointing error is defined as the difference between the response angle and the source angle in degrees azimuth. A large variation in pointing behavior across participants can be seen, especially at more eccentric angles. Additionally, a shift in the pointing direction towards the left side (negative angles) can be seen in the virtual environment relative to the real environment. The statistical analysis of the responses showed a significant difference between the pointing in the real environment versus pointing in the virtual environment $[F_{1,2563} = 166.294, p < 0.0001]$. The post-hoc comparison estimated the effect size between the real and the virtual environment to be 1.76° $[t_{2563} = 12.896, p < 0.0001]$. Additionally, an effect of the stimulus location was found $[F_{12,2563} = 49.953, p < 0.0001]$, but no significant interaction between the environment and the stimulus location $[F_{12,2551} = 1.309, p = 0.2058]$.

5.3.2 Effect of Ambisonics order

Fig. 5.4 shows the signed localization error, that is the difference between the response angle and the source angle in degrees azimuth, as a function of the source angle for the different Ambisonics orders. As no interaction between the conditions and the Ambisonics order was found, data from all conditions are included in this figure. The localization error was found to vary with the stimulus location $[F_{24,39806} = 142.178, p < 0.0001]$ and Ambisonics order $[F_{3,39806} = 29.631, p < 0.0001]$. Moreover, a significant interaction between the Ambisonics

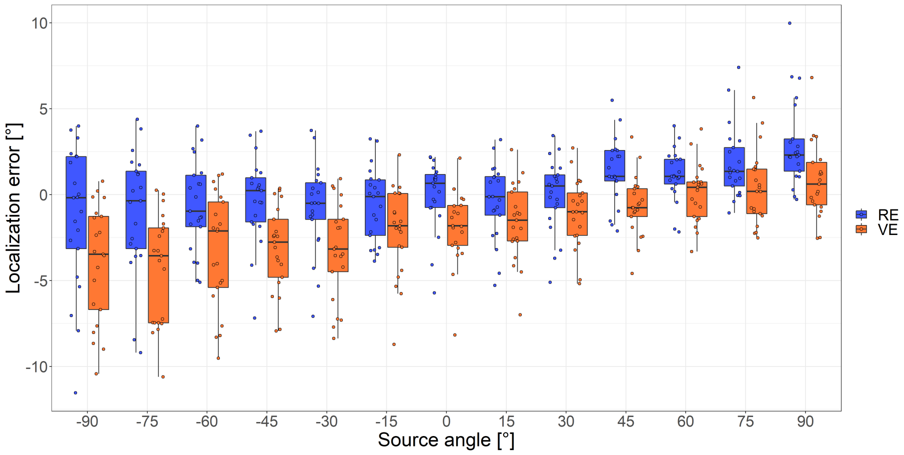


Figure 5.3: The pointing bias in the visual pointing task is shown per angle for both the virtual (VE) and the real (RE) environment. The dots indicate the mean pointing error per person and the boxplot shows the distribution. The boxes extend from the first to the third quartile, with the median shown as the center black line.

order and the stimulus location was found [$F_{72,39806} = 73.144, p < 0.0001$]. A discrepancy between 1st order Ambisonics responses and higher order Ambisonics responses can be seen. This difference was significant at all angles, except at 7.5°-22.5° (1st - 3rd, [7.5°: $t_{39806} = -0.618, p = 1.00$], [15°:$t_{39806} = -1.372, p = 1.00$], [22.5°: $t_{39806} = -2.175, p = 0.1778$]; 1st - 5th, [7.5°: $t_{39806} = -1.207, p = 1.00$], [15°: $t_{39806} = -1.372, p = 1.00$]; 1st - 11th, [7.5°: $t_{39806} = -1.015, p = 1.00$]). As can be seen in Fig. 5.4, for the 1st-order Ambisonics results, the absolute localization error increased with the absolute source azimuth. Such consistent increase of error with azimuth was not found when higher Ambisonics orders were used to simulate the sources. For all Ambisonics orders, at the outermost angles, i.e., ±82.5-90°, the simulated sources were perceived insufficiently lateralized. Here again results were most pronounced when 1st-order Ambisonics was used. With the 1st-order Ambisonics, the localization error reached ±30° at ±90° azimuth, i.e., participants perceived the source two entire loudspeakers closer to the center. For sources reproduced using higher Ambisonics orders this discrepancy was highly reduced, although not fully diminished. Few differences were found between the perceived location of the presented sources with the Ambisonics orders larger than one; a small decrease in error was observed when increasing the Ambisonics order at the outer source angles (3rd - 5th, [90°: $t_{39806} = -4.172, p = 0.0002$]; 3rd - 11th, [82.5°: $t_{39806} = -3.889, p = 0.0006$], [90°: $t_{39806} = -4.099, p = 0.0002$]). Besides that, there was a small difference between 3rd- and 5th-order at -67.5° [$t_{29806} = 2.727, p = 0.0384$] and 5th- and 11th-order Ambisonics at 52.5° [$t_{29806} = -2.770, p = 0.0337$]. At all other angles no significant difference within the higher Ambisonics orders were found.

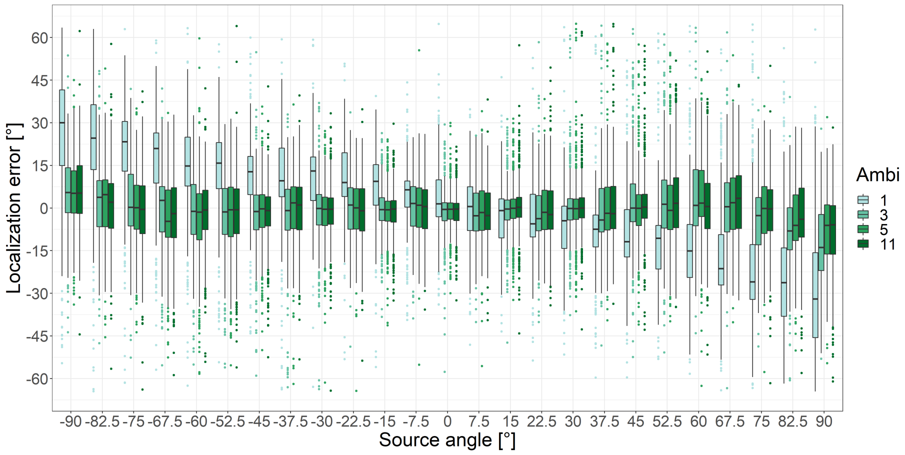


Figure 5.4: Here the perceived source angle is plotted as a function of the stimulus position. Data from all acoustic conditions are included in the figure, except for 450 outliers (1.125%) that occurred outside of the figure boundaries (localization error larger than ± 65°). Data are separated by the Ambisonics order used to produce the stimuli. The boxes extend from the first to the third quartile, with the median shown as the center black line. Responses that exceeded 1.5 times the interquartile range are considered outliers and are indicated as dots.

5.3.3 Effect of the HMD

The localization error in the blindfolded conditions is shown in Fig. 5.5. Data without the HMD are shown in blue and data with the HMD are shown in orange. The localization error varied with presentation angle [$F_{24,39806} = 142.178, p < 0.0001$] and depended on the condition, i.e., with or without the HMD [$F_{3,39806} = 59.077, p < 0.0001$]. Additionally, a significant interaction was found between the presentation angle and the condition [$F_{72,39806} = 12.819, p < 0.0001$]. At negative angles, i.e., in the left hemisphere, the localization error tended to be more negative, i.e., sources were perceived more to the left, when wearing the HMD, compared to when participants were not wearing the HMD. At positive angles, i.e., in the right hemisphere, instead sounds were perceived more to the right when wearing the HMD. The post-hoc analysis showed that the increase in the perceived lateralization of the sound sources when wearing the HMD was larger in the right hemisphere than in the left hemisphere. In the right hemisphere the difference in the conditions was significant at all angles ([7.5° – 67.5°: $t_{39806}, p < 0.0001$; 75°: $t_{39806}, p = 0.0001$; 82.5°: $t_{39806}, p = 0.0073$; 90°: $t_{39806}, p = 0.0152$]). In the left hemisphere, the effect of the HMD only reached statistical significance at -82.5° [$t_{39806} = 3.209, p = 0.0080$] and -37.5° [$t_{39806} = 2.692, p = 0.0427$]. Furthermore, the difference between the conditions with and without HMD was less pronounced in the left hemisphere; the maximum difference between the RE (without HMD) and VE (with HMD) was 3.5° in the left hemisphere, and between 3.7° and 8.4° in the right hemisphere.

5.3.4 Effect of visual information

Fig. 5.6 shows the effect of the HMD when visual information is present. Data from the condition without the HMD is shown in blue and data from the condition with the HMD is shown in orange. Although, at most angles, the disparity between the two conditions was reduced either partially or fully, compared to the disparity in Fig. 5.5, significant differences between the RE and VE condition remained. In the left hemisphere a significant difference was still found at -60° azimuth [$t_{39806} = 3.181, p = 0.0088$], while in the right hemisphere significant differences remained at several angles: [0°:$t_{39806} = -3.386, p = 0.0043$], [7.5°: $t_{39806} = -4.238, p = 0.0001$], [22.5°: $t_{39806} = -3.629, p = 0.0017$], [45°:

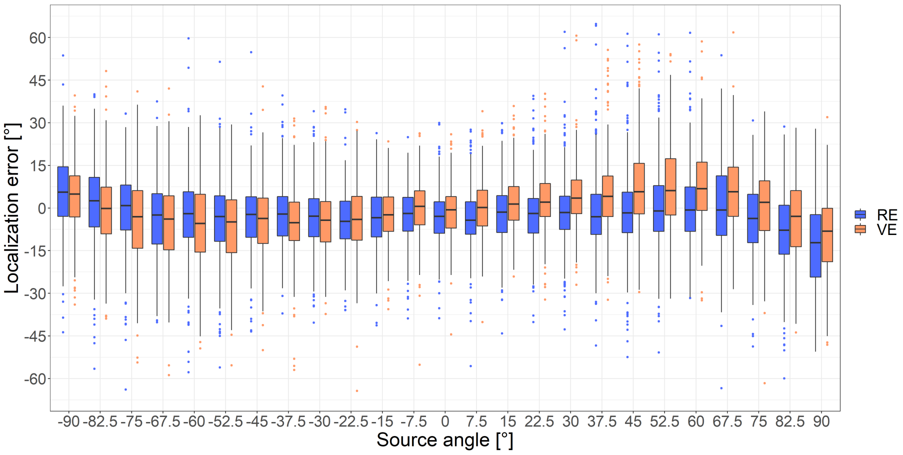


Figure 5.5: The localization error in the blindfolded conditions with (orange) and without (blue) HMD. Due to the discrepancy in the first versus higher order Ambisonics data, only higher order (3rd-, 5th-, 11th-) Ambisonics data is included in the figure. The boxes extend from the first to the third quartile, with the median perceived response shown with black lines. Responses that exceeded 1.5 times the interquartile range are considered outliers and are indicated as dots. 158 outliers (0.527%) are not shown as they occurred outside of the figure boundaries (localization error larger than ± 65°).

$t_{39806} = -2.643, p = 0.0493$], [52.5°: $t_{39806} = -5.223, p < 0.0001$], [60°: $t_{39806} = -3.677, p = 0.0014$], [67.5°: $t_{39806} = -2.822, p = 0.0287$]. These difference between Fig. 5.5 and 5.6 shows that visual information of the loudspeaker locations affected the localization error.

Moreover, a pattern in the localization error was found, which was consistent with participants pointing at visual loudspeakers locations. From Fig. 5.6 it can be seen that the error is smaller at integer multiples of 15° than at the other angles. As sound was presented not only at exact loudspeaker positions (integer multiples of 15°), but also halfway in between, it was hypothesized that the visual information of the loudspeaker location might have an effect on the response pattern. To more clearly investigate this behavior, data from Fig. 5.6 (right hemisphere only) was replotted in Fig. 5.7.

Fig. 5.7 shows the data from Fig. 5.6, as a violin plot, where the probability density of the responses is shown per azimuth angle. Darker colors indicate

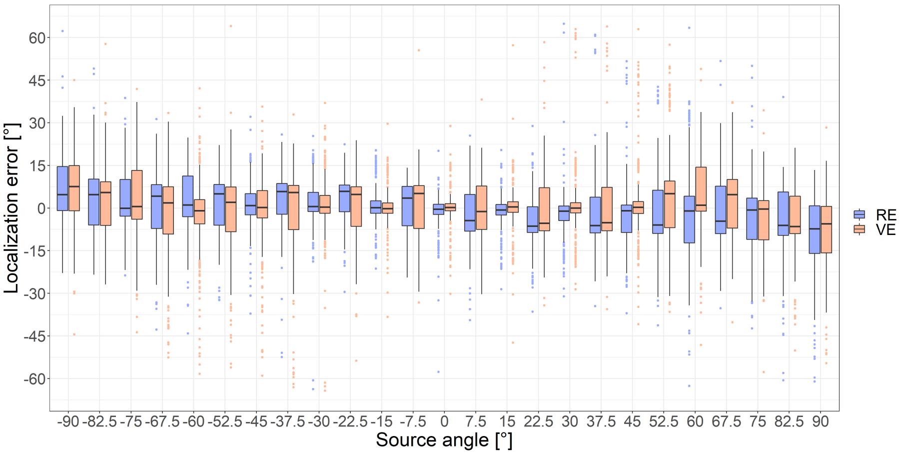


Figure 5.6: Boxplot of the (pointing bias corrected) localization error with (orange) and without HMD (blue) in the conditions with visual information. Only higher order Ambisonics data is included in the figure. The boxes extend from the first to the third quartile, with the median perceived response shown with black lines. Responses that exceeded 1.5 times the interquartile range are considered outliers and are indicated as dots. 142 outliers (0.473%) are not shown as they occurred outside of the figure boundaries (localization error larger than ± 65°).

that the sound was simulated at an angle with a loudspeaker, while the lighter colors indicate that sound was presented halfway in-between loudspeakers. At small angles, when sound sources were simulated at loudspeaker locations, the errors were unimodally distributed around 0° localization error. When sound was instead simulated in between loudspeakers, responses were bimodally distributed, i.e., responses were split between the two closest loudspeakers. At larger source angles, multiple peaks can be seen in the distributions. The centers of these peaks remain consistent with loudspeaker locations.

5.3.5 Mean results

Fig. 5.8 shows the mean absolute localization error for the acoustic conditions, separated by Ambisonics order and condition. The benefit of using higher order Ambisonics, compared to 1st-order Ambisonics is clearly visible in the reduced localization error. Additionally, visual information, indicated in light colors, decreases the localization error when higher order Ambisonics is used. Finally,

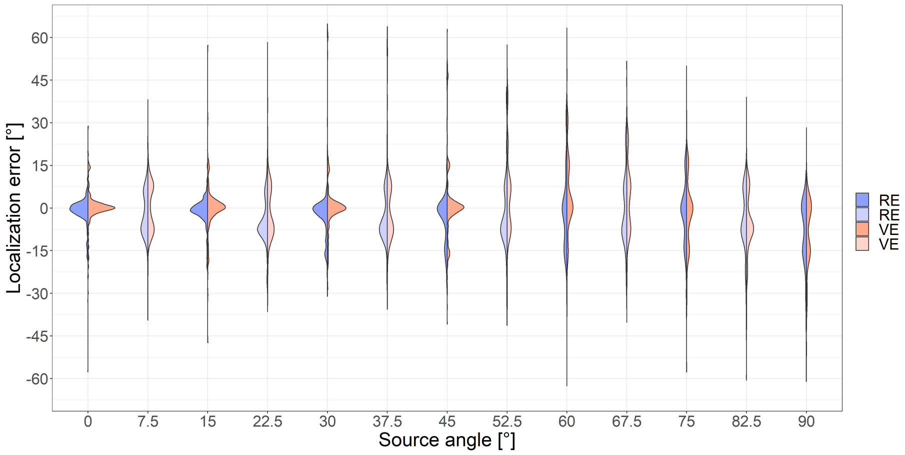


Figure 5.7: Violin plot of the (pointing bias corrected) localization error with (orange) and without HMD (blue) in the conditions with visual information, right hemisphere data only. The distributions of the localization error are shown as function of the azimuth location when visual information is presented. Both the environment (RE and VE) as well as the presence of a loudspeaker at the azimuth location are color-coded. The violin plot shows the spread of the responses per angle and per condition, in the form of a sideways histogram. Only higher order Ambisonics data, right hemisphere, are included. 86 outliers (1.103%) are not shown as they occurred outside of the figure boundaries (localization error larger than ±65°).

the effect of the HMD can be seen when comparing the blue and the orange boxes, as an increase in the localization error for the higher Ambisonics orders.

5.4 Discussion

5.4.1 Shift in the perceived location due to the HMD

As in Ahrens et al. (2019) and Gupta et al. (2018) we found that the HMD increased the perceived lateralization of the stimuli. Interestingly, despite the symmetric setup, this effect was found to be stronger in the right hemisphere. In the left hemisphere, there was a similar trend in the data, but it was much smaller and not significant. In contrast, Ahrens et al. (2019) found the larger effect in the left hemisphere (although significance levels were only reached at few stimulus positions, which might be related to fewer participants). In both the RE blindfolded and VE blindfolded indications of bias were found. In

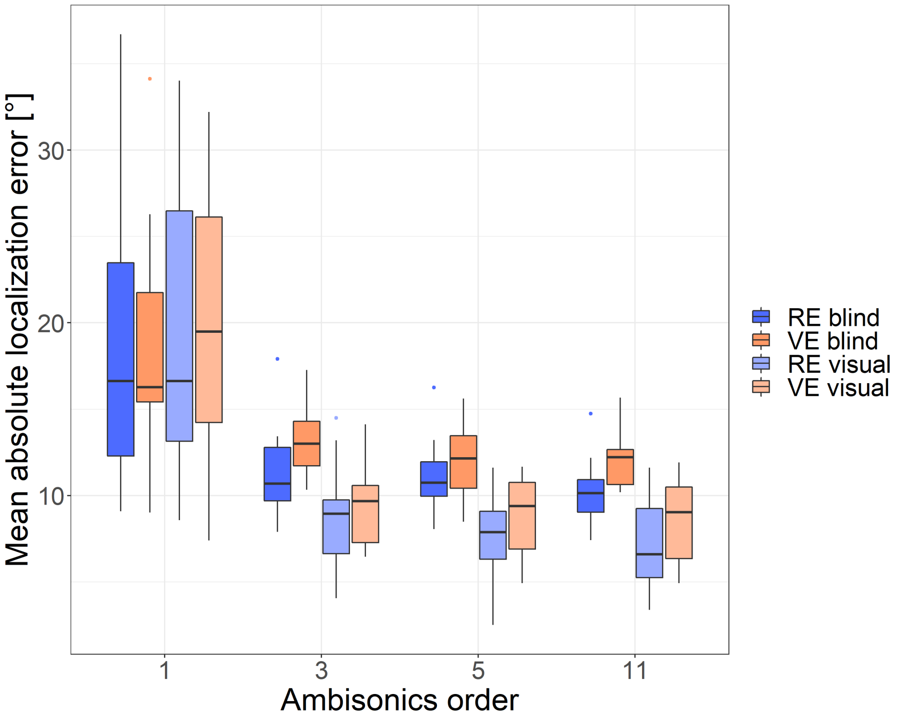


Figure 5.8: The mean absolute localization error per condition, separated by Ambisonics order. The average localization error was grouped by angle, condition, and azimuth, after which the mean localization error was calculated. The boxes extend from the first to the third quartile, with the median perceived response shown with black lines. Responses that exceeded 1.5 times the interquartile range are considered outliers and are indicated as dots.

the RE participants tended to point slightly more to the left, whereas in the VE participants tended to point slightly more towards the right. Together these biases might have increased the effect of the HMD in the right hemisphere, while decreasing the effect in the left hemisphere. Interestingly this difference remained also in the visual condition, although strongly reduced. The distribution of the responses showed that, when in doubt, participants tended to favor the closest loudspeaker to the left more often in the RE, whereas in the VE they tended to favor the closest loudspeaker to the right. To confirm this, we calculated the percentage of responses that occurred to the right of the actual position. In the right hemisphere, 38.05% of the responses occurred to the right of the speaker in the RE, compared to 49.85% in the VE. That the effect occurred in both the blindfolded and the visual condition suggests that there was some consistent effect of the HMD. It could be that the HMD affects pointing also when there is no visual information, potentially due the size and weight of the HMD.

5.4.2 Ambisonics order does not influence the effect of the HMD

Although it was hypothesized that the effect of the HMD might vary with Ambisonics order, no such interaction was found. Localization itself, however, was clearly affected by Ambisonics order. Firstly, when using 1st-order Ambisonics, stimuli were consistently perceived slightly more towards the center, especially at the most extreme angles. This difference between the intended location and the perceived location was large enough that responses were regularly shifted by an entire loudspeaker at the outer angles in the condition where visual information was available. This suggests that the 1st-order Ambisonics system could not fully achieve the intended lateralization. This "under lateralization" of the stimuli was also found by Pulkki and Hirvonen (2005), who also found that the lateralization appeared capped between 60°-70° azimuth. As in previous studies ((Bates et al., 2007; Bertet et al., 2013; Pulkki and Hirvonen, 2005; Thresh et al., 2017), increasing the Ambisonics order from 1st- to 3rd-, greatly improved lateralization. Increasing the Ambisonics order further, however, did not improve localization much, again in line with previous studies (Thresh et al., 2017). Similar effects of the Ambisonics order have also been shown for speech intelligibility (Ahrens et al., 2020). However, in the current experiment participants were seated in the center of the array. For off-center listenening positions, studies have found that increasing Ambisonics order improves localization accuracy (Stitt et al., 2013, 2014) also beyond the third order (Frank et al., 2008). Moreover, the current study tested localization in anechoic conditions. Previous studies showed that reverberation can mask some of the errors that are due to the Ambisonics reproduction (Ahrens et al., 2020; Oreinos and Buchholz, 2015). Thus, adding reverberation might further affect localization performance and interactions with the Ambisonics orders (Sampedro Llopis et al., 2019).

5.4.3 Some compensation from visual information

Visual information strongly affected the responses of the participants and reduced the effect of the HMD at many locations. However, due to stimuli being presented also in between loudspeakers, it did not always improve localization. Similarly, the tendency to point at loudspeakers also increased the effect of the HMD at some locations where the HMD shift affected which loudspeaker was perceived as the nearest (see for example 5.6, 52.5°). In the right hemisphere visual information reduced the effect of the HMD at almost all angles. In the left hemisphere, the results were mostly unaffected by visual information, as the difference between the VE and RE were already very small. Nevertheless, significant differences between localization with and without the HMD remained. However, these remaining differences can also be the result of the correction applied.

5.4.4 Limited impact on VR

As visual information compensated for most of the effect of the HMD, it is likely that in most use-cases, such as playing audio-visual recordings, VR games etc. the effects of the HMD will be negligible. Even if the effects are not fully compensated for, it is likely that presence (“the feeling of being there”), one of the key factors of VR, remains unaffected. Presence has been shown to be facilitated by audio (Hruby, 2019; Kern and Ellermeier, 2020; Larsson and Västfjäll, 2007; Nordahl, 2004), especially spatialized audio (Hendrix and Barfield, 1995; Riecke et al., 2009). However, the accuracy of the spatialized audio was found not to influence presence Riecke et al. (2009). Moreover, the shift in the perceived location in the frontal area (where we are most sensitive to spatial audio-visual disparities) is within the spatial distance where visual and auditory stimuli are integrated (Godfroy et al., 2003; Lewald and Guski, 2003; Stenzel, 2017; Thurlow and Jack, 1973). As a result, integration of the audio-visual scenes should be mostly unaffected. However, because a shift in the perceived location of the auditory stimulus is present, it will be important to take the effect of the HMD into account in experiments where the exact positioning of the stimuli is relevant, such as audio (-visual) localization experiments.These results only extend to the combination of the HMD with loudspeaker reproduced Ambisonics, headphone reproductions will not be shifted in the same way.

5.5 Conclusion

In line with previous studies, we found that the HMD increased the perceived lateralization of auditory stimuli, especially in the right hemisphere. In the left hemisphere the effect was much smaller and only significant at a few angles. However, significant effects mostly occurred in the right hemisphere. Although an interaction between the Ambisonics order used to present the stimuli and the effect of the HMD was hypothesized, no such interaction was found. Localization itself, however, was found to be strongly affected by the Ambisonics order. Sounds presented with 1st-order Ambisonics were generally perceived to be originating from a more central location. This “under lateralization” of the stimuli increased with azimuth, reaching an error of up to 30° at a source angle of 90°. Increasing the Ambisonics order from 1st- to 3rd-order greatly improved the accuracy of the reproduction, however increasing the order beyond this only minimally affected the localization accuracy at a few angles. Finally, visual information led to a compensation for most of the effect of the HMD, but not fully, as significant differences between the RE and VE conditions remained. The impact of this shift in the perceived location is likely small, but in cases where the exact location of the stimuli is important, it will be important to account for effects of the HMD on the auditory stimuli.

Acknowledgments

This work was carried out as part of the research activities at the Centre for Applied Hearing Research (CAHR) at DTU, supported by GN Hearing, Oticon and WSA. Audiograms were measured by audiologist Rikke Skovhøj Sørensen. The model of the room and loudspeaker array used in the experiment was created by Kasper Duemose Lund.

6

Increase of the spatial integration window with age but not hearing loss[a]

Abstract

Audio and visual information can be integrated over a range of temporal and spatial stimulus disparities. The temporal disparities over which stimuli are integrated, the ‘temporal integration window’, has been shown to increase with age and hearing loss. However, it has remained unclear how age and hearing loss influence the spatial disparity over which stimuli are integrated, i.e., the spatial integration window. The present study examined the spatial window of audio-visual integration in eight young normal-hearing, six older normal-hearing and seven older hearing-impaired participants. To assess the spatial integration window, the just-noticeable difference (JND) in angle was measured for each participant in four conditions and at five azimuthal angles (±30°, ±15° and 0°), using a left-right discrimination task. These four conditions represented an audio-only condition, a visual-only condition, a congruent audio-visual condition and an incongruent audio-visual condition. The audio-only and visual-only conditions served as a reference for the unimodal performance, the congruent audio-visual condition served

---

[a] This chapter is based on Huisman et al., (2021), submitted.

to confirm the beneficial effects of integration and the incongruent condition served as a measure of the spatial integration window. In this incongruent audio-visual condition, the visual stimulus was always presented from the same location, halfway in between the auditory stimuli, to ensure that participants could not rely on visual information to perform the task. With this JND paradigm, an estimation of the spatial window was obtained by varying the distance between the auditory stimuli, and thereby the distance between the auditory and visual stimuli. On repeated successful trials, the spatial distance was decreased, increasing the probability of integration and highly increasing the difficulty of the task. When the distance between the auditory stimuli was large enough, such that the auditory stimuli were no longer integrated with the visual stimuli, listener judgment was assumed to be based on audio-only processing. Hence, an increase in the JND in the incongruent condition, compared to the audio-only condition, was considered to relate to width of the spatial integration window. Results showed that the spatial integration window increased with age but was unaffected by hearing loss. This increase of the spatial integration window with age was significant at 0° azimuth whereas at other angles, the results in the incongruent audio-visual condition did not differ from the audio-only performance.

6.1 Introduction

Sensory systems decline with age (see Fozard (1990) for a review). According to the World Health Organization (WHO, 2021), the prevalence of hearing loss steeply increases after the age of 60 years, with about 25% of adults between 70-75 years and more than 50% of adults above 85 years suffering from a hearing loss. Age-related hearing loss is typically also more severe, with 58% of people with moderate or severe hearing loss being adults above 60 years old. This hearing loss not only affects auditory perception, but also the way that auditory and visual information is integrated.

Due to the differences in sensory and neural processing of the stimuli in the auditory and visual system, the perceived timing and location of congruent audio-visual stimuli may not be congruent in the brain. As such, signals can be integrated even if there is a difference in the stimulus onset and/or the stimulus location. The probability of integration decreases with increasing temporal and spatial disparities (e.g., Godfroy et al., 2003; Jack and Thurlow, 1973; Jackson, 1953; Lewald and Guski, 2003; Stenzel et al., 2019; Thurlow and Jack, 1973). The range of stimulus onset asynchronies and the spatial disparities where the probability of integration is higher than 50% has been referred to as the temporal and spatial integration window, respectively (e.g., Diederich and Colonius, 2004; Meredith, 2002; Powers et al., 2009; Rohe and Noppeney, 2015).

Multiple studies have shown that localization accuracies affect the size of the integration window. In the spatial domain, through a variation of the reliability of the visual stimuli, Rohe and Noppeney (2015) found that spatial windows were larger when the reliability of the visual stimulus was reduced. In the temporal domain, Bidelman (2016) found that musicians, who were better and faster at processing concurrent audio-visual cues than non-musicians, exhibited smaller temporal windows than the non-musicians. Similarly, Di Luzio et al. (2021) found smaller temporal windows in gamers, compared to non-gamers. The decreased reliability of the auditory system in temporal tasks has been associated with an increased temporal window (Schormans and Allman, 2018), although not consistently (Başkent and Bazo, 2012), and visual impairments have been associated with an increased spatial window (Leo et al., 2008). However, these studies on reliability and impairment only varied the reliability of the dominant sensory system. Therefore, it is unclear if it is the reliability of the dominant system that causes this increase in the integration windows, if both senses contribute, or if the size depends on the relative reliability of the auditory and visual stimuli. Studies that did investigate the effect of impairment in the non-dominant sense (Narinesingh et al., 2017; Richards et al., 2016) only did so in the temporal domain, where they found increased temporal integration windows in participants with a visual disorder, suggesting that deficits in the non-dominant sense can also affect the integration window. Similar experiments have not been performed in the spatial domain, so it is still unclear if a hearing loss affects the window of integration in the spatial domain, where it is the visual sensory system that generally dominates the integrated percept.

The impact of hearing impairment on auditory localization is limited. Hearing-impaired listeners have reported degraded spatial awareness and localization (Gatehouse and Noble, 2004; Olsen et al., 2012) even after compensation for reduced audibility (Glyde et al., 2013). However, as sensorineural hearing loss commonly causes reduced audibility particularly at high frequencies, the consequences of this hearing loss on spatial perception is mainly associated with degraded elevation acuity and increased susceptibility to front-back confusions (Häusler et al., 1983; Noble et al., 1994, 1997; Otte et al., 2013; Rakerd et al., 1998). In fact, when stimuli are presented from the front direction, a bilateral sensorineural hearing loss has been found to only mildly affect horizontal localization performance, with many hearing-impaired listeners (even those with poor speech discrimination performance) showing near-normal or even normal localization performance (Häusler et al., 1983; Lorenzi et al., 1999; Otte et al., 2013; Rakerd et al., 1998). Thus, effects of a hearing loss on the spatial integration window might be minor. However, indications of a potential detrimental effect do exist. Already at early stages of a (mild-to-moderate) hearing loss, evidence of cross-modal cortical reorganization has been found (Campbell and Sharma, 2014; Puschmann and Thiel, 2017), resulting in an increased susceptibility to cross-modal distractors (Puschmann et al., 2014). Moreover, an increased susceptibility to the ventriloquist effect in acute and severe (but not moderate) unilateral hearing-impaired listeners has been found (Venskytis et al., 2019). However, as pointed out by the authors, the visual acuity of the participants was not measured. Thus, it is not clear to what extent the increased susceptibility was due to their hearing impairment.

When exploring effects of hearing loss on the spatial integration window, age needs to be taken into account, as studies, especially in the temporal domain, have also found effects of age on the size of the integration windows (Bedard and Barnett-Cowan, 2016; De Boer-Schellekens and Vroomen, 2014; Diederich et al., 2008; Hernández et al., 2019; Scurry et al., 2020; Setti et al., 2011), for a review see Baum and Stevenson (2017). While the increased width of the temporal integration window with age has been assumed to reflect a decrease of the visual and auditory reliabilities, a large-scale study of 2920 participants by Hirst et al. (2019) found that, while sensory deficits were directly associated with an increased visual bias (measured in the temporal domain), they could not account for the full effect of age, i.e. the increased width of the temporal integration window was associated not only with age-related sensory deficits, but also just with age. The effect of age, however, has received much less attention in the spatial domain. Moreover, the two studies that investigated the effect of age on the spatial integration window showed conflicting results. Dobreva et al. (2012)

found an increased visual bias in older participants relative to young participants, both in terms of horizontal and elevation localization. However, Stawicki et al. (2019) did not find differences between young and older participants when investigating spatial integration windows in older normal-hearing adults with no or minimal sensory deficits. The difference in hearing abilities - the older adults in the study by Dobreva et al. (2012) had clinically normal age-matched thresholds, as opposed to normal-hearing thresholds as in Stawicki et al - may have contributed to the differences in results, but it is unclear which factors are mainly responsible for the discrepancies between the studies.

To investigate if the spatial integration window is increased in older and hearing-impaired listeners, eight young normal-hearing, six older normal-hearing and seven older hearing-impaired participants performed an adaptive left-right discrimination task, which measured the just noticeable difference (JND) in angle in four conditions and at five azimuthal angles. Two unimodal conditions (audio-only and visual-only) were considered to compare unimodal and bimodal performance within the groups and to compare the unimodal performance between the groups. The first bimodal condition presented the auditory and visual stimuli congruently and served to confirm benefits of audio-visual integration. The second bimodal condition presented the auditory and visual stimuli incongruently to get a measure of the AV integration window. Such incongruent stimulus presentation can result in response biases (Chen and Vroomen, 2013), where the participants respond to the visual stimulus instead of the audio-visual stimulus. To avoid these response biases, the visual stimuli were always presented from the same location, halfway in between the auditory stimuli, while only the position of the auditory stimuli was varied, as in Bertelson and Aschersleben (1998), Bertelson et al. (2000), Stawicki et al. (2019), and Vroomen et al. (2001). In this way, when the spatial separation between the auditory stimuli was small, the stimuli were more likely to be integrated, which then reduced the perceived distance between the auditory stimuli, as their perceived location was pulled towards the visual stimuli. Hence, integration should result in an increased JND. Only when the distance between the auditory stimuli is larger than the spatial integration window should participants be able to consistently perform the left-right discrimination task. An increased integration window should thus result in an increased threshold in the incongruent audio-visual condition.

To further confirm integration, reaction times were measured and tested against the race-model (Miller, 1982; Raab, 1962). This race-model assumes independent effects of the audio and visual stimuli on the reaction times. Integrated stimuli have processing times that are shorter than predicted by this race model. Therefore, invalidation of the race-model can be used as a further confirmation of integration in both the incongruent and congruent audio-visual condition.

The participants in the three groups were selected based on their age and hearing loss. For each participant, an audiogram and visual acuity (VA) were measured. The hypothesis was that, after taking into account the VA of the participants, age and hearing loss would increase the spatial integration window, such that OHI participants would show the largest JND in the audio-visual incongruent condition, followed by the ONH and the YNH participants.

6.2 Methods

6.2.1 Participants

Eight young normal hearing (YNH, six males, average age 25 ± 3), eight older normal hearing (ONH; one male, average age 62 ± 12) and eight hearing impaired

(OHI; six males, average age 74 ± 7) listeners were recruited from the Hearing Systems' database. Due to COVID-19 restrictions, participant recruitment was limited. As a result of this, an average age difference of 12 years between the ONH and OHI groups was allowed, instead of fully age-matching them. Data from participant 18 (ONH) were excluded from the analysis due to an issue with the eye-tracking system. Additionally, due to extreme outliers in the AV (congruent) performance data, compared to the unimodal data (i.e., congruent thresholds more than 10 times unimodal thresholds) and thresholds larger than 2 standard deviations above the group mean, the data from participant 8 (ONH) and 22 (OHI) were rejected. Eight participants remained for the ONH group, six participants (1, male, average age 64 ± 13) for the ONH group and seven (6 males, 73 ± 6.5) for the OHI group.

Auditory thresholds were measured at octave frequencies between 125 Hz and 8 kHz. Participants in the YNH group had all hearing thresholds at 20 dB hearing level (HL) or below. The ONH had no more than one 25 dB threshold between 125 and 6 kHz (all others, except the 8 kHz threshold needed to be 20 dB or below). The symmetric mild-to-moderate sloping hearing loss of the OHI group was classified based on the lowest root-mean-square error of the participants audiogram and the Bisgaard standard audiograms (Bisgaard, Vlaming, and Dahlquist, 2010), following van Beurden et al. (van Beurden et al., 2018). Three participants had a mild sloping (N2) hearing loss, the other three had moderate sloping (N3) hearing loss. The audiograms of the participants are shown in Fig. 6.1 (middle and right panel). The three groups (YNH, ONH, and OHI) are color coded. The horizontal dashed line indicates the 20 dB HL. As can be seen from Fig. 6.1, although the YNH and ONH are both (mostly) normal hearing, the YNH generally had better audiograms.

The visual acuity of the participants was measured using a logMAR visual chart. Participants that used corrective lenses during the experiment performed this test with these corrective lenses. Thus, the visual acuity ratings shown in Fig. 6.1 (left panel) are the corrected ratings. Again, the groups are color coded and the threshold for normal acuity is indicated as the black dashed line. Scores below this line indicate below-normal performance. As can be seen, several older participants and one YNH scored below normal in this task. The aim was to have enough participants, such that VA could be factored out. The VA did not vary significantly across groups [$F_{2,18} = 2.627, p = 0.0998$].

The experimental procedure was approved by the Science-Ethics Committee for the Capital Region of Denmark (H-16036391) and all participants provided written informed consent.

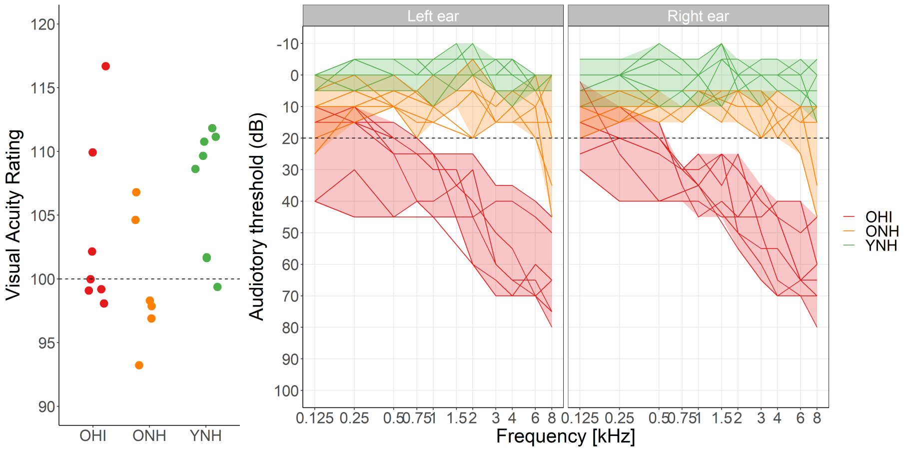


Figure 6.1: Visual acuity and audiograms. The groups are colour coded green (YNH), orange (ONH) and red (OHI). The dotted black lines show the normal thresholds. The left panel shows the visual acuity for each participant. The right panel shows the audiogram for the left and right ears. Each line shows an audiogram, while the coloured area indicated the range of thresholds measured at each value. Some, but not all, participants were measured at two additional frequencies, namely 750 Hz and 1500 Hz. These measurements are included in the graph and the outlined area, which is why at those frequencies the audiogram of a participant is outside the coloured area (since for this person there are no results at those frequencies).

6.2.2 Apparatus

A loudspeaker array, 4.8m in diameter, containing 24 equidistantly spaced loudspeakers (KEF LS50) was used to produce the auditory stimuli. The auditory stimuli were generated in MATLAB (The Mathworks, Natick, MA) and reproduced and moved around using 11th order, horizontal only ambisonics reproduction. The audio signals were sent to the amplifiers (Sonible GmbH, Graz, Austria) that drive the loudspeakers via a TESIRA biamp DSPs with TESIRA SOC-4 audio DSP cards (biamp Systems, Beaverton, OR).

Visual information was created and displayed in UNITY3D (Unity Technologies, San Francisco, CA) and presented to the participants using an HMD (HTC Vive Pro Eye VR system; HTC Corporation, New Taipei City, Taiwan). The visual environment was a simplified version of the experimental room, without loudspeaker array or foam wedges. The virtual environment was calibrated using three HTC VIVE trackers, which repositioned the virtual environment if a discrepancy larger than 1 cm occurred (for details, see Ahrens et al. (2019). The two computers running the sound system and the virtual environment communicated via an OSC server.

Participants used two handheld HTC VIVE controllers to respond and proceed through the experiment. Participants could respond ‘left’ and ‘right’ by pressing a button on either the left or the right controller.

6.2.3 Stimuli

The auditory stimuli consisted of a 10-Hz pink noise burst train, containing three short noise bursts 22 ms in length, with 5 ms on/off ramps. These stimuli were generated using MATLAB (The Mathworks, Natick, MA) at a sound pressure level (SPL) of 80 dB. This level was roved for each individual noise burst (uniform distribution from ±1 dB). The stimuli were spatialized using 11th order ambisonics in the horizontal plane. The visual stimuli were 10Hz flash trains. Each light was shown for 2 frames on the HMD, refreshing at 90Hz, corresponding to 22ms. The stimuli, created in UNITY, were warm-white (#F1EED4) point light sources, with a range of 3 cm and an intensity of 0.95 presented at approximately eye height at a distance of 2.4m. The stimulus design is shown in Fig. 6.2.

The stimuli were presented around ±30°, ±15° and 0° azimuth. Each trial introduced a small jitter (uniform distribution between ±5°) to ensure that participants would not remember specific locations. Similarly, the pause in between the two subsequent stimuli trains was uniformly distributed between 0.3 and 0.5 seconds.

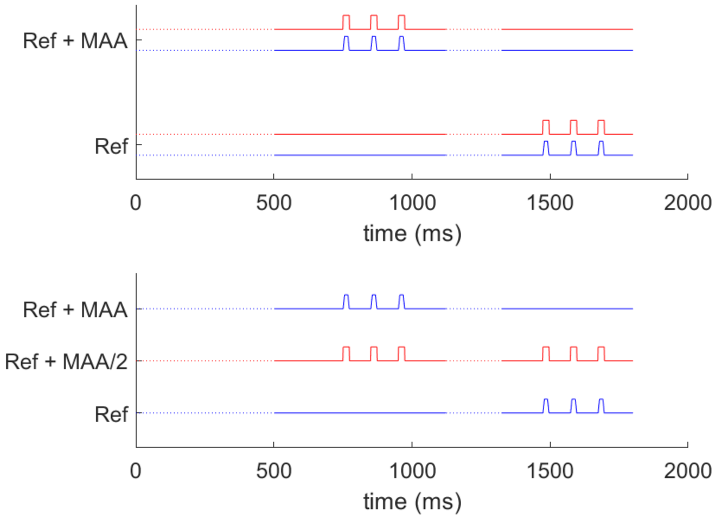


Figure 6.2: The timing and positioning of the auditory (blue) and visual stimuli (red). Dotted lines indicate jitter in the timing. These timings varied on each trial. The solid lines indicate fixed timing. A congruent stimulus presentation is shown in the upper panel, with both the auditory and visual stimulus moving. The lower panel shows an incongruent presentation. While the auditory stimulus moves, the visual stimulus is presented twice at the same location. This ensures that participants cannot rely on visual information to complete the task.

6.2.4 Experimental conditions

The experiment consisted of four conditions: audio, visual, audio-visual congruent and audio-visual incongruent. The experimental runs were divided into three blocks. The unimodal conditions were measured in separate blocks, while the audio-visual congruent and incongruent conditions were combined into a single block. This was done to reduce adaptation to incongruencies. The experiment was split up into two sessions, one measuring the unimodal blocks and the other measuring the audio-visual block. The order of these sessions and the order of the unimodal blocks within the sessions were counterbalanced across participants.

The experiment was a left-right discrimination task; participants were asked to indicate in each trial if the second of two stimulus presentations occurred to the left or right of the first. For each angle and condition, the minimum detectable angle was determined through an adaptive procedure. At the start of a run (1 run results in a threshold estimate for 1 condition at 1 position) the step size was 10°. Using a 1-up 2-down tracking rule Levitt (1971), this step size was varied by multiplying or dividing it by a pre-set coefficient. After each reversal, this coefficient was decreased such that the adaptive procedure narrowed in on the threshold of 70.7% correct. This coefficient started at 2 and exponentially decayed to 1.1. An example of the adaptive procedure is shown in Fig. 6.3

Within a block, the positions were presented interleaved, with separate adaptive tracking for the different positions. A block was completed once all runs had reached 9 reversals. Until this end of the block, each set of 20 trials was generated with 4 trials per position, which, in the audio-visual blocks, were split such that 2 trials were congruent, and 2 trials were incongruent at each position. Positions and conditions where 9 reversals had already been reached continued the adaptive tracking. This was done to ensure that, throughout the entire experiment, participants could not anticipate the position or condition of the stimuli.

The participants were instructed to respond as fast and accurately as possible and to look forward during trials. To verify that participants were indeed not moving their eyes during the trials, an eye tracker was used to initiate each trial (participants were asked to look at the focus point 2.4m in front of them at eye height before each trial) and track their gaze during the trial. If their gaze left the focus point during a trial or an eye blink was detected, the trial was considered invalid, and the adaptive procedure was not updated. Note that the focus point disappeared (visually) at the start of a trial and only reappeared for the participants after they had indicated their response. The start of stimuli presentation in each trial was jittered by 0.25 to 0.75 seconds to remove anticipatory effects. At the start of each block, participants performed at least 20 training trials to ensure they correctly understood the task and to reduce any training effects.

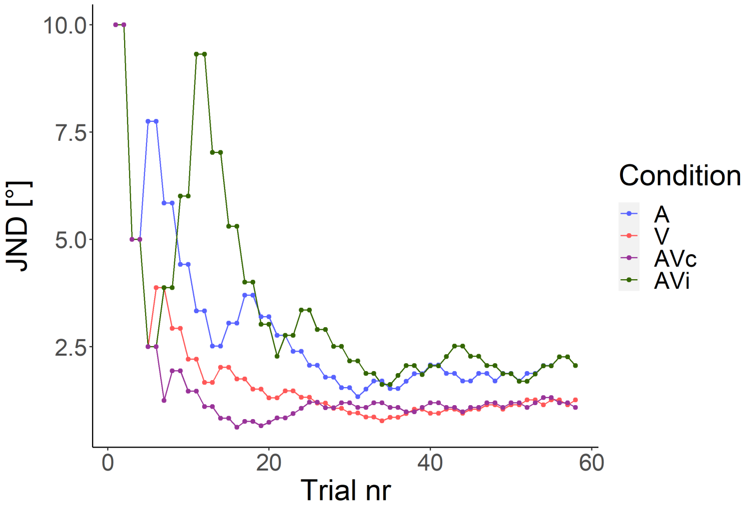


Figure 6.3: The adaptive track for participant 21 at 15° azimuth. The four conditions are indicated in blue (auditory), red (visual), purple (congruent) and green (incongruent). This participant did not show much susceptibility to the ventriloquist effect, as can be seen from the similar incongruent and auditory only data.

6.2.5 Analysis

For the analysis, the JND was calculated per angle, condition, and participant as the mean JND at the last 6 reversals. An ANOVA was performed using the statistical software R (R Core Team, 2020) to investigate how VA, angle, condition, and participant group affected the JND and the reactions time. Bonferroni corrected post-hoc analyses of within factor comparisons were performed to compare the effect of the various angles, groups, and conditions. Finally, the reaction times were also tested against a race-model as in Ulrich et al. (2007) to ascertain integration.

6.3 Results

Fig. 6.4 shows the average JND per angle for the four conditions. The data obtained with the OHI participants are shown in the left panel, the ONH data are shown in the middle panel and the YNH data are represented in the right panel. It can be seen that thresholds varied with condition [$F_{3,348} = 59.530, p < 0.0001$], but that the same general trends were found in all participant groups. The obtained thresholds were generally smallest in the audio-visual (purple) or visual-only (red) condition, increased by about 1-2° in the audio-only (blue) condition and then increased further in the audio-visual incongruent (green) condition (the amount of which varied per angle). The post-hoc comparison confirmed these differences between the conditions. When averaged across the other factors, the thresholds in the audio-visual congruent condition were best, although visual performance was not significantly worse [$t_{348} = 0.014, p = 1.00$]. The thresholds in these two conditions were lower than those in the audio-only condition [A – AV: $t_{348} = 3.991, p = 0.0005$; A – V: $t(348) = 4.005, p = 0.0005$] and, finally, audio-visual incongruent thresholds were significantly increased compared to all other conditions, as can be seen in Fig. 6.4 [A – AVi: $t_{348} = -7.494, p < 0.001$; V – AVi: $t_{348} = -11.485, p < 0.001$; AVc – AVi: $t_{348} = -11.499, p < 0.001$].

However, the effect of condition varied with other factors. A significant interaction between the effect of condition and angle [$F_{6,348} = 8.108, p < 0.0001$] was found and a three-way interaction between the effect of condition, angle and group [$F_{12,348} = 1.983, p = 0.0249$] was found. When split by angle and group, the difference between audio-visual incongruent and audio-only was significant only at 0° azimuth for the OHI [$t_{348} = -6.490, p < 0.0001$] and ONH groups [$t_{348} = -4.895, p = 0.0001$], but not for YNH group [$t_{348} = -1.059, p = 1.00$]. At 15°, this difference was no longer significant for any of the groups [OHI: $t_{348} = -2.386, p = 1.00$, ONH: $t_{348} = -0.629, p = 1.00$; YNH: $t_{348} = -2.700, p = 0.4806$] and this was also the case at 30° [OHI: $t(348) = -1.384, p = 1.00$, ONH: $t_{348} = -0.921, p = 1.00$; YNH: $t_{348} = -1.169, p = 1.00$]. Moreover, while there were significantly increased thresholds for the audio condition compared to the audio-visual and visual conditions, on average, the unimodal and AV congruent conditions did not differ significantly at any angle or any group in this three-way interaction analysis.

The main effect of angle was also significant [$F_{2,348} = 5.470, p = 0.0048$]. While there was a trend of increasing JND with increasing eccentricity in both the unimodal and the congruent audio-visual condition, this effect of angle was only found to be significant in the visual condition [$t_{348} = -2.570, p = 0.0318$]. As for the incongruent audio-visual condition, the largest JND was found quite consistently at 0° azimuth (the deviation in the YNH group was due to an extreme outlier at 15°). It decreased sharply at ±15°, approaching the audio condition, and remained similar or increased slightly again at ±30°. The analysis revealed that these changes with angle were significant, as an increased threshold at

0° was found when comparing to both 15° [$t_{348} = 5.560, p < 0.0001$] and 30° [$t_{348} = 4.976, p < 0.0001$] azimuth.

The effect of angle also varied per group [$F_{4,348} = 3.520, p = 0.0078$]. Specifically, the thresholds of the OHI group increased more with angle than the other two groups [OHI – ONH: $t_{348} = 3.2789, p = 0.0034$; OHI – YNH: $t_{348} = 3.234, p = 0.0040$]. This can be seen in Fig. 6.4, where the 'v' shape of the unimodal conditions was steeper in the OHI group than in the ONH and YNH group.

A significant difference in the thresholds between groups was found [$F_{2,348} = 7.511, p < 0.0001$]. Considering the best thresholds, 0.284° ± 0.210°, 0.209° ± 0.149° and 0.225° ± 0.216°, respectively, for OHI, ONH and YNH, results were very close to each other. However, in the incongruent audio-visual condition, the average thresholds differed significantly, with the average thresholds increasing up to 6.65° ± 4.37°, 4.87° ± 3.37° and 3.00° ± 1.15° for the OHI, ONH and YNH groups, respectively. As mentioned before, for YNH this was not a significant increase compared to the audio-only performance. Regarding the incongruent AV condition, the analysis revealed an effect of age [ONH – YNH: $t_{348} = 3.681, p = 0.0178$; OHI – YNH:$t_{348} = 4.970, p = 0.0001$], but no effect of hearing loss [OHI – ONH: $t_{348} = 0.918, p = 1.0000$]. At other angles, no differences were found between conditions.

Fig. 6.5 shows the effect of the VA of the participants on the JND for 0° azimuth (left panel), ±15° (middle panel) and ±30° (right panel). JNDs generally decreased with increasing VA [$F_{1,348} = 27.480, p < 0.0001$]. As a significant interaction between the angle and the effect of VA was found [$F_{12,348} = 1.983, p = 0.0249$], the results were analyzed separately in terms of angle instead of group. A linear fit of the results per angle and per condition illustrates the relation between VA and JND. At 0°, the results in the unimodal and audio-visual congruent conditions were not affected by VA. For the audio-visual incongruent a negative

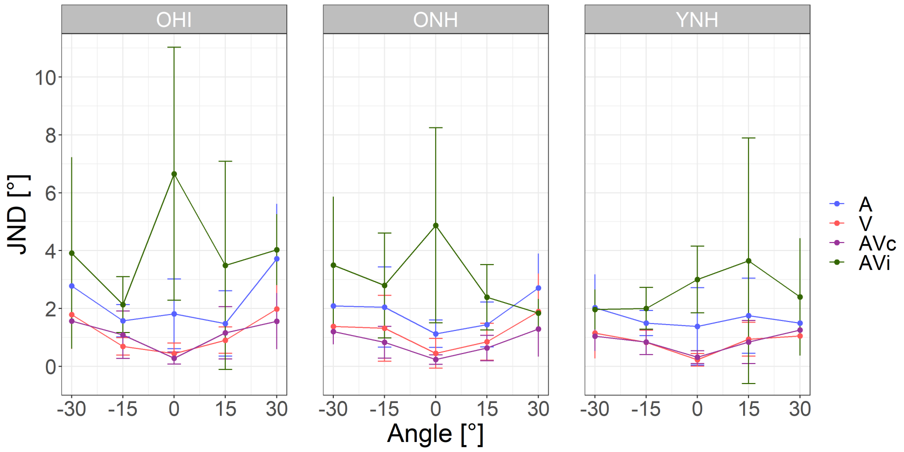


Figure 6.4: Average JND as a function of angle, split per group. Audio (blue), visual (red), audio-visual congruent (AVc) and audio-visual incongruent (AVi) JNDs are plotted per angle for OHI (left) ONH (middle) and YNH (right).

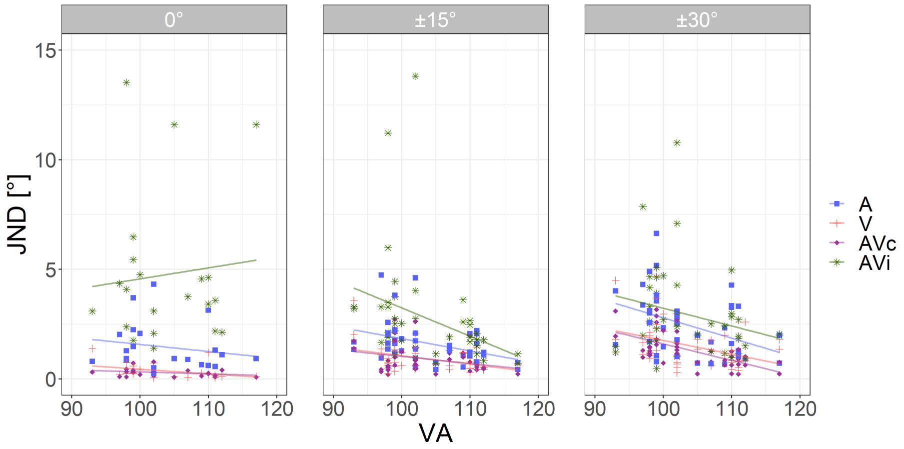


Figure 6.5: Average JND as a function of VA, for 0°, ±15° and ±30°. The coloured lines show the linear fit per condition. The slope of the curves increases with angle for the unimodal and audio-visual congruent condition. Each point in the plot corresponds to individual thresholds.

trend can be seen at 0° azimuth. However, with increasing angle, the slope of the linear fit steepened in all conditions.

Fig. 6.6 shows the reaction times. A large variation of the reaction times was found and no consistent pattern of the reaction time per condition [$F_{3,348} = 1.673 p = 0.1726$] or angle [$F_{2,348} = 0.018, p = 0.9826$] was found. There was, however, a significant main effect of group [$F_{2,348} = 17.765, p < 0.001$]. The

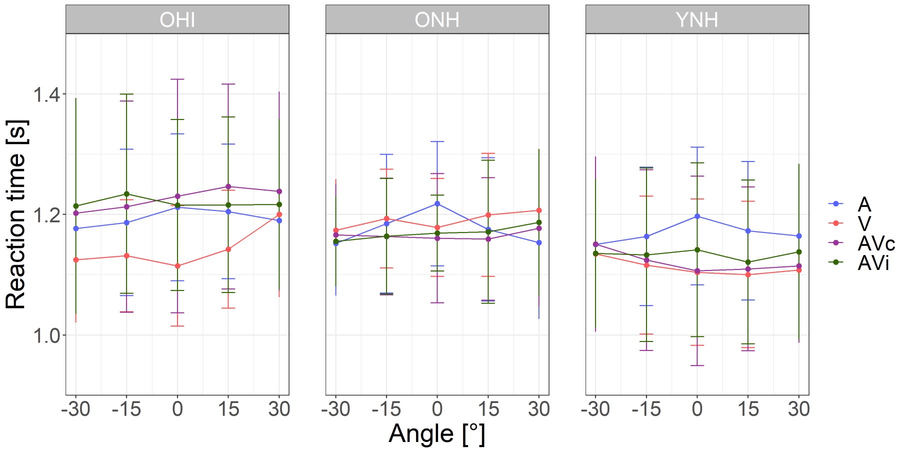


Figure 6.6: Average reaction times per angle, group and condition. The average reaction time was calculated as the average of the participants average reaction times.

average reaction time was largest for the OHI group (1.19 ± 0.26 s), followed by the ONH group (1.18 ± 0.23 s) and then the YNH group (1.13 ± 0.21 s). This difference between the YNH and OHI group [$t_{348} = 3.436, p = 0.0020$] and between the YNH and ONH groups [$t_{348} = 3.882, p = 0.0004$] was significant, whereas the difference in reaction time between the OHI and the ONH groups [$t_{348} = -1.067, p = 0.8597$] was not, demonstrating again an effect of age but not hearing loss.

Since a main effect of VA [$F_{1,348} = 7.893, p = 0.0052$] and a significant interaction between VA and group was found [$t_{348} = 9.889, p < 0.0001$], Fig. 6.7 shows the reaction times as a function of VA split by group. As indicated by the linear fits, the reaction times generally increased with VA for both the OHI (left panel) and ONH (middle panel) groups. The YNH group (right panel) was unaffected by VA.

There was no consistent decrease in reaction times in the AV conditions, as compared to the unimodal conditions. Indeed, while for some participants evidence for integration was found, the pooled responses did not invalidate race model for separate processes .

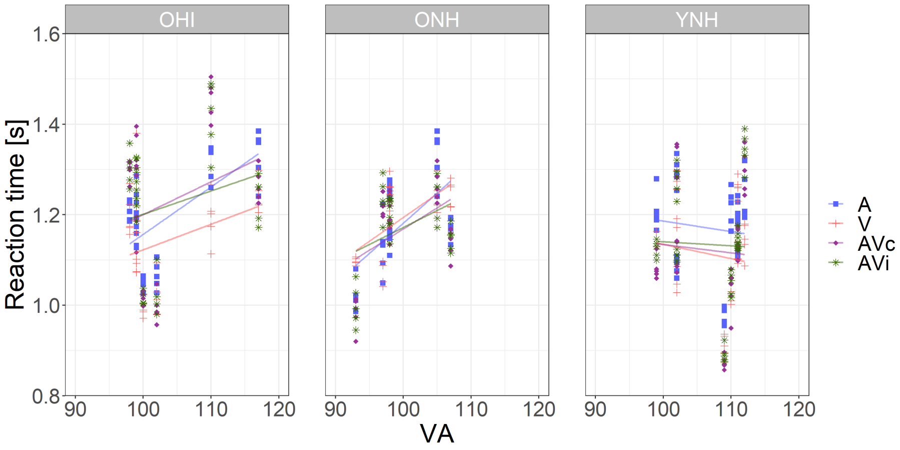


Figure 6.7: Average reaction times as a function of VA. Data from all angles are included. The average reaction time was calculated as the average of the participants average reaction times per angle and condition. The coloured lines show the linear fit per condition.

6.4 Discussion

The present study investigated if age and hearing loss affect the spatial integration windows, using a left-right discrimination task with unimodal audio and visual, and bimodal congruent and incongruent conditions. The results showed that the integration window increased with age, but, although the average JND in the audio-visual incongruent condition was larger for the OHI compared to the ONH, this difference was not consistent enough to conclude an effect of hearing loss. Similarly, the average auditory JND was also larger, on average, for the OHI compared to the other groups, but this difference was not significant, nor any other unimodal comparison, potentially due to the low number of participants per group.

While an effect of age has been previously reported in the temporal domain, results in the spatial domain were conflicting. While Dobreva et al. (2012) found evidence for an effect of age, Stawicki et al. (2019) did not. In the present study, controlling for both auditory and visual sensory deficits, the spatial integration window was found to increase with age. It is possible that the discrepancy in the results of the present study and Dobreva et al. (2012) versus the results presented in Stawicki et al. (2019) is due to the sensory deficits, as Stawicki et al. (2019) recruited adults without any auditory or visual sensory deficit. However, as mentioned in the introduction it has been shown that sensory deficits cannot not account for the full effect of age (Hirst et al., 2019). Moreover, Stawicki et al. (2019) appear to have measured audiograms up to 6 kHz (no frequency labels given to verify) and all but one participant in the present study had normal hearing up to 6 kHz. In the present study participants were not excluded based on visual deficits, but Dobreva et al. (2012) also excluded participants with visual impairments and still found an effect of age (although they were more lenient with the audiograms). Thus, differences in sensory deficits are unlikely to be the cause of the different pattern of results across studies. Another possible explanation for the different results is potential differences in the cognitive health of the participants. Cognitive impairments and disorders have been linked to increases in the spatial window (Chan et al., 2015; Foss-Feig et al., 2010; Haß et al., 2017). However, none of the experiments, including the present study, tested the cognitive skills abilities of the participants.

The increase in the spatial window mirrors the increase of the temporal integration window, which suggest that shifts in the audio-visual processing occur with age in both domains. The increased windows suggests that older adults, regardless of sensory deficits, will likely be more susceptible to integrating incongruent information, making it more difficult to ignore irrelevant/distracting information. How this further affects older adults is unclear. Despite the increased integration windows found here and in other studies (Bedard and Barnett-Cowan, 2016; De Boer-Schellekens and Vroomen, 2014; Diederich et al., 2008; Dobreva et al., 2012; Hernández et al., 2019; Scurry et al., 2020), and despite enhanced benefits from integration (Laurienti et al., 2006; Peiffer et al., 2007), some studies found that older adults were less likely to integration integrate compared to younger adults (Diederich et al., 2008; Scurry et al., 2020). On the other hand, other studies found opposite results, with increased integration in older adults (DeLoss et al., 2013; Dobreva et al., 2012; Setti et al., 2011). Either way, it is clear that audio-visual integration changes as we age and this affects elderly, as for example a link has been suggested between inefficient multisensory integration and an increased susceptibility to falling (Setti et al., 2011).

Interestingly, the effect of age that was found in the present study was only significant at 0° azimuth. At increased eccentricities, thresholds in the audio-visual incongruent condition rapidly approached audio-only performance, especially in the YNH group. The ventriloquist effect has previously been shown to decrease with angle (Charbonneau et al., 2013; Hairston et al., 2003), however, the shift was not quite as strong as found here. Charbonneau et al. (2013) showed that the decrease of the ventriloquist effect with eccentricity paralleled the lowering of the relative reliability of the unimodal stimuli. In the present study, however, this shift in relative reliability was not found. This is probably the case because thresholds were measured at lower eccentricities, compared to Charbonneau et al. Thus, eccentricity related lowering of localization reliability cannot explain the sharp decrease in the susceptibility to the ventriloquist effect in the present study. Their setup, however, used a fixed audio-visual disparity of 15° and did not eliminate a response bias. Perhaps this difference in methods is what caused the difference in the effect of eccentricity, or maybe, as with age, there is an effect of eccentricity by itself.

In line with previous studies (e.g., DeLoss et al., 2013; Diederich et al., 2008; Dobreva et al., 2012; Laurienti et al., 2006; Mahoney et al., 2011) older adults had slower reaction times. However, while some participants did show reaction times in the audio-visual tasks that were consistent with audio-visual integration, this was not found at the group level. The average reaction times per condition, as shown in Fig. 6.6, also show this clearly as there is no consistent reduction in the reaction time in the audio-visual congruent compared to the unimodal conditions in the ONH and OHI groups. It is possible that this lack of improved reaction times is the results of the trade-off between reaction speed and integration found in older adults (Jones et al., 2019).

While an effect of age and visual acuity was found, an effect of hearing loss cannot be concluded based on the current data. There was a trend towards an increased spatial integration window, on average, for the OHI group, at most angles, but this was not significant. If hearing loss affects the spatial integration window then it does so to a much lesser extent than age and visual acuity. This is consistent with what would be predicted based on the audio-only JNDs, which also were not significantly increased in the OHI compared to the ONH group, although the average audio-only JNDs were generally highest in this group. The difference in the audio-only JND's between the OHI and ONH group match well with other studies, which found an increase of about 1-2 degrees. This limited increase might, however, not accurately represent the difficulties that hearing impaired face in spatial localization, as results between localization and left-right discrimination tasks have been found to vary.

Since results were gathered in the midst of the corona pandemic, the number of participants that could be recruited for this study was limited. It is possible that with more participants, the differences between the OHI and ONH groups would become significant. The limitation of the low number of participants can also be seen in the effect of VA. The results show that participants with better VA generally showed a smaller JND in all conditions, including the audio-only condition where VA was not expected to affect the results. Moreover, in the OHI group, a positive correlation between the detection thresholds and the VA was found as outliers at the outer extreme ranges highly influenced the linear regression. Similarly, regarding the reaction times, an effect of VA was found.

Overall, older hearing-impaired listeners were found to have increased integration windows in the spatial domain. This increased integration window was only found at 0° azimuth, suggesting that, as with age, there might be an effect of eccentricity beyond the shift in sensory reliability. While congruence judgement tasks have been used to map how congruence judgement changes with eccentricity, elevation and spatial disparity, a similar effort with the spatial integration could provide a very interesting comparison and could explore how eccentricity affects integration. Finally, while the present study did not find a significant effect of hearing loss, future studies may investigate effects of hearing loss in experimental settings and paradigms that more closely resemble real world settings.

6.5 Conclusion

The present study investigated how age and hearing loss affect the spatial integration window. In line with previous studies in the temporal domain, an increase in the spatial window with age was found. This difference between the young and older normal hearing listeners was only significant at 0-degrees azimuth. When the stimuli were presented at increased eccentricities, the audio-visual incongruent performance rapidly approached the audio-only performance, and no significant differences were found. In addition to an effect of age, an effect of the visual acuity of the participants was found. However, no significant difference between the older normal hearing- and hearing-impaired listeners was found. Trends in the data were promising and future studies could reassess the effects of hearing loss and could further explore what causes the increased size of integration windows with age.

7

Investigating target search with audio, visual, or audio-visual information in virtual reality

Abstract

The auditory system is thought to guide visual localization. In this study, we investigated if indeed audio-visual localization can be explained by a combination of auditory guidance and visual localization and how this is affected by auditory distractors.

To study audio-visual localization behavior and how it changes with increasing auditory distractors, seven normal-hearing listeners participated in an auditory, visual and audio-visual search task with a varying number of auditory distractors (0, 1, 2, 3, 5, 7 or 11). Participants were presented with an audio-only, visual-only or audio-visual target, which was then moved to an unknown position somewhere around them. Their task was then to find this target as quickly as possible. Behavioral features, such as eye-gaze and head-rotation, were tracked and analyzed.

The results demonstrated that when the number of auditory distractors was lower than seven, audio-visual area localization (the time it took to get the target within the field of view) could be well explained by auditory area localization. However, for seven and eleven

distractors, audio-visual area localization times were smaller than those in the audio-only conditions. Head-motion data was similar in the audio-visual and auditory conditions when the number of auditory distractors was low, however, as the number of distractors increased, participants rotated their heads into the wrong direction more often. The distribution of these data showed similarities to those both in the auditory and in the visual condition.

The target localization time (the time it took to localize the target when it was already within the field of view) was consistently faster in the audio-visual than the visual-only target localization times, as well as the audio-only localization times.

Together, the head-motion data and localization times supports a more complex interaction between the auditory and visual systems when localizing stimuli than the traditionally held view that the auditory system guides visual localization. Further, this interaction is influenced by the number of auditory distractors that are present.

7.1 Introduction

The visual system derives the location of objects from the position where light hits the retina, which is highly accurate in the fovea, but localization accuracy decreases steeply towards the periphery (van Opstal, 2016). On the other hand, the auditory system is less accurate, but it is not limited to one direction because it uses interaural time and level differences of the sound, and pinnae cues to deduce the location of the sound source, see Blauert (1997) for a review. Further, although localization performance of the auditory system degrades at increased angles, the decay of localization acuity with angle is less steep than for vision (Freeman et al., 2018). Together, these systems are assumed to work in a complementary manner where auditory localization guides the visual system (Heffner and Heffner, 2016; Heffner and Heffner, 1992; Perrott et al., 1990).

Indeed, many studies have shown faster visual target acquisition when auditory cues have been presented (e.g., Begault and Pittman, 1996; Bouchara et al., 2013; McIntire et al., 2010; Tannen et al., 2004; Tannen et al., 2000; Vu et al., 2006). Differences in search strategies, as indicated by a change in the participants head motion, have also been observed when audio cues are available (Nelson et al., 1998; Perrott et al., 1990). Further, when auditory information was presented, Perrott et al. (1990) observed participants close their eyes during the majority of trials. In Nelson et al. (1998) participants initially moved their heads in a circular search pattern (scanning the entire field) when only visual or non-localized audio was presented, but this switched to a much more swift and direct movement when the auditory stimulus was localized. Interestingly, in both the spatialized and non-spatialized audio, there was a delay before participants started moving their heads, which was not present in the visual only trials (Nelson et al., 1998).

Moreover, spatialized auditory information also strongly decreased the effect of visual distractors Bolia et al. (1999), Perrott et al. (1991), and Rudmann and Strybel (1999). On the other hand, auditory distractors very quickly degraded visual search; While visual search times were improved when two or fewer distractor sound sources were presented, more auditory distractors degraded visual search compared to the no-audio condition (Brungart et al., 2010). While this decrease in benefit and even hindrance of visual target search with increasing number of auditory distractors has been shown in reaction times, it is unclear how it affects localization behavior.

To study how audio-visual localization behavior changes with increasing auditory distractors, we investigated behavioral patterns in normal-hearing listeners in a realistic audio-visual search task with a varying number of auditory distractors. Immediately prior to a trial, participants were presented with an auditory-only, visual-only or audio-visual target. During the trial, this target was presented at a random position somewhere around the participant. The participants' task was to find this target as fast as possible. Behavioral features, such as eye-gaze and head-rotation were tracked and analyzed.

Assuming that the auditory system is responsible for guiding the visual system (e.g., Heffner and Heffner, 2016; Heffner and Heffner, 1992; Perrott et al., 1990), the way in which participants approach the task would be to use the auditory system to find the target area (so that the head can be moved to bring the target within the field of view) and then rely on both systems (although predominantly the visual system) to fully locate the target. To confirm this, tracking data was used to estimate the time it took participants to find the target area and the time it took to fully locate the target. Moreover, a varying number of distractors were presented to see how localization behavior changes when auditory localization becomes more difficult. It was hypothesized that the time it takes participants to find an audio-visual target can be explained by a combination of the auditory area search time (to guide the visual system) and the visual target search time (once the target is within the field of view). However, as the number of distractors decreases, participants might shift to more visual-only localization behavior.

7.2 Methods

7.2.1 Participants

Seven self-reported normal-hearing and (corrected-to) normal-seeing students at the Technical University of Denmark participated in the experiment. The participants provided written informed consent and received an hourly compensation of 126 DKK for their participation. The experimental procedure was approved by the Science-Ethics Committee for the Capital Region of Denmark (H-16036391).

7.2.2 Apparatus

The auditory stimuli were presented to the participants via 24 loudspeakers (KEF LS50), placed in a 4.8m diameter ring with the listeners placed in the center. The loudspeakers received the audio signals from a sonible d:24 amplifier (Sonible GmbH, Graz, Austria) via TESIRA biamp DSPs with TESIRA SOC-4 digital-to-analog converters (biamp Systems, Beaverton, OR). For the presentation of the visual stimuli, an HTC Vive Pro Eye head mounted display (HMD) was used (HTC Corporation, New Taipei City, Taiwan). The visual stimuli were presented on white blocks located at the real-world positions of the loudspeakers as shown in Fig. 7.1. These white blocks were also the response options for the participants. Participants could indicate the perceived location of the target by pointing at a block and pressing a button on a handheld HTC Vive controller. Visual feedback was given in the form of a virtual laser pointer. The virtual environment was almost an exact 1:1 model of the experimental room, with the exception of the white blocks in the virtual world replacing the loudspeakers in the real world. The entire visual environment was created and displayed in UNITY3D (Unity Technologies, San Francisco, CA) and calibrated to the real world using HTC

Vive trackers (Ahrens et al., 2019). An example of the visual environment and the visual stimuli is shown in Fig. 7.1.

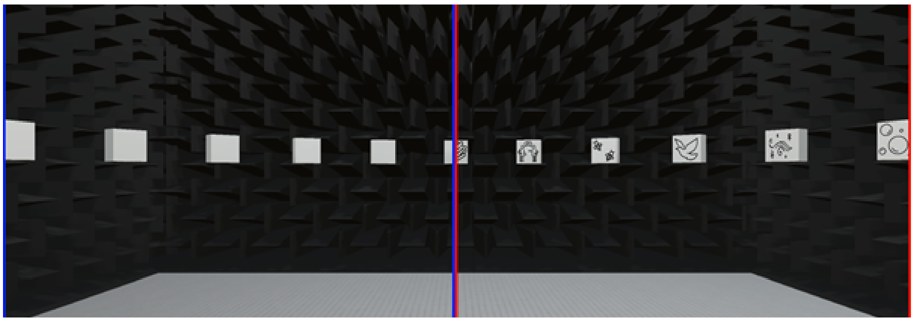


Figure 7.1: The environment as presented via the virtual reality glasses. The visual information in the auditory condition (no visual stimuli active) is shown on the left (blue outlined) and the environment with visual stimuli active is shown on the right (red outlined). Each box corresponds to a loudspeaker location in the real world. Matching auditory stimuli were played at the target and distractor locations.

The visual environment and the audio system were controlled by separate computers with commands from MATLAB to update the visual stimuli and responses from participants being sent back and forth between these two computers via an OSC server. Besides response data, motion tracking data from the participant, namely eye-gaze data, head-position and controller position data were collected with a frame-rate of approximately 90 Hz.

7.2.3 Stimuli

24 stimuli were used in the experiment. These stimuli were selected from a collection of sound textures (freesound.org). The auditory stimuli were chosen to be subjectively unique, recognizable, and repeatable over time. For each stimulus, the audio file was repeated and concatenated to increase the duration to 5 minutes. A-weighted loudness equalization was used to present all stimuli at approximately 70 dB SPL. Matching black and white visual icons from the Noun Project (www.thenounproject.com) were matched to the acoustic stimuli to create 24 audio-visual stimuli, e.g., a clock, bees, horses, blizzard, a ringing

telephone, fire, clapping.

7.2.4 Procedure

The participants in the experiment were instructed to find the target as quickly as possible. The target localization task was specifically designed to be a relatively easy task, as the interest was in how participants found the target, rather than if they could find it. At the start of a trial, the target sound was played for 3 seconds such that the participant knew which stimulus was the target for that trial. If the trial was an audio-visual trial, then the icon was also displayed during these 3 seconds. After these 3 seconds, the screen went black and all visual blocks, corresponding to the 24 stimuli, were shuffled. Then, participants had to find the target, point the handheld controller on the block positioned at that location and press a button.

Three target modalities were investigated: audio-only, visual-only or audio-visual. In the case of an audio-only target, all boxes, which indicated the potential target locations, remained blank (see Fig. 7.1, left). On each trial, one of the 24 stimuli was randomly assigned as the target for that trial. A subset of the remaining 23 stimuli was randomly chosen to be the auditory distractors in the audio-only and audio-visual conditions, whereas in the visual-only and audio-visual conditions all 24 visual stimuli were presented (i.e., 23 visual distractors). The acoustic stimuli continued playing for the entire duration of the trial until the participant made a localization decision or until the maximum duration of 5 minutes was reached. In the audio-only and the audio-visual conditions, the number of auditory distractors was varied with either 0, 1, 2, 3, 5, 7 or 11 distractors, spaced equidistantly from the target. In the audio-visual conditions, 23 visual distractors were always present. The position of the target was varied per trial from -120° (left of the listener) to 120° (right of the listener) in 15° steps. Each location was repeated 5 times in each condition. The target stimulus was also varied per trial. While the target selection was balanced, in that each stimulus was presented the same number of times, the combination of angle and stimulus was randomized. Thus, each condition (audio-only and audio-visual each presented with 0, 1, 2, 3, 5, 7 or 11 distractors, and visual-only) contained 85 trials, for a total of 1275 trials per participant.

At the start of the experiment, participants were informed that they would be shown a target and then would have to find this specific stimulus as quickly as possible. They were also informed about the horizontal intervals where the target could (-120° to 120°) and could not appear (-135° to -180° and 135° to 180°) and that stimuli would always be congruent in the audio-visual conditions. How they completed the task was left completely up to the participants.

7.2.5 Analysis

Since eye- and head-tracking data were collected, both localization performance and localization behaviour were analyzed. Mixed linear models were fitted to the localization error and the response times using the statistical software R (R Core Team, 2020) and the "lmerTest" package (Kuznetsova et al., 2017). FFor the localization data the absolute target angle, the number of auditory distractors and the target modality (audio-only, visual-only or audio-visual) were included as fixed effects, while the target stimulus, participants and repetitions were included as random effects. Post-hoc analyses of within factor comparisons were performed to examine the effect of different conditions and the number of distractors. In this analysis, two trials where the response time exceeded one minute were excluded.

Summary statistics were calculated from the tracking data. For this tracking data, only correct responses were analyzed. Data from 539 trials, corresponding to 6% of the total number of trials, were not included in this analysis. Another trial was removed from this analysis as the search times exceeded one minute.

First, two intervals were calculated based on the eye-gaze data; the time to find the approximate target area and the time to locate the target within this area. These intervals will be referred to as the 'area localization time' and the 'target localization time'. The area localization time was defined as the time it took participants to orient their eyes to within ±22.5° around the target. This specific range was based on the average response times of the participants, which greatly increased between ±15° and ±30° degrees, suggesting that the field of view of the participants was between ±15° and ±30°. The target localization time was defined as the time it took participants to identify the exact position after finding the right approximate position. This was defined as the time in between orienting their eyes to the right area and clicking on the target.

In addition to the eye-gaze data, we also analyzed head-rotations. As the head-rotation was less variable than the eye-gaze, the head-motion was used to investigate how often participants initially searched into the wrong direction and how far they searched into the wrong direction before turning towards the target location. These data were also analyzed with a mixed linear model with the same factors: stimulus angle, the number of distractors and the condition (audio-only, visual-only or audio-visual) as fixed effects and the target stimulus, participants and repetitions as random effects.

7.3 Results

7.3.1 Localization error

Fig. 7.2 shows the localization error for the auditory (left panel), visual (right panel) and audio-visual (middle panel) target localization as a function of the absolute target angle. The data are grouped by the number of auditory distractors. Progressively darker colours are used to indicate an increasing number of distractors. Data from the left hemisphere were mirrored. Participants correctly identified the target in the majority of the trials, especially in the visual-only and audio-visual conditions where only 5 and 11 incorrect trials occurred, respectively (see red and purple dots in Fig. 7.2). However, in the audio-only target localization conditions, as the angle and number of distractors increased, the number of incorrect trials also increased. Furthermore, the size of the error in these incorrect trials also increased with the number of auditory distractors. When less than three distractors were included, errors were concentrated between ±15°, but as the number of distractors increased, the error increased up to ±180°. It can also be seen that, when more than three auditory distractors were present, the localization error also varied with target angle in the audio-only conditions. Errors not only became more likely, but also larger as the targets occurred at increased angles.

The statistical analysis confirmed these results. A main effect of target modality [$F_{2,8762} = 52.8725, p < 0.0001$], the number of auditory distractors [$F_{6,8762} = 29.5490, p < 0.0001$], and the target angle [$F_{8,8774} = 5.5774, p < 0.0001$] was found. In addition, significant interactions were found between the target angle and the number of auditory distractors [$F_{48,8770} = 1.8185, p = 0.0005$], between the target angle and the target modality [$F_{16,8772} = 3.6972, p < 0.0001$], between the target modality and the number of auditory distractors [$F_{6,8766} = 26.3228, p < 0.0001$], and a three-way interaction between the target modality, the number of auditory distractors and the target angle [$F_{48,8771} = 1.6119, p = 0.0047$]. As seen in Fig. 7.2, only in the audio-only conditions was the localization error affected by the target angle and the number of auditory distractors. At angles larger than ±30°, an increased localization error was found in the condition with eleven auditory distractors as compared to all conditions with fewer auditory distractors [0 – 11: $t_{8771} = 4.842, p < 0.0001$; 7 – 11: $t_{8771} = 3.206, p = 0.0284$]. From ±90°, the condition with seven auditory distractors was also significantly larger compared to conditions with three or less distractors [0 – 7: $t_{8772} = 3.566, p < 0.0076$; 3 - 7: $t_{8772} = 3.227, p = 0.0264$] etc. Only the audio-only conditions with two or less auditory distractors remained unaffected by the target angle.

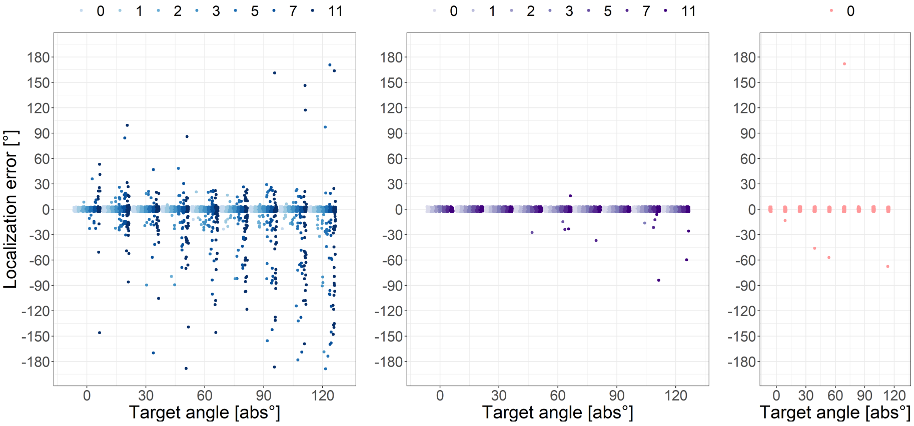


Figure 7.2: The localization error as a function of the absolute target angle for the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel, the data are grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The localization error is jittered ±0.75 degrees in the horizontal and ±3 degrees in the vertical direction to reduce overlap due to the discrete response options.

7.3.2 Response time

Fig. 7.3 shows the total response time as a function of angle for the audio-only (left, blue), audio-visual (purple, center) and visual-only (right, red) conditions. Within each panel the data were grouped by the number of auditory distractors. As can be seen in this figure, response times varied with target modality [$F_{2,8851} = 469.9542, p < 0.0001$], target angle [$F_{8,8854} = 37.1992, p < 0.0001$] and with the number of auditory distractors [$F_{6,8851} = 277.3453, p < 0.0001$]. In addition, significant interactions between the target modality and the target position

[$F_{16,8854} = 3.2021, p < 0.0001$] and between the target modality and the number of auditory distractors [$F_{6,8851} = 100.8071, p < 0.0001$] were found.

In most cases, the response times were largest when the target consisted of only the visual stimulus. The exceptions to this were the conditions with an audio-only target and five or more auditory distractors, which at small angles had larger response times compared to the visual target response times. Audio-visual response times were generally smallest, although when there were no auditory distractors, the response times in the audio-only and audio-visual conditions were similar [0: $t_{8851} = 1.296, p = 0.5849$]. However, as the number of auditory distractors increased, so did the difference in the response times between these two conditions [1: $t_{8851} = 2.697, p = 0.0211$; 11: $t_{8851} = 29.711, p < 0.0001$] as response times strongly increased with the number of auditory distractors in the audio-only conditions [0-11: $t_{8851} = -37.248, p < 0.0001$], whereas the audio-visual conditions were only marginally affected by the number of auditory distractors [0-11: $t_{8851} = -8.721, p < 0.0001$].

In all conditions, response times increased with the target angle. The strongest increase in response times with the target position can be seen in the visual condition. While response times were larger in the audio-only conditions compared to the audio-visual conditions at all angles [0°: $t_{8851} = 8.017, p < 0.0001$; ±120°: $t_{8852} = 8.433, p < 0.0001$], the audio-visual response times were more strongly affected by the target angle [A, 0° - ±120°: $t_{8853} = -6.924, p < 0.0001$; AV, 0° - ±120°: $t_{8853} = -9.300, p < 0.0001$].

7.3.3 Motion behavior

Fig. 7.4 shows the area localization time, defined in the methods section, as a function of the stimulus angle for the audio-only (blue), audio-visual (purple) and visual-only (red) conditions. In each panel, the data are again grouped

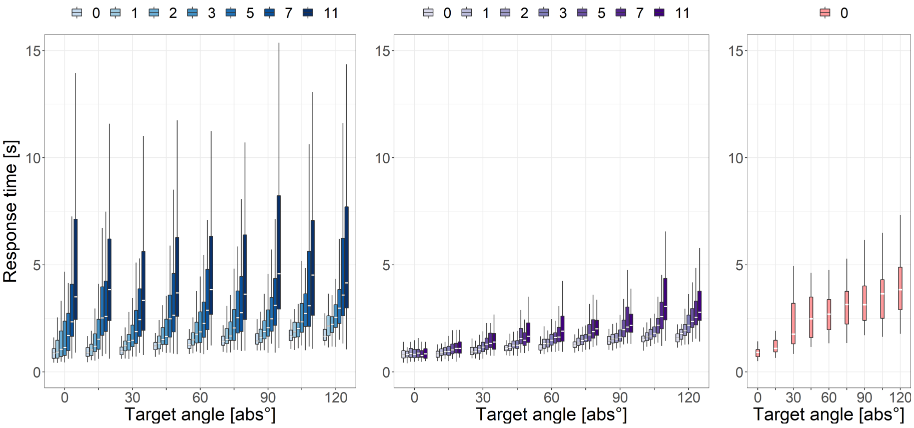


Figure 7.3: A boxplot of the response time as a function of the absolute target angle for the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel the data are grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The boxes extend from the first to the third quartile, with the median shown as the center white line. Whiskers extent to 1.5 times the inter-quartile range. The outliers are not included, see figure 7 in the supplementary figures for the full figure with outliers.

by the number of auditory distractors. Since the area was defined as ±22.5° around the target, target angles between 0° and ±15° were not included in this analysis. As can be seen in Fig. 7.4, search times varied with target modality [$F_{2,6817} = 360.2356, p < 0.0001$], with reaction times being highest, on average, when only visual information was presented. The only exception occurred at ±120°, when the target modality was audio-only and there were 11 distractors present.

Audio-only and audio-visual area localization times were very similar in most conditions. However, the audio-only area localization time was more affected by an increasing number of audio-distractors. The statistical analysis supports this; In addition to the main effect of modality and the main effect of auditory distractors [$F_{6,6862} = 205.9503, p < 0.0001$], an interaction between the target modality and the number of auditory distractors was found [$F_{6,6862} = 24.6446, p < 0.0001$]. The post-hoc comparison revealed that when no distractors were present and performance between the various modalities can be compared, visual-only search times were indeed worse compared to audio-only [$t_{6817} = -21.911, p < 0.0001$] and audio-visual [$t_{6817} = 22.019, p < 0.0001$] search times. However, no difference between the audio and audio-visual search times was found [0: $t_{6817} = 0.075, p = 1.00$]. Only when seven or eleven distractors were present were audio-only search times significantly worse as compared to audio-visual search times [7: $t_{6821} = 2.999, p = 0.0081$; 11: $t_{6822} = 13.830, p < 0.0001$].

An increase of the area localization time with angle and an interaction between the target modality and the effect of target angle [$F_{12,6817} = 3.9866, p < 0.0001$] was found in all conditions [$F_{6,6817} = 83.4562, p < 0.0001$]. In the visual-only condition the median area localization time increased from half a second at ±30° to over 2 seconds at ±105° and ±120°. The effect of angle was much more similar for the audio-only and audio-visual conditions.

Finally, the effect of angle also depended on the number of auditory distractors [$F_{36,6817} = 5.4848, p < 0.0001$], with localization times increasing more strongly with angle as the number of auditory distractors increased. When there are no distractors present, the difference between search times at ±30° and ±120° was only about half a second. When eleven distractors were present, the difference was more than twice as long.

Fig. 7.5 shows the target localization time as a function of target angle. The data were grouped by the target modality and the number of auditory distractors. Again, results between ±15° were not included here. Comparing the target modalities, it can be seen that search times varied with target modality [$F_{2,6783} = 339.6744, p < 0.0001$]. Audio-visual search times were, on average, the smallest and, depending on the number of auditory distractors in the audio-only conditions, either the audio-only (more than three distractors) or visual search times (less than two distractors) were largest.

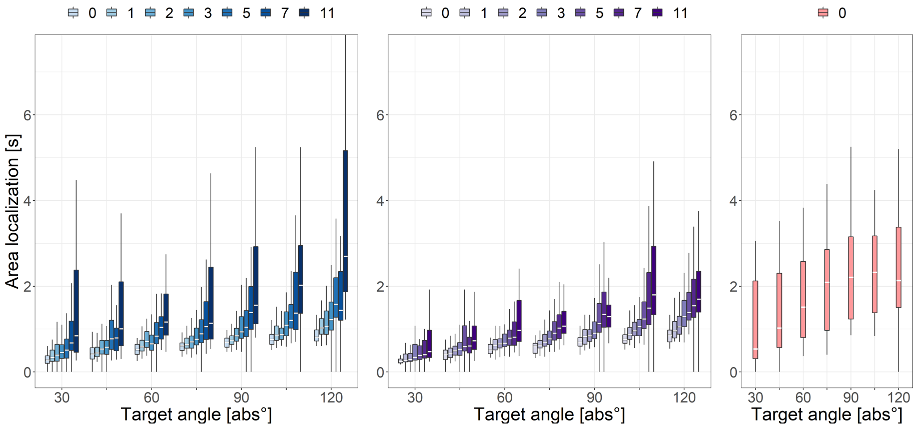


Figure 7.4: A boxplot of the area localization time as a function of the absolute target angle for the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The boxes extend from the first to the third quartile, with the median shown as the center white line. Whiskers extent to 1.5 times the inter-quartile range. For the last boxplot in the audio-only conditions the whisker is outside the figure border to ensure a legible figure scale. The outliers are not included, see figure 8 in the supplementary figures for the full figure.

The auditory and audio-visual target search times were not significantly different from each other when there were no distractors [$t_{6780} = 2.146, p = 0.0956$]. However, as the number of distractors increased, the difference between these conditions also increased, as the time it took participants to find the target depended strongly on the number of distractors. The difference between the audio-only and audio-visual conditions was already significant with a single distractor [$t_{6780} = 3.841, p = 0.0004$].

A main effect of the number of auditory distractors [$F_{6,6781} = 96.5980, p < 0.0001$] and an interaction between the target modality and the number of auditory distractors [$F_{6,6780} = 50.0570, p < 0.0001$] was found. In both the audio-only and audio-visual conditions, localization times increased with the number of auditory distractors. However, in the audio-visual conditions, only the difference between the low and high number of auditory distractors was significant and even the largest difference [$0-11 : t_{6780} = -3.915, p = 0.0019$] was still

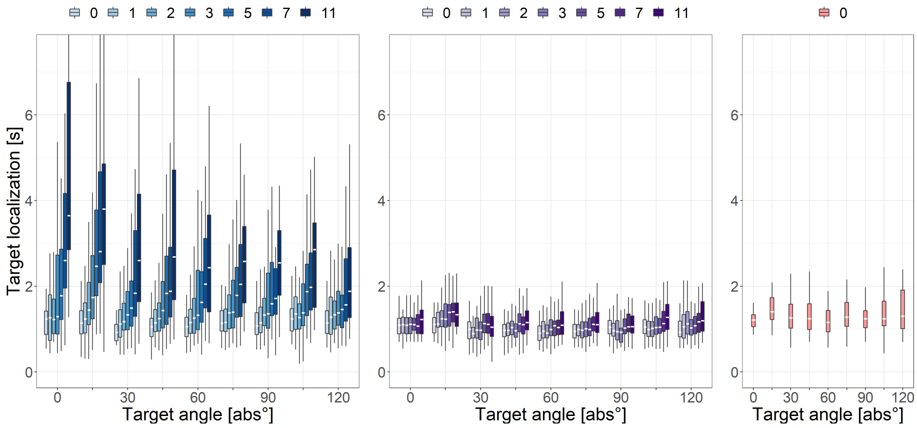


Figure 7.5: A boxplot of the target localization time as a function of the absolute target angle for the auditory (left, blue), audio-visual (middle, purple) and visual (right, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The boxes extend from the first to the third quartile, with the median shown as the center white line. Whiskers extent to 1.5 times the inter-quartile range. For some boxplots in the auditory conditions the whiskers are outside the figure border to ensure a legible figure scale. The outliers are not included, see figure 9 in the supplementary figures for the full figure.

small. On the other hand, in the audio-only condition, larger differences were observed and all except three comparisons (0-1, 1-2 and 3-5) were significant.

The main effect of angle was not significant [$F_{6,6780} = 2.0157, p = 0.0601$], but the interaction between target angle and target modality [$F_{12,6780} = 1.8926, p = 0.0305$] and the three-way interaction between all the factors [$F_{36,6780} = 1.4815, p = 0.0319$] were significant. As shown in Fig. 7.5, when the target was audio-only and there were many distractors, a decreasing trend was found where the search times decreased with increasing angle. This same trend was not found in the audio-visual or visual conditions.

Although visual-only and audio-visual target search times were much more similar than the visual-only and audio-visual area localization times, visual-only search times were still generally larger at all angles. Moreover, this difference was significant at ±45° [$t_{6780} = 2.597, p = 0.0283$] and ±120° [$t_{6780} = 3.342, p = 0.0025$].

7.3.4 Head rotation

Fig. 7.6 shows the maximum head-rotation of the participants in the wrong direction as a function of the target position, with data grouped by modality and number of auditory distractors. The averages are shown as black dots. The statistical analysis revealed an effect of target modality [$F_{2,8299} = 1053.0613, p < 0.0001$], target angle [$F_{8,8299} = 25.9143, p < 0.0001$], the number of auditory distractors [$F_{6,8299} = 65.6662, p < 0.0001$] and interactions between the target modality and the target angle [$F_{2,8299} = 23.5963, p < 0.0001$], between the target modality and the number of auditory distractors [$F_{6,8299} = 6.5823, p < 0.0001$] and, finally, between the number of auditory distractors and the target angle [$F_{48,8299} = 1.7660, p = 0.0009$].

In the audio-only (left, blue) and audio-visual (middle, purple) conditions, the data-points were most densely clustered around 0°, indicating that the participants generally immediately turned towards the correct direction or only towards the wrong direction by a few degrees. However, in both the audio-only and audio-visual conditions, there were many trials where participants turned to the wrong direction. The increase in the average rotation with increasing number of auditory distractors shows that the trials where participants did not immediately search in the correct direction occurred more often when there were more auditory distractors present. This effect of audio-distractors was especially noticeable in the audio-visual conditions, which had, on average, larger rotations into the wrong direction compared to the auditory conditions. This difference between the auditory and audio-visual conditions was significant when there were five or more distractors [5: $t_{8299} = -4.992, p < 0.0001$; 11: $t_{8299} = -4.628, p < 0.0001$].

Similarly, the average rotation into the wrong direction also increased with target angle. As mentioned, this effect of target angle varied per modality. The average rotation increased with the target angle, more so in the visual and the audio-visual conditions than in the auditory conditions. While at small angles the results did not vary with target modality, from ±60° degrees on the difference between the audio-only and audio-visual conditions became significant [±60°: $t_{8299} = -2.901, p = 0.0112$; ±120°: $t_{8299} = -5.839, p < 0.0001$]. The results in the visual-only condition were more strongly dependent on angle, as compared to the audio-visual condition with no distractors. In the visual condition, differences between the conditions were already significant at ±15° [±15°: $t_{8299} = 5.153, p < 0.0001$; ±120°:$t_{8299} = 19.030, p < 0.0001$].

Considering the effect of the number of auditory distractors and the target angle, a positive interaction was found where the rotation into the wrong direction increased more strongly with angle for the conditions with more auditory distractors.

Unlike the audio-only and audio-visual conditions, which had most data-clustering around 0°, in the visual condition, data points clustered around both 0° and 100°. In this condition, there was a clear increase in the average maximum rotation into the wrong direction with target angle. When the target occurred within the immediate field of view, at ±0° or ±15°, participants rarely rotated into the wrong direction or, if they did, generally no further than ±30°. At increased target angles, however, the averages fall about halfway in between the data clusters, while most data were clustered around 0° and 100°, reflecting that in about half of the trials, participants initially rotated into the wrong direction. The lack of data points in between 0° and 100° shows that when participants initially turned their head in the wrong direction, they searched all option in one hemisphere before moving towards the other.

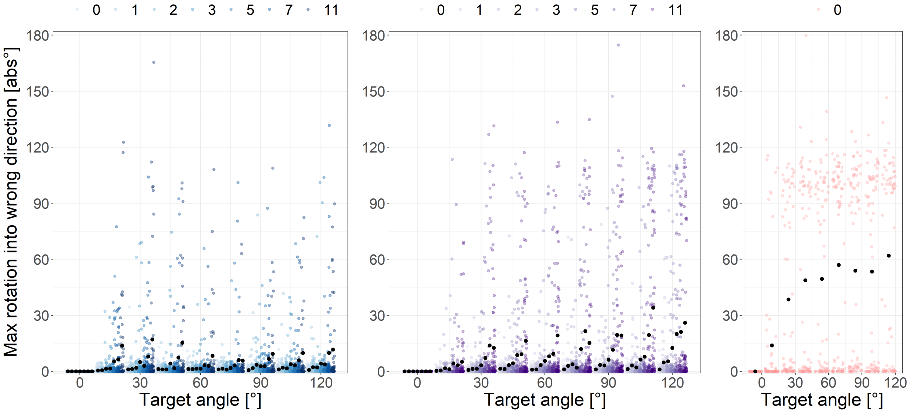


Figure 7.6: Maximum head rotation into the wrong direction as a function of the target position in the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The black dots indicate the average per angle, target modality and number of auditory distractors.

7.4 Discussion

In this study, we investigated whether audio-visual localization behavior can be explained by a combination of auditory area search and visual target search and how it is affected by auditory distractors. As the goal was to study how participants localized the stimuli rather than if they could, the task was relatively easy. In the visual and audio-visual conditions, the localization error was insignificant. Significant localization errors were only found in the most difficult audio-only conditions, where were many audio distractors were present. In analyzing the area localization and the target localization times, the results did not fully match our hypothesis. For up to seven distractors, no significant difference was found between the audio-only and the audio-visual area localization times, as was hypothesized. However, the visual-only target localization times were generally larger than the audio-visual target localization times and this difference was also significant at several angles. Moreover, when there were more than seven distractors, audio-only area localization times also could no longer explain the audio-visual area localization times. In fact, in the conditions with seven or eleven distractors, audio-visual area localization times were faster than both the audio-only and the visual-only search times.

Thus, while part of the results of the area localization times are in line with what would be expected if auditory localization guided visual localization (Heffner and Heffner, 2016; Heffner and Heffner, 1992; Perrott et al., 1990), our results suggest an influence of visual information also on the area-localization, at least when there are many auditory distractions. The head-movements of the participants further support this influence of visual information on the area localization task. When the number of auditory distractors was small, the average rotation into the wrong direction and the distribution of the data in the audio-visual conditions was similar to that in the audio-only conditions. However, as the number of auditory distractors increased, a significant difference in the average rotation between the audio-only and audio-visual conditions was found. Most importantly, the incorrect rotation was larger in the audio-visual conditions and the distribution of the data in those conditions was similar to both the distributions in the audio-only and visual-only conditions. Two clusters were found around 0° and 100° as in the visual-only conditions, but the majority of the data points occurred around 0° and a few data points occurred in between these clusters as in the audio-only conditions. This shows, both in the head motions and in the search times, that with an increase in the number of auditory distractors, audio-visual area localization is influenced by visual information.

Similarly, audio-visual target localization times within this field of view were smaller than both audio-only and visual-only target localization times. This is not surprising, as many studies have shown decreased localization times for congruent audio-visual stimuli, due to multisensory integration (e.g., Diederich and Colonius, 2004; Mahoney et al., 2011; Miller, 1982; Schröger and Widmann, 1998). However, it again does not support the idea of the auditory system guiding visual localization. Instead, our results support the audio-visual guidance of audio-visual localization. When the number of auditory distractors were low, this audio-visual guidance could be well explained by pure auditory guidance, however, as the number of auditory distractors increased a shift towards audio-visual guidance was found. A limitation of the current study, however, is the definition of the area localization time and the target localization time, as the field of view was based on the overall reaction times. The larger visual target localization times can also be explained by an overestimation of the field of view; an overestimation would confound part of the area-localization time, where the visual search times were larger, with the target search time.

Overall, the results show a strong integration between the auditory and visual system in localizing stimuli and the interactions of these systems depended on the number of auditory distractors.

7.5 Supplementary figures

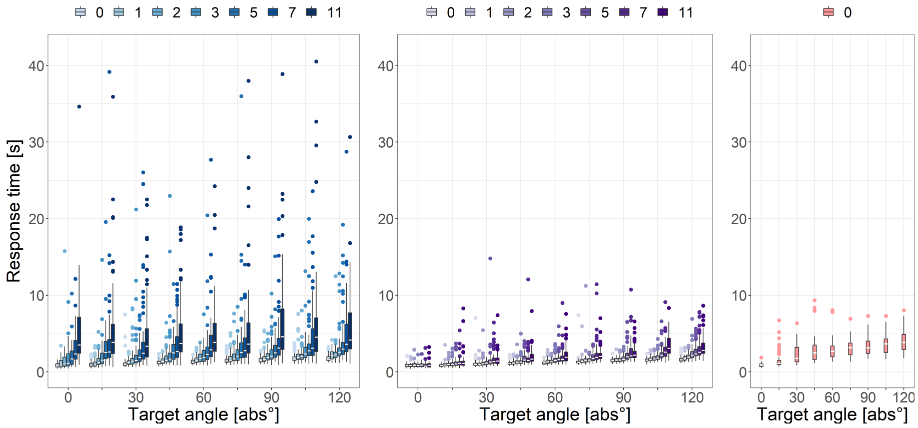


Figure 7.7: A boxplot of the response time as a function of the absolute target angle for the audio-only (upper, blue), audio-visual (middle, purple) and visual-only (lower, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The localization error is jittered in the vertical direction to reduce overlap due to the discrete response options. The boxes extend from the first to the third quartile, with the median shown as the center black line. Whiskers extent to 1.5 times the inter-quartile range.

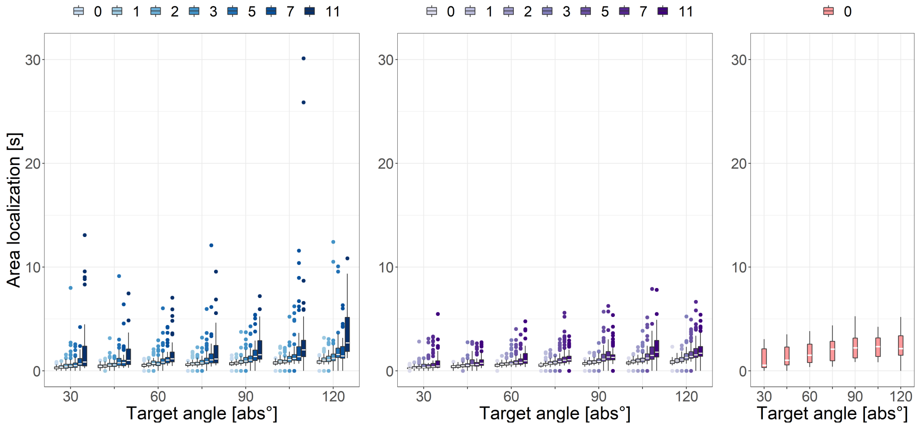


Figure 7.8: A boxplot of the area localization time as a function of the absolute target angle for the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The boxes extend from the first to the third quartile, with the median shown as the center black line. Whiskers extent to 1.5 times the inter-quartile range. The outliers are shown separately.

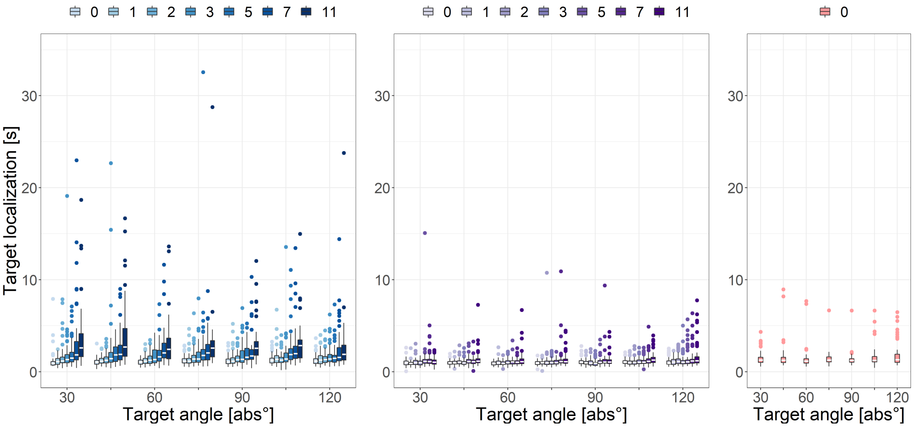


Figure 7.9: A boxplot of the target localization time as a function of the absolute target angle for the audio-only (left, blue), audio-visual (middle, purple) and visual-only (right, red) conditions. In each panel the data were grouped by the number of auditory distractors, with increasing numbers of auditory distractors being indicated by darker colors. The boxes extend from the first to the third quartile, with the median shown as the center black line. Whiskers extent to 1.5 times the inter-quartile range. The outliers are shown separately.

8

Overall discussion

8.1 Summary of main results

Many studies have shown detrimental effects of hearing loss on localization performance (e.g., Akeroyd and Whitmer, 2016; Häusler et al., 1983; Lorenzi et al., 1999; Noble et al., 1994; Otte et al., 2013; Rakerd et al., 1998). However, it is unclear how a hearing loss impacts hearing-impaired adults in their daily life, as most experimental tasks and settings do not reflect daily experiences. In the real world, more information, such as visual cues, self-motion cues etc., is available to solve problems that hearing-impaired listeners might face in laboratory settings. As such, the long-term goal is to investigate spatial hearing in (aided) hearing-impaired listeners in ecologically valid environments. To move towards this goal, in this thesis the focus was on investigating the influence of visual information on spatial hearing in normal-hearing and hearing-impaired listeners.

In *Chapter 3*, a Gaussian clustering and categorization method was used to distinguish audio-visual responses (reflecting integration) from visual responses (reflecting a response biases). While the analysis method was able to filter out consistent visual responses, the instability of the clusters, especially those clusters where the difference between visual and audio-visual information was unclear and where an objective separation method was most needed, meant that this method was not widely applicable without further improvements.

However, some very interesting aspects were found in the data. The probability of integration was largest when the stimuli were slightly misaligned at 3 degrees spatial separation, with the visual stimulus occurring slightly further outwards compared to the auditory stimulus. Furthermore, the spatial integration window ranged from -11.9 to 28.2 degrees and, as visible from the shift in the optimal integration point and the asymmetry of the spatial integration window, stimuli were more likely to be integrated when the visual stimulus occurred at a position that was slightly more eccentric than that of the audio stimulus.

In *Chapter 4*, the effect of stimulus realism on the ventriloquist effect was evaluated. Previous studies suggested facilitative effects of stimulus realism on integration. However, as temporal correlation between the auditory and visual stimuli often increase with stimulus realism, it was unclear if these effects were due to temporal correlation of the stimuli or stimulus realism. The results from *Chapter 4* were not fully decisive, with contrasting differences between the stimuli occurring depending on the specific stimulus angle and relative stimulus positioning. When the auditory stimulus occurred at 0 degrees azimuth and the visual stimulus occurred towards the left, the probability of integration was significantly smaller in the most realistic condition (with a ball falling down and making an impact sound as it bounced). However, when the visual stimulus instead occurred on the right, the visual bias was smallest (even negative) for one of the least realistic conditions (noise + flash with distractor). As in *Chapter 3*, the relative stimulus positioning affected the probability of integration. Overall, no consistent effect of stimulus realism was found, certainly not to the extent of what some previous studies suggested.

In *Chapter 5*, we investigated how the head mounted display (HMD) affected Ambisonics sound source localization. As in previous studies that investigated the effect of the HMD on single loudspeaker sound source localization, a shift in the perceived localization was found, where stimuli were perceived at increased eccentricities when wearing the HMD. This shift was largest when the sounds were presented around ±52.5 degrees and the shift was larger in the right hemisphere than in the left hemisphere. The localization performance was strongly affected by the Ambisonics order used to reproduce the sound; When stimuli were presented using first order Ambisonics, stimuli were perceived significantly closer to the center than when stimuli were presented using higher-order Ambisonics. Within the higher-order Ambisonics (3rd, 5th and 11th), however, the difference was minimal. In contrast to the localization performance, the shift in the perceived location as a result of the HMD was independent of the Ambisonics order. When visual information about the location of the loudspeakers was available to the participant, the responses of participants were clearly biased towards the loudspeakers. In cases where sounds were simulated at a position of a loudspeaker, the addition of visual information generally improved localization, although there were cases where the perceived location of the sound matched better with a neighboring loudspeaker, resulting instead in an increase of the localization error. Similarly, for sounds that were simulated half-way in between loudspeakers, adding visual information typically increased the localization error. These results showed again, how strongly visual information can bias sound localization. Thus, while the HMD shifted the perceived location of the sound sources, visual information generally compensated for this shift.

Chapter 6 investigated how age and hearing loss affect the spatial integration window. In all three groups, namely older normal-hearing (ONH), young normal-hearing (YNH) and older hearing-impaired (OHI), similar trends were found. The just-noticeable-difference in angle was smallest in the audio-visual and visual conditions, with thresholds being just slightly smaller, on average, in the audio-visual conditions. The minimum audible angle was about 2 degrees larger than the audio-visual and visual thresholds. Finally, the incongruent audio-visual condition, which was used to estimate the spatial integration window, had the largest thresholds. A significant increase in the spatial integration window was found when comparing the OHI and ONH groups with the YNH groups, i.e., an effect of age, but not when comparing the OHI and the ONH groups, i.e., no effect of hearing loss was found. This difference between the groups was only significant when the stimuli were presented around 0 degrees. When stimuli were presented at increased angles, the difference between the auditory-only and incongruent audio-visual condition was not significant, i.e., participants were able to ignore the visual stimulus. While no effect of hearing loss was found, the visual acuity of the participants did affect the spatial integration window, with better visual acuity predicting better thresholds and a smaller integration window at most angles. Measurements of the reaction time showed that the older participants (ONH and OHI groups) generally had longer reaction times. In contrast to expectations, reaction times were not shorter in the audio-visual conditions (both congruent and incongruent). In fact, no consistent pattern was found between various conditions with regards to reaction times.

Finally, *Chapter 7* moved towards investigating congruent spatial localization behavior in more realistic settings. The experiment with normal-hearing listeners showed that while localization accuracy was generally highest when visual information was available, auditory information improved response times more strongly. Only when there were many auditory distractors present was visual localization faster than auditory localization. Visual localization was also more strongly influenced by angle. When the visual target was presented at angles larger than 15 degrees, a large step in the response times was found. Head-rotation data also found that participants approached auditory localization and visual localization differently. In the audio-only conditions, the participants generally immediately rotated their heads towards the right direction. In contrast, in the visual condition, participants moved their heads towards the wrong direction in about half of the trials. Moreover, the head-rotation data suggested that participants searched all options in a single direction before searching in the other, whereas in the audio and audio-visual conditions participants switched to the other hemisphere at any point. Audio-visual localization was the fastest. The behavioral data suggested that both auditory and visual localization contributed to the audio-visual search task, in both the area localization and the target localization. As the number of audio-distractors increased, the patterns observed in the head-rotation data of the audio-visual conditions shifted towards those observed in the head-rotation data of the visual-only condition.

8.1.1 Using VR and a loudspeaker array for ecologically valid audio-visual studies

Using VR with a loudspeaker array combines experimental control, both over the visual and acoustic scene, and ecological validity. Because of this, it has a lot of potential in perceptual studies. However, some care needs to be taken in the application, especially with older participants.

We chose to use this combination of VR with the loudspeaker array to achieve the longer-term goal of testing spatial localization in more complex and realistic settings in both hearing-impaired, but also aided hearing-impaired listeners. Particularly for the latter, sound reproduction via loudspeakers is preferable to using headphones. This is because most hearing aids make use of multiple microphones and it is challenging to simulate the appropriate acoustic signal at each microphone on the hearing aid. Thus, to test aided perception using headphone reproduction, one also has to simulate the effects of the hearing aid processing. The combination of VR and the loudspeaker array used in the studies in this thesis offered both the realistic and controllable audio and visual environment, and it can accommodate aided hearing-impaired listeners in potential future studies.

However, while indeed providing this realistic and controllable environment, there are some challenges with the setup which need to be taken into account for futures studies. First, as established in earlier studies (Ahrens et al., 2019; Genovese et al., 2018; Gupta et al., 2018) and also in *Chapter 5* with Ambisonics sound reproduction, there is a small shift in the perceived location of sound sources when loudspeakers are used to produce sound, due to the size of the HMD. In most use cases, visual information can correct these small shifts, as was also seen in *Chapter 5*. This is further corroborated by the localization results in *Chapter 7* where no consistent localization error was found in the audio-only conditions with few distractors, regardless of angle. However, in precise localization experiments, such as experiment 1 and 2, this might have slightly increased the unimodal auditory bias (since both the auditory bias and HMD shift results in perceived sound sources being shifted towards the periphery). Moreover, shifts might have occurred also in *Chapter 6*, thus affecting the perceived distance between the two auditory stimuli, although the jittering of the stimuli and the randomizing of the initial direction of the stimuli should have reduced most effects on the results.

Second, there are some challenges in using VR specifically with older adults (Seifert and Schlomann, 2021). There is a 'digital gap' between younger and older generations, where younger generations are more used to working with and make more use of newer technology (Hunsaker and Hargittai, 2018; Seifert et al., 2017). Due to this gap, working with VR is generally less intuitive to older adults and, more importantly, they are more susceptible to cyber sickness (Maneuvrier et al., 2020; Rebenitsch and Owen, 2016; Stauffert et al., 2020). Indeed, two participants (both older women) reported some discomfort due to the VR experience.

Third, it is unclear to what extent VR experiences generalize, as some differences in VR and real-world perception have been found. Clemenson et al. (2020), for example, found that the way we navigate new locations can differ between real-world and virtual experience, distance perception is regularly reported as shorter in VR environments than in real-world environments (for a review see Renner et al. (2013)) and in *Chapter 5*, a small difference in pointing at real world and virtual sources was found.

For the purpose of ecologically valid audio-visual experiments, especially with aided hearing-impaired listeners, the benefits of the combination of VR and a loudspeaker array will likely outweigh these downsides; Having run multiple experiments both with younger and older, normal-hearing and hearing-impaired adults, the experiences with this setup were very positive for the majority of the participants. Several participants expressed excitement over having experienced VR and some noted that the VR experience was part of their motivation for joining in these experiments. Moreover, in experiments such as described in the *Chapter 6* and *Chapter 7*, the shift of the HMD is unlikely to affect the results, negating most of the downsides. However, audio-visual experiments that use absolute localization, such as described in *Chapter 3* and *Chapter 4*, might be better studied using a different combination, if possible.

Thus, there is a lot of potential for the use of VR in perceptual experiments; It combines experimental control and ecological validity and, possibly in part due to its relative novelty, is exciting to most participants. However, for experiments relying on absolute auditory localization, the effects of the shifts in the perceived location need to be considered. Moreover, care needs to be taken when working with older participants.

8.1.2 The effect of absolute and relative stimulus positioning

Together, the studies in thesis have demonstrated that both absolute and relative stimulus positioning affect the probability of integration. Considering the relative stimulus positioning, in chapter *Chapter 3* and chapter *Chapter 4* integration was found to be more likely when the visual stimuli occurred at more peripheral locations relative to the auditory stimuli (except for at 0 degrees in *Chapter 4*). As discussed in those chapters, this could be due to the peripheral auditory bias and the central visual bias decreasing the perceived distance between these stimuli. Such biases, where the auditory stimuli were perceived more towards the periphery then they actually were, were found both in previous studies (e.g., Ahrens et al., 2019; Odegaard et al., 2015; Parise et al., 2012), but also in *Chapter 3*, *Chapter 4* and *Chapter 5*. Similarly, in *Chapter 3* and *Chapter 4* visual localization biases were found, where the visual stimuli were perceived more closer to the center, consistent with Mateeff and Gourevich (1983), Odegaard et al. (2015), and Parise et al. (2012). However, as noted also in *Chapters3*, Godfroy et al. (2003) found no effect of the relative stimulus positioning on the probability of the audio and visual stimuli being perceived as coming from the same position. In neither their nor our experiments, however, was the effect of the localization biases the focus of the experiment. Odegaard et al. (2015) studied how these biases affected congruent integration and found that the combined percept was less biased (although a significant centric bias was still found), but they did not include incongruent stimulus presentations. In fact, to our knowledge, no study has specifically focused on whether these unimodal biases affect the spatial integration window. The results of such a study could be very interesting since it is commonly assumed that the spatial integration window is symmetric. However, results by Godfroy et al. (2003) and *Chapter 3* and *Chapter 4* do not support this. An asymmetric spatial integration window is not necessarily surprising, given that the temporal integration window is not symmetric. As mentioned in *Chapter 3*, several studies have found that people are more tolerant of auditory lagging temporal disparities (Bhat et al., 2015; Slutsky and Recanzone, 2001; Wassenhove et al., 2007). Moreover, as auditory localization biases have been associated with increased localization variance (Garcia et al., 2017), localization biases themselves might also be associated with increased spatial integration windows.

Besides an effect of relative stimulus positioning, *Chapter 3*, *Chapter 4* and *Chapter 6* also demonstrated an effect of stimulus eccentricity. As stimuli were presented further away from the center integration became less likely, although in *Chapter 4* this effect of stimulus eccentricity depended on the relative stimulus positioning. The effect of stimulus eccentricity was particularly noticeable in *Chapter 6*, where the difference between auditory and incongruent audio-visual localization was only significant at 0 degrees azimuth. Similar effects of eccentricity, as found in *Chapter 3* and partially *Chapter 4* have also been found previously (Charbonneau et al., 2013; Hairston et al., 2003), with the decrease in the probability of integration reflecting a change in the relative reliability (Charbonneau et al., 2013).

8.1.3 The influence of vision on hearing of hearing-impaired listeners

The results from this thesis showed the importance of visual information on auditory localization. *Chapter 3*, *Chapter 4* and *Chapter 6* focused on the ventriloquist paradigm and showed how strongly biased auditory localization can be towards visual information, while *Chapter 6* focused on combined localization and showed a more balanced combination of the sensory systems. Even in *Chapter 5*, where the focus was on determining how the HMD affected Ambisonics sound source localization, an influence of vision on auditory localization was seen. Particularly *Chapter 7* showed, though, how important it is to evaluate spatial localization in ecologically valid settings. Based on the results of the first few chapters, one might underestimate the importance of auditory localization, since visual information appears to be the main dominating factor in localization. However, the last experiment revealed how much faster audio-visual localization is than either unimodal localization strategy and that participants adjust their localization behavior based on the available information.

Although a large impact of visual information was found on auditory localization, *Chapter 6* found no difference in the results of older normal-hearing adults and older hearing-impaired adults. These null-results were in line with several other integration studies with hearing-impaired participants. In the temporal domain, Başkent and Bazo (2012) found no effect of hearing loss and even CI users performed similarly to normal-hearing listeners (Butera et al., 2018) in a temporal order judgement task. Audio-visual speech perception was also not improved in hearing-impaired listeners (Tye-Murray et al., 2007). Then again, as also mentioned in *Chapter 2*, other studies did find effects of hearing-loss on audio-visual speech (Puschmann et al., 2019; Rosemann and Thiel, 2018; Schulte et al., 2020; Stropahl and Debener, 2017), spatial integration (Venskytis et al., 2019) and Puschmann et al. (2014) found an increased susceptibility to cross-modal distractors.

What causes these differences across studies regarding the effects of hearing-loss remains unclear. In the case of the spatial integration, unimodal performance may explain the results; In *Chapter 6*, the auditory and visual localization performance was similar between the older normal-hearing and older hearing-impaired listeners. Similarly, in Venskytis et al. (2019) participants that did not show poorer auditory localization performance also did not show an increased visual bias. However, the unimodal performance cannot explain the apparent contradictions in all cases; although Başkent and Bazo (2012) found a large individual variability in their study, there was no correlation in the results with either age or hearing loss. Moreover, the same target group (older adults with a mild to moderate sensorineural hearing loss) showed diminished integration in one McGurk study (Musacchia et al., 2009), but an increased susceptibility to the McGurk effect in another (Rosemann and Thiel, 2018) and showed no difference from the results obtained with normal-hearing participants in another audio-visual speech integration study (Tye-Murray et al., 2007). These results show that there is still a lot of research required to better understand if and how hearing loss affects audio-visual integration.

8.2 Perspectives

The current results have been very promising for hearing-impaired listeners, as they showed no difference in auditory or audio-visual performance relative to the performance obtained in the normal-hearing group. However, more research is needed to better understand if and how a hearing loss affects audio-visual integration, not just in the spatial, but also the temporal domain, both in non-speech as wel as speech studies. A first step would be to extend the paradigm in *Chapter 7* to hearing-impaired participants. At increased angles (outside the visual field) where auditory information could be most vital, hearing loss has been found to more strongly impact localization performance than in the case of stimuli presented from the front (Häusler et al., 1983; Rakerd et al., 1998). Additionally, in *Chapter 7*, a change in localization behavior was already observed, where participants showed more signs of visual localization behavior when the number of distractors increased. Exploring if and how these behavioral changes occur in hearing-impaired participants could provide more insights into the characteristics of audio-visual integration in hearing-impaired listeners, the localization challenges in hearing-impaired-listeners and the behavioral adjustments that they might undertake to compensate for changes in auditory performance in the periphery. Although the original goal was to investigate both hearing-impaired and aided hearing-impaired listeners, no experiments with aided hearing-impaired listeners have been performed. Studies with aided hearing-impaired listeners have so far reported less integration in the temporal domain (Hirst et al., 2020), but an increased temporal integration window (Gieseler et al., 2018), i.e., they integrated less in total, but integrated over a longer range of temporal disparities. These are both indications of reduced auditory reliability. Depending on the results of running the experiment in *Chapter 7* with hearing-impaired listeners), the next step could be to test this with aided hearing-impaired listeners and continue moving towards investigating spatial hearing of hearing-impaired and aided-hearing-impaired listeners in ecologically valid experiments.

8.3 Conclusions

The findings presented throughout the chapters of the thesis suggest the following:

- The spatial audio-visual integration window is asymmetric, with integration being more likely when the auditory stimuli are presented closer to the center and the visual stimuli are presented more towards the periphery.

- Audio-visual integration decreases when the stimuli are presented further away from the center.
- The VR headset can shift the perceived location of auditory stimuli when they are presented with a loudspeaker array.
- Mild-to-moderate hearing loss does not affect the spatial audio-visual integration window.
- Age, however, increases the spatial audio-visual integration window.
- Localization behavior changes depending on the reliability of the auditory and visual information.

Overall, these findings provide a starting point for investigating audio-visual localization in more ecologically valid settings, which may, in the future, help better understand the challenges that hearing-impaired and aided-hearing-impaired listeners face, which is likely to help in the design of new hearing-aid processing algorithms or deciding between already existing algorithms.

Bibliography

Acerbi, L., K. Dokka, D. E. Angelaki, and W. J. Ma (2018). “Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception”. In: *PLoS Computational Biology* 14.7. Ed. by S. J. Gershman, e1006110.

Ahrens, A., K. D. Lund, M. Marschall, and T. Dau (2019). “Sound source localization with varying amount of visual information in virtual reality”. In: *PLoS ONE* 14.3, pp. 1–19.

Ahrens, A., M. Marschall, and T. Dau (2020). “The effect of spatial energy spread on sound image size and speech intelligibility”. In: *The Journal of the Acoustical Society of America* 147.3, pp. 1368–1378.

Akeroyd, M. A. and W. M. Whitmer (2016). *Hearing Aids*. Vol. 56, pp. 10–16.

Alais, D. and D. Burr (2004). “The Ventriloquist Effect Results from Near-Optimal Bimodal Integration”. In: *Current Biology* 14.3, pp. 257–262.

Altieri, N. and D. Hudock (2014). “Hearing impairment and audiovisual speech integration ability: A case study report”. In: *Frontiers in Psychology* 5.JUL, pp. 1–10.

Andersen, T. S. and P Mamassian (2008). “Audiovisual integration of stimulus transients”. In: *Vision Research* 48, pp. 2537–2544.

Anderson, P. W. and P. Zahorik (2014). “Auditory/visual distance estimation: Accuracy and variability”. In: *Frontiers in Psychology* 5.SEP, pp. 1–11.

Arnold, D. H., K. Petrie, C. Murray, and A. Johnston (2019). “Suboptimal human multisensory cue combination”. In: *Scientific Reports* 9.1.

Bailey, H. D., A. B. Mullaney, K. D. Gibney, and L. D. Kwakye (2018). “Audiovisual Integration Varies with Target and Environment Richness in Immersive Virtual Reality”. In: *Multisensory Research* 31.7, pp. 689–713.

Başkent, D. and D. Bazo (2012). “Erratum: Audiovisual asynchrony detection and speech intelligibility in noise with moderate to severe sensorineural hearing impairment (Ear Hear (2011) 32 (582-592))”. In: *Ear and Hearing* 33.2, p. 303.

Bates, E., G. Kearney, D. Furlong, and F. Boland (2007). “Localization accuracy of advanced spatialisation techniques in small concert halls”. In: *The Journal of the Acoustical Society of America* 121.5, pp. 3069–3070.

Battaglia, P. W., R. A. Jacobs, and R. N. Aslin (2003). “Bayesian integration of visual and auditory signals for spatial localization”. In: *Journal of the Optical Society of America A* 20.7, p. 1391.

Baum, S. H. and R. A. Stevenson (2017). “Shifts in Audiovisual Processing in Healthy Aging”. In: *Current Behavioral Neuroscience Reports* 4.3, pp. 198–208.

Bedard, G. and M. Barnett-Cowan (2016). “Impaired timing of audiovisual events in the elderly”. In: *Experimental Brain Research* 234.1, pp. 331–340.

Begault, D and M. T. Pittman (1996). “Three-dimensional audio versus head-down traffic alert and collision avoidance system displays.” In: *The International journal of aviation psychology* 6 1, pp. 79–93.

Bernstein, I. H., P. K. Chu, P Briggs, and D. L. Schurman (1973). “Stimulus intensity and foreperiod effects in intersensory facilitation”. In: *Quarterly Journal of Experimental Psychology* 25.2, pp. 171–181.

Bertelson, P. and G. Aschersleben (1998). “Automatic visual bias of perceived auditory location”. In: *Psychonomic Bulletin and Review* 5.3, pp. 482–489.

Bertelson, P., J. Vroomen, B. De Gelder, and J. Driver (2000). “The ventriloquist effect does not depend on the direction of deliberate visual attention”. In: *Perception & Psychophysics* 62.2, pp. 321–332.

Bertelson, P., J. Vroomen, G. Wiegeraad, and B. De Gelder (1994). “Exploring the relation between McGurk interference and ventriloquism”. In: *Icslp 94. 1994 International Conference on Spoken Language Processing*. Vol. 2, pp. 559–562.

Bertet, S., J. Daniel, E. Parizet, and O. Warusfel (2013). “Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources”. In: *Acta Acustica united with Acustica* 99.4, pp. 642–657.

Bhat, J., L. M. Miller, M. A. Pitt, and A. J. Shahin (2015). “Putative mechanisms mediating tolerance for audiovisual stimulus onset asynchrony”. In: *Journal of Neurophysiology* 113.5, pp. 1437–1450.

Bidelman, G. M. (2016). “Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion”. In: *Experimental Brain Research* 234.10, pp. 3037–3047.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bizley, J. K., B. G. Shinn-Cunningham, and A. K. Lee (2012). “Nothing is irrelevant in a noisy world: Sensory illusions reveal obligatory within-and across-modality integration”. In: *Journal of Neuroscience* 32.39, pp. 13402–13410.

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

Bolia, R. S., W. R. D’Angelo, and R. L. McKinley (1999). “Aurally aided visual search in three dimensional space”. In: February.

Bolognini, N., F. Rasi, and E. Làdavas (2005). “Visual localization of sounds”. In: *Neuropsychologia* 43.11, pp. 1655–1661.

Bonath, B. et al. (2007). “Neural Basis of the Ventriloquist Illusion”. In: *Current Biology* 17.19, pp. 1697–1703.

Bosen, A. K., J. T. Fleming, S. E. Brown, P. D. Allen, W. E. O’Neill, and G. D. Paige (2016). “Comparison of congruence judgment and auditory localization tasks for assessing the spatial limits of visual capture”. In: *Biological Cybernetics* 110.6, pp. 455–471.

Bouchara, T., C. Jacquemin, and B. F. Katz (2013). “Cueing multimedia search with audiovisual blur”. In: *ACM Transactions on Applied Perception* 10.2.

Brungart, D, N Iyer, and B Simpson (2010). “Aurally Aided Visual Search with Multiple Audio Cues”. In.

Bruns, P. (2019). “The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update”. In: *Frontiers in Integrative Neuroscience* 13.

Bruns, P., M. Maiworm, and B. Röder (2014). “Reward expectation influences audiovisual spatial integration”. In: *Attention, Perception, and Psychophysics* 76.6, pp. 1815–1827.

Butera, I. M., R. A. Stevenson, B. D. Mangus, T. G. Woynaroski, R. H. Gifford, and M. T. Wallace (2018). “Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants”. In: *Scientific Reports* 8.1, pp. 1–12.

Caclin, A., S. Soto-Faraco, A. Kingstone, and C. Spence (2002). “Tactile "capture" of audition”. In: *Perception and Psychophysics* 64.4, pp. 616–630.

Campbell, J. and A. Sharma (2014). “Cross-modal re-organization in adults with early stage hearing loss”. In: *PLoS ONE* 9.2.

Chan, J. S. et al. (2021). “Predictive Coding Over the Lifespan: Increased Reliance on Perceptual Priors in Older Adults—A Magnetoencephalography and Dynamic Causal Modeling Study”. In: *Frontiers in Aging Neuroscience* 13.April, pp. 1–14.

Chan, J. et al. (2015). “Expanded Temporal Binding Windows in People with Mild Cognitive Impairment”. In: *Current Alzheimer Research* 12.1, pp. 61–68.

Chan, Y. M., M. J. Pianta, and A. M. McKendrick (2014). “Reduced audiovisual recalibration in the elderly”. In: *Frontiers in Aging Neuroscience* 6.AUG, pp. 1–9.

Charbonneau, G., M. Veronneau, C. Boudrias-Fournier, F. Lepore, and O. Collignon (2013). “The ventriloquist in periphery: Impact of eccentricity-related reliability on audio-visual localization”. In: *Journal of Vision* 13, 20.12, pp. 1–14.

Chen, L. and J. Vroomen (2013). “Intersensory binding across space and time: A tutorial review”. In: *Attention, Perception, and Psychophysics* 75.5, pp. 790–811.

Chen, Y.-C. and C. Spence (2017). “Assessing the Role of the ‘Unity Assumption’ on Multisensory Integration: A Review”. In: *Frontiers in Psychology* 8.MAR.

Chuen, L. and M. Schutz (2016). “The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues”. In: *Attention, Perception, and Psychophysics* 78.5, pp. 1512–1528.

Chung, K., A. C. Neuman, and M. Higgins (2008). “Effects of in-the-ear microphone directionality on sound direction identification”. In: *The Journal of the Acoustical Society of America* 123.4, pp. 2264–2275.

Clemenson, G. D., L. Wang, Z. Mao, S. M. Stark, and C. E. L. Stark (2020). “Exploring the Spatial Relationships Between Real and Virtual Experiences: What Transfers and What Doesn’t”. In: *Frontiers in Virtual Reality* 1, p. 17.

Colonius, H. and A. Diederich (2011). “Computing an optimal time window of audiovisual integration in focused attention tasks: Illustrated by studies on effect of age and prior knowledge”. In: *Experimental Brain Research* 212.3, pp. 327–337.

Colonius, H. and A. Diederich (2020). “Formal models and quantitative measures of multisensory integration: a selective overview”. In: *European Journal of Neuroscience* 51.5, pp. 1161–1178.

De Boer-Schellekens, L. and J. Vroomen (2014). “Multisensory integration compensates loss of sensitivity of visual temporal order in the elderly”. In: *Experimental Brain Research* 232.1, pp. 253–262.

DeLoss, D. J., R. S. Pierce, and G. J. Andersen (2013). “Multisensory Integration, Aging, and the Sound-Induced Flash Illusion”. In: *Psychol Aging* 28.3, pp. 802–812.

Delong, P. et al. (2018). “Invisible Flashes Alter Perceived Sound Location”. In: *Scientific Reports* 8.1, p. 12376.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series B (methodological)* 39.1, pp. 1–38.

Di Luzio, P., S. Borgomaneri, S. Sanchioni, A. Tessari, and V. Romei (2021). “Exposure to first-person shooter videogames is associated with multisensory temporal precision and migraine incidence”. In: *Cortex* 134, pp. 223–238.

Dias, J. W., C. M. McClaskey, and K. C. Harris (2021). “Audiovisual speech is more than the sum of its parts: Auditory-visual superadditivity compensates for age-related declines in audible and lipread speech intelligibility”. In: *Psychology and Aging* 36.4, pp. 520–530.

Diederich, A. and H. Colonius (2004). “Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time”. In: *Perception and Psychophysics* 66.8, pp. 1388–1404.

Diederich, A., H. Colonius, and A. Schomburg (2008). “Assessing age-related multisensory enhancement with the time-window-of-integration model”. In: *Neuropsychologia* 46.10, pp. 2556–2562.

Dobreva, M. S., W. E. O’Neill, and G. D. Paige (2012). “Influence of age, spatial memory, and ocular fixation on localization of auditory, visual, and bimodal targets by human subjects”. In: *Experimental Brain Research* 223.4, pp. 441–455.

Donohue, S. E., M. G. Woldorff, and S. R. Mitroff (2010). “Video game players show more precise multisensory temporal processing abilities”. In: *Attention, Perception, and Psychophysics* 72.4, pp. 1120–1129.

Dufour, A., O. Després, and T. Pebayle (2002). “Visual and auditory facilitation in auditory spatial localization”. In: *Visual Cognition* 9.6, pp. 741–753.

Echevarria Sanchez, G. M., T. Van Renterghem, K. Sun, B. De Coensel, and D. Botteldooren (2017). “Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space”. In: *Landscape and Urban Planning* 167, pp. 98–107.

Elliott, D. B. (2016). “The good (logMAR), the bad (Snellen) and the ugly (BCVA, number of letters read) of visual acuity measurement”. In: *Ophthalmic and Physiological Optics* 36.4, pp. 355–358.

Ernst, M. O. and M. S. Banks (2002). “Humans integrate visual and haptic information in a statistically optimal fashion”. In: *Nature (london)* 415.6870, pp. 429–433.

Ernst, M. O. and H. H. Bülthoff (2004). “Merging the senses into a robust percept”. In: *Trends in Cognitive Sciences* 8.4, pp. 162–169.

Favrot, S. E. and J. Buchholz (2010). “LoRA: A Loudspeaker-Based Room Auralization System”. In: *Acustica United With Acta Acustica* 96.2, pp. 364–375.

Fendrich, R. and P. M. Corballis (2001). “The temporal cross-capture of audition and vision”. eng. In: *Perception and Psychophysics* 63.4, pp. 719–725.

Fleming, J. T., A. L. Noyce, and B. G. Shinn-Cunningham (2020). “Audio-visual spatial alignment improves integration in the presence of a competing audiovisual stimulus”. In: *Neuropsychologia* 146.September 2019, p. 107530.

Forster, B., C. Cavina-Pratesi, S. M. Aglioti, and G. Berlucchi (2002). “Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time”. In: *Experimental Brain Research* 143.4, pp. 480–487.

Foss-Feig, J. H. et al. (2010). “An extended multisensory temporal binding window in autism spectrum disorders”. In: *Experimental Brain Research* 203.2, pp. 381–389.

Fozard, J. L. (1990). “Vision and Hearing in Aging”. In: *Handbook of the Psychology of Aging*. Ed. by J. E. Birren and K. W. Schaie. Third Edit. Academic Press, pp. 150–170.

Frank, M., F. Zotter, and A. Sontacchi (2008). “Localization experiments using different 2D ambisonics Decoders”. In: *25th Tonmeistertagung – VDT International Convention*.

Frassinetti, F., N. Bolognini, D. Bottari, A. Bonora, and E. Làdavas (2005). “Audiovisual Integration in Patients with Visual Deficit”. In: *Journal of Cognitive Neuroscience* 17.9, pp. 1442–1452.

Frassinetti, F., F. Pavani, and E. Làdavas (2002). “Acoustical vision of neglected stimuli: Interaction among spatially converging audiovisual inputs in neglect patients”. In: *Journal of Cognitive Neuroscience* 14.1, pp. 62–69.

Freeman, L. C. A., K. C. Wood, and J. K. Bizley (2018). “Multisensory stimuli improve relative localisation judgments compared to unisensory auditory or visual stimuli”. In: *The Journal of the Acoustical Society of America* 143.6.

Frens, M. A., A. J. Van Opstal, and R. F. Van Der Willigen (1995). “Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements”. In: *Perception & Psychophysics* 57.6, pp. 802–816.

Garcia, S. E., P. R. Jones, G. S. Rubin, and M. Nardini (2017). “Auditory Localisation Biases Increase with Sensory Uncertainty”. In: *Scientific Reports* 7.1, p. 40567.

Gatehouse, S. and I. Noble (2004). “The Speech, Spatial and Qualities of Hearing Scale (SSQ)”. In: *International Journal of Audiology* 43.2, pp. 85–99.

Genovese, A., G. Zalles, G. Reardon, and A. Roginska (2018). “Acoustic perturbations in HRTFs measured on Mixed Reality Headsets”. In: *Proceedings of the Aes International Conference — 2018*. Audio Engineering Society, pp. 268–283.

Gerzon, M. A. (1973). “Periphony: With-Height Sound Reproduction.” In: *AES: Journal of the Audio Engineering Society* 21.1, pp. 2–10.

Gieseler, A., M. A. Tahden, C. M. Thiel, and H. Colonius (2018). “Does hearing aid use affect audiovisual integration in mild hearing impairment?” In: *Experimental Brain Research* 236.4, pp. 1161–1179.

Glyde, H., S. Cameron, H. Dillon, L. Hickson, and M. Seeto (2013). “The effects of hearing impairment and aging on spatial processing”. In: *Ear and Hearing* 34.1, pp. 15–28.

Godfroy-Cooper, M., P. M. Sandor, J. D. Miller, and R. B. Welch (2015). “The interaction of vision and audition in two-dimensional space”. In: *Frontiers in Neuroscience* 9.SEP.

Godfroy, M., C. Roumes, and P. Dauchy (2003). “Spatial variations of visual - Auditory fusion areas”. In: *Perception* 32.10, pp. 1233–1245.

Gondan, M., B. Niederhaus, F. Rösler, and B. Röder (2005). “Multisensory processing in the redundant-target effect: A behavioral and event-related potential study”. In: *Perception and Psychophysics* 67.4, pp. 713–726.

Gori, M., G. Sandini, C. Martinoli, and D. C. Burr (2014). “Impairment of auditory spatial localization in congenitally blind human subjects”. In: *Brain* 137.1, pp. 288–293.

Gupta, R., R. Ranjan, J. He, and W. S. Gan (2018). “Investigation of effect of VR/AR headgear on Head related transfer functions for natural listening”. In: *Proceedings of the Aes International Conference — 2018*. Audio Engineering Society, pp. 130–139.

Hairston, W. D., M. T. Wallace, J. W. Vaughan, B. E. Stein, J. L. Norris, and J. A. Schirillo (2003). “Visual Localization Ability Influences Cross-Modal Bias”. In: *Journal of Cognitive Neuroscience* 15.1, pp. 20–29.

Hartmann, W. M. (1999). “How we localize sound”. In: *Physics Today* 52.11, pp. 24–29.

Haß, K. et al. (2017). “Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion”. In: *Cognitive Neuropsychiatry* 22.2, pp. 145–158.

Häusler, R., S. Colburn, and E. Marr (1983). “Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests”. In: *Acta Oto-Laryngologica* 96.SUPPL. 400.

Heffner, H. E. and R. S. Heffner (2016). “The Evolution of Mammalian Sound Localization”. In: *Acoustical Today* 12.1.

Heffner, R. S. and H. E. Heffner (1992). “Visual factors in sound localization in mammals”. In: *The Journal of Comparative Neurology* 317.3, pp. 219–232.

Hein, G., O. Doehrmann, N. G. Müller, J. Kaiser, L. Muckli, and M. J. Naumer (2007). “Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas”. In: *Journal of Neuroscience* 27.30, pp. 7881–7887.

Helbig, H. B. and M. O. Ernst (2007). “Optimal integration of shape information from vision and touch”. In: *Experimental Brain Research* 179.4, pp. 595–606.

Heller, M. A. (1982). “Visual and tactual texture perception: Intersensory cooperation”. In: *Perception & Psychophysics* 31.4, pp. 339–344.

Hendrix, C. and W. Barfield (1995). “Presence in virtual environments as a function of visual and auditory cues”. In: *Proceedings - Virtual Reality Annual International Symposium*, pp. 74–82.

Hernández, B., A. Setti, R. A. Kenny, and F. N. Newell (2019). “Individual differences in ageing, cognitive status, and sex on susceptibility to the sound-induced flash illusion: A large-scale study”. In: *Psychology and Aging* 34.7, pp. 978–990.

Heron, J., D. Whitaker, and P. V. McGraw (2004). “Sensory uncertainty governs the extent of audio-visual interaction”. In: *Vision Research* 44.25, pp. 2875–2884.

Hirst, R. J., A. Setti, R. A. Kenny, and F. N. Newell (2019). “Age-related sensory decline mediates the Sound-Induced Flash Illusion: Evidence for reliability weighting models of multisensory perception”. In: *Scientific Reports* 9.1, pp. 1–12.

Hirst, R. J. et al. (2020). “The effect of eye disease, cataract surgery and hearing aid use on multisensory integration in ageing”. In: *Cortex* 133, pp. 161–176.

Howard, I. P. and W. B. Templeton (1966). *Human Spatial Orientation.* Oxford, England: Wiley.

Hruby, F. (2019). “The Sound of Being There: Audiovisual Cartography with Immersive Virtual Environments”. In: *KN - Journal of Cartography and Geographic Information* 69.1, pp. 19–28.

Hunsaker, A. and E. Hargittai (2018). “A review of Internet use among older adults”. In: *New Media & Society* 20.10, pp. 3937–3954.

Huyse, A., J. Leybaert, and F. Berthommier (2014). “Effects of aging on audiovisual speech integration”. In: *The Journal of the Acoustical Society of America* 136.4, pp. 1918–1931.

Jack, C. E. and W. R. Thurlow (1973). “Effects of Degree of Visual Association and Angle of Displacement on the Ventriloquism effect”. In: *Perceptual and Motor Skills* 37.3, pp. 967–979.

Jackson, C. V. (1953). “Visual Factors in Auditory Localization”. In: *Quarterly Journal of Experimental Psychology* 5.2, pp. 52–65.

Jensen, N. S., T. Neher, S. Laugesen, R. B. Johannesson, and L. Kragelund (2013). “Laboratory and field study of the potential benefits of pinna cue-preserving hearing aids”. In: *Trends in Amplification* 17.3, pp. 171–188.

Jones, S. A., U. Beierholm, D. Meijer, and U. Noppeney (2019). “Older adults sacrifice response speed to preserve multisensory integration performance”. In: *Neurobiology of Aging* 84, pp. 148–157.

Keidser, G., A. O’Brien, J. U. Hain, M. McLelland, and I. Yeend (2009). “The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users”. In: *International Journal of Audiology* 48.11, pp. 789–803.

Keidser, G. et al. (2006). “The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers”. In: *International Journal of Audiology* 45.10, pp. 563–579.

Kern, A. C. and W. Ellermeier (2020). “Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence”. In: *Frontiers in Robotics and AI* 7.February, pp. 1–13.

Kessling, P. and T. Görne (2018). “Studio for immersive media research and production: Immersive Audio Lab at HAW Hamburg”. In: *145th Audio Engineering Society International Convention, Aes 2018*. Audio Engineering Society.

Klemm, O (1909). “Lokalisation von Sinneseindriicken bei disparaten Nebenreizen.” In: *Psychol. Stud.* 5, pp. 73–162.

Koppen, C., A. Alsius, and C. Spence (2008). “Semantic congruency and the Colavita visual dominance effect”. In: *Experimental Brain Research* 184.4, pp. 533–546.

Körding, K. P., U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum, and L. Shams (2007). “Causal inference in multisensory perception”. In: *PLoS ONE* 2.9.

Kramer, A., B. Röder, and P. Bruns (2020). “Feedback Modulates Audio-Visual Spatial Recalibration”. In: *Frontiers in Integrative Neuroscience* 13.January, pp. 1–15.

Kuling, I. A., A. Kohlrausch, and J. F. Juola (2013). “Quantifying temporal ventriloquism in audiovisual synchrony perception”. In: *Attention, Perception, and Psychophysics* 75.7, pp. 1583–1599.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2017). “lmerTest Package: Tests in Linear Mixed Effects Models”. In: *Journal of Statistical Software, Articles* 82.13, pp. 1–26.

Larsson, P. and D. Västfjäll (2007). “When what you hear is what you see: Presence and auditory-visual integration in virtual environments”. In: *Proceedings of the 10th Annual International Workshop on Presence*. September 2016, pp. 11–18.

Laurienti, P. J., J. H. Burdette, J. A. Maldjian, and M. T. Wallace (2006). “Enhanced multisensory integration in older adults”. In: *Neurobiology of Aging* 27.8, pp. 1155–1163.

Laurienti, P. J., R. A. Kraft, J. A. Maldjian, J. H. Burdette, and M. T. Wallace (2004). “Semantic congruence is a critical factor in multisensory behavioral performance”. In: *Experimental Brain Research* 158.4, pp. 405–414.

Leo, F., N. Bolognini, C. Passamonti, B. E. Stein, and E. Làdavas (2008). “Cross-modal localization in hemianopia: New insights on multisensory integration”. In: *Brain* 131.3, pp. 855–865.

Levitt, H. (1971). “Transformed Up-Down Methods in Psychoacoustics”. In: *The Journal of the Acoustical Society of America* 49.2B, pp. 467–477.

Lewald, J. and R. Guski (2003). “Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli”. In: *Cognitive Brain Research* 16.3, pp. 468–478.

Lippert, M., N. K. Logothetis, and C. Kayser (2007). “Improvement of visual contrast detection by a simultaneous sound”. In: *Brain Research* 1173, pp. 102–109.

Lorenzi, C., S. Gatehouse, and C. Lever (1999). “Sound localization in noise in hearing-impaired listeners”. In: *The Journal of the Acoustical Society of America* 105.6, pp. 3454–3463.

Lovelace, C. T., B. E. Stein, and M. T. Wallace (2003). “An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection”. In: *Cognitive Brain Research* 17.2, pp. 447–453.

Magnotti, J. F., W. J. Ma, and M. S. Beauchamp (2013). “Causal inference of asynchronous audiovisual speech”. In: *Frontiers in Psychology* 4.NOV, pp. 1–10.

Mahoney, J. R., P. C. C. Li, M. Oh-Park, J. Verghese, and R. Holtzer (2011). "Multisensory integration across the senses in young and old adults". In: *Brain Research* 1426, pp. 43–53.

Makous, J. C. and J. C. Middlebrooks (1990). "Two-dimensional sound localization by human listeners". In: *Journal of the Acoustical Society of America* 87.5, pp. 2188–2200.

Maneuvrier, A., L. M. Decker, H. Ceyte, P. Fleury, and P. Renaud (2020). "Presence Promotes Performance on a Virtual Spatial Cognition Task: Impact of Human Factors on Virtual Reality Assessment". In: *Frontiers in Virtual Reality* 1, p. 16.

Mateeff, S. and A. Gourevich (1983). "Peripheral vision and perceived visual direction". In: *Biological Cybernetics* 49.2, pp. 111–118.

McGovern, D. P., E. Roudaia, J. Stapleton, T. M. McGinnity, and F. N. Newell (2014). "The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration". In: *Frontiers in Aging Neuroscience* 6.SEP, pp. 1–9.

McIntire, J. P., P. R. Havig, S. N. Watamaniuk, and R. H. Gilkey (2010). "Visual search performance with 3-D auditory cues: Effects of motion, target location, and practice". In: *Human Factors* 52.1, pp. 41–53.

Mégevand, P., S. Molholm, A. Nayak, and J. J. Foxe (2013). "Recalibration of the Multisensory Temporal Window of Integration Results from Changing Task Demands". In: *PLoS ONE* 8.8.

Meijer, D., S. Veselič, C. Calafiore, and U. Noppeney (2019). "Integration of audio-visual spatial signals is not consistent with maximum likelihood estimation". In: *Cortex* 119, pp. 74–88.

Meredith, M. A. (2002). "On the neuronal basis for multisensory convergence: A brief overview". In: *Cognitive Brain Research* 14.1, pp. 31–40.

Miller, J. (1982). “Divided attention: Evidence for coactivation with redundant signals”. In: *Cognitive Psychology* 14.2, pp. 247–279.

Miller, J. (1986). “Timecourse of coactivation in bimodal divided attention”. In: *Perception & Psychophysics* 40.5, pp. 331–343.

Molholm, S., W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe (2002). “Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study”. In: *Cognitive Brain Research* 14.1, pp. 115–128.

Montagne, C. and Y. Zhou (2016). “Visual capture of a stereo sound: Interactions between cue reliability, sound localization variability, and cross-modal bias”. In: *The Journal of the Acoustical Society of America* 140.1, pp. 471–485.

Morein-Zamir, S., S. Soto-Faraco, and A. Kingstone (2003). “Auditory capture of vision: Examining temporal ventriloquism”. eng. In: *Cognitive Brain Research* 17.1, pp. 154–163.

Moro, S. S., L. R. Harris, and J. K. Steeves (2014). “Optimal audiovisual integration in people with one eye”. In: *Multisensory Research* 27.3-4, pp. 173–188.

Moro, S. S. and J. K. Steeves (2018). “Audiovisual plasticity following early abnormal visual experience: Reduced McGurk effect in people with one eye”. In: *Neuroscience Letters* 672.December 2017, pp. 103–107.

Mozolic, J. L., C. E. Hugenschmidt, A. M. Peiffer, and P. J. Laurienti (2008). “Modality-specific selective attention attenuates multisensory integration”. In: *Experimental Brain Research* 184.1, pp. 39–52.

Murray, M. M. et al. (2005). “Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment”. In: *Cerebral Cortex* 15.7, pp. 963–974.

Musacchia, G., L. Arum, T. Nicol, D. Garstecki, and N. Kraus (2009). “Audiovisual deficits in older adults with hearing loss: Biological evidence”. In: *Ear and Hearing* 30.5, pp. 505–514.

Musicant, A. D. and R. A. Butler (1984). “The influence of pinnae-based spectral cues on sound localization”. In: *Journal of the Acoustical Society of America* 75.4, pp. 1195–1200.

Narinesingh, C., H. C. Goltz, R. A. Raashid, and A. M. Wong (2015). “Developmental Trajectory of McGurk Effect Susceptibility in Children and Adults With Amblyopia”. In: *Investigative ophthalmology & visual science* 56.3, pp. 2107–2113.

Narinesingh, C., H. C. Goltz, and A. M. Wong (2017). “Temporal binding window of the sound-induced flash illusion in amblyopia”. In: *Investigative Ophthalmology and Visual Science* 58.3, pp. 1442–1448.

Navarra, J., A. Vatakis, M. Zampini, S. Soto-Faraco, W. Humphreys, and C. Spence (2005). “Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration”. In: *Cognitive Brain Research* 25.2, pp. 499–507.

Nelson, W. T., L. J. Hettinger, J. A. Cunningham, B. J. Brickman, M. W. Haas, and R. L. McKinley (1998). “Effects of Localized Auditory Information on Visual Target Detection Performance Using a Helmet-Mounted Display”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40.3, pp. 452–460.

Noble, W., D. Byrne, and B. Lepage (1994). “Effects on sound localization of configuration and type of hearing impairment”. In: *Journal of the Acoustical Society of America* 95.2, pp. 992–1005.

Noble, W., D. Byrne, and K. Ter-Horst (1997). “Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners”. In: *The Journal of the Acoustical Society of America* 102.4, pp. 2343–2352.

Noel, J. P., K. Modi, M. T. Wallace, and N. Van der Stoep (2018). “Audiovisual integration in depth: multisensory binding and gain as a function of distance”. In: *Experimental Brain Research* 236.7, pp. 1939–1951.

Noesselt, T., D Bergmann, M. Hake, H Heinze, and R Fendrich (2008). “Sound increases the saliency of visual events”. In: *Brain Research* 1220, pp. 157–163.

Nordahl, R. (2004). “Self-induced footsteps sounds in virtual reality: Latency, recognition, quality and presence”. In: *Proceedings of the Eight Annual International Workshop Presence, London, UK* 5, pp. 5–6.

Ocklenburg, S., M. Hirnstein, M. Hausmann, and J. Lewald (2010). “Auditory space perception in left- and right-handers”. In: *Brain and Cognition* 72.2, pp. 210–217.

Odegaard, B. and L. Shams (2016). “The Brain’s Tendency to Bind Audiovisual Signals Is Stable but Not General”. In: *Psychological Science* 27.4, pp. 583–591.

Odegaard, B., D. R. Wozny, and L. Shams (2015). “Biases in Visual, Auditory, and Audiovisual Perception of Space”. In: *PLoS Computational Biology* 11.12, pp. 1–23.

Odgaard, E, Y Arieh, and L Marks (2003). “Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias”. In: *Perception \& Psychophysics* 65, pp. 123–132.

Odgaard, E. C., Y. Arieh, and L. E. Marks (2004). “Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation”. In: *Cognitive, Affective and Behavioral Neuroscience* 4.2, pp. 127–132.

Olsen, S. O., L. H. Hernvig, and L. H. Nielsen (2012). “Self-reported hearing performance among subjects with unilateral sensorineural hearing loss”. In: *Audiological Medicine* 10.2, pp. 83–92.

Oreinos, C. and J. M. Buchholz (2015). “Objective analysis of ambisonics for hearing aid applications: Effect of listener’s head, room reverberation, and directional microphones”. In: *The Journal of the Acoustical Society of America* 137.6, pp. 3447–3465.

Otte, R. J., M. J. Agterberg, M. M. Van Wanrooij, A. F. Snik, and A. J. Van Opstal (2013). “Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance”. In: *JARO - Journal of the Association for Research in Otolaryngology* 14.2, pp. 261–273.

Pápai, M. S. and S. Soto-Faraco (2017). “Sounds can boost the awareness of visual events through attention without cross-modal integration”. In: *Scientific Reports* 7.January, pp. 1–13.

Parise, C. V., C. Spence, and M. O. Ernst (2012). “When correlation implies causation in multisensory integration”. In: *Current Biology* 22.1, pp. 46–49.

Pavani, F. et al. (2017). “Spatial and non-spatial multisensory cueing in unilateral cochlear implant users”. In: *Hearing Research* 344, pp. 24–37.

Peiffer, A. M., J. L. Mozolic, C. E. Hugenschmidt, and P. J. Laurienti (2007). “Age-related multisensory enhancement in a simple audiovisual detection task”. In: *NeuroReport* 18.10, pp. 1077–1081.

Perrott, D, T Sadralodabai, K Saberi, and T Strybel (1991). “Aurally Aided Visual Search in the Central Visual Field: Effects of Visual Load and Visual Enhancement of the Target”. In: *Human Factors: The Journal of Human Factors and Ergonomics Society* 33, pp. 389–400.

Perrott, D. R., K. Saberi, K. Brown, and T. Z. Strybel (1990). “Auditory psychomotor coordination and visual search performance”. In: *Perception & Psychophysics* 48.3, pp. 214–226.

Pick, H. L., D. H. Warren, and J. C. Hay (1969). “Sensory conflict in judgments of spatial direction”. In: *Perception & Psychophysics* 6.4, pp. 203–205.

Plass, J., E. Guzman-Martinez, L. Ortega, S. Suzuki, and M. Grabowecky (2017). “Automatic auditory disambiguation of visual awareness”. In: *Attention, Perception, and Psychophysics* 79.7, pp. 2055–2063.

Pomper, U. and M. Chait (2017). “The impact of visual gaze direction on auditory object tracking”. In: *Scientific Reports* 7.1, pp. 1–16.

Powers, A. R., A. Hillock-Dunn, and M. T. Wallace (2016). “Generalization of multisensory perceptual learning”. In: *Scientific Reports* 6.March, pp. 1–9.

Powers, A. R., A. R. Hillock, and M. T. Wallace (2009). “Perceptual training narrows the temporal window of multisensory binding”. In: *Journal of Neuroscience* 29.39, pp. 12265–12274.

Pulkki, V. and T. Hirvonen (2005). “Localization of virtual sources in multichannel audio reproduction”. In: *IEEE Transactions on Speech and Audio Processing* 13.1, pp. 105–119.

Puschmann, S., P. Sandmann, A. Bendixen, and C. M. Thiel (2014). “Age-related hearing loss increases cross-modal distractibility”. In: *Hearing Research* 316, pp. 28–36.

Puschmann, S. and C. M. Thiel (2017). “Changed crossmodal functional connectivity in older adults with hearing loss”. In: *Cortex* 86, pp. 109–122.

Puschmann, S. et al. (2019). “Hearing-impaired listeners show increased audiovisual benefit when listening to speech in noise”. In: *NeuroImage* 196.April, pp. 261–268.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Raab, D. H. (1962). "Statistical facilitation of simple reaction times". In: *Transactions of the New York Academy of Sciences* 24.5 Series II, pp. 574–590.

Radeau, M. and P. Bertelson (1977). "Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations". In: *Perception & Psychophysics* 22.2, pp. 137–146.

Rakerd, B., T. J. Vander Velde, and W. M. Hartmann (1998). "Sound localization in the median sagittal plane by listeners with presbyacusis." In: *Journal of the American Academy of Audiology* 9.6, pp. 466–479.

Rebenitsch, L. and C. Owen (2016). "Review on cybersickness in applications and visual displays". In: *Virtual Reality* 20.2, pp. 101–125.

Renner, R. S., B. M. Velichkovsky, and J. R. Helmert (2013). "The perception of egocentric distances in virtual environments - A review". In: *ACM Computing Surveys* 46.2, pp. 1–40.

Richards, M. D., H. C. Goltz, and A. M. Wong (2018). "Optimal audiovisual integration in the ventriloquism effect but pervasive deficits in unisensory spatial localization in amblyopia". In: *Investigative Ophthalmology and Visual Science* 59.1, pp. 122–131.

Richards, M., H. Goltz, and A. Wong (2016). "The audiovisual temporal binding window in unilateral amblyopia: monocular and binocular effects". In: *Journal of Vision* 16.12, p. 143.

Riecke, B. E., A. Väljamäe, and J. Schulte-Pelkum (2009). "Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality". In: *ACM Transactions on Applied Perception* 6.2.

Roach, N. W., J. Heron, and P. V. McGraw (2006). "Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration".

In: *Proceedings of the Royal Society B: Biological Sciences* 273.1598, pp. 2159–2168.

Rohe, T. and U. Noppeney (2015). “Sensory reliability shapes perceptual inference via two mechanisms”. In: *Journal of Vision* 15.5, pp. 1–16.

Rohe, T. and U. Noppeney (2018). “Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Attention”. In: *eneuro* 5.1, e0315–17.2018.

Rosemann, S. and C. M. Thiel (2018). “Audio-visual speech processing in age-related hearing loss: Stronger integration and increased frontal lobe recruitment”. In: *Neuroimage* 175, pp. 425–437.

Rudmann, D and T Strybel (1999). “Auditory Spatial Facilitation of Visual Search Performance: Effect of Cue Precision and Distractor Density”. In: *Human Factors: The Journal of Human Factors and Ergonomics Society* 41, pp. 146–160.

Sampedro Llopis, H., F. K. Pind Jörgensson, and C.-H. Jeong (2019). “Effects of the order of Ambisonics on localization for different reverberant conditions in a novel 3D acoustic virtual reality system”. In: *Proceedings of 23rd International Congress on Acoustics* 2019-, pp. 747–754.

Schormans, A. L. and B. L. Allman (2018). “Behavioral plasticity of audiovisual perception: rapid recalibration of temporal sensitivity but not perceptual binding following adult-onset hearing loss”. In: *Frontiers in Behavioral Neuroscience* 12.October, pp. 1–18.

Schröger, E. and A. Widmann (1998). “Speeded responses to audiovisual signal changes result from bimodal integration”. In: *Psychophysiology* 35.6, pp. 755–759.

Schulte, A., C. M. Thiel, A. Gieseler, M. Tahden, H. Colonius, and S. Rosemann (2020). “Reduced resting state functional connectivity with increasing age-related hearing loss and McGurk susceptibility”. In: *Scientific Reports* 10.1, p. 16987.

Schwarz, G. (1978). “Estimating the dimension of a model”. In: *Annals of Statistics* 6.2, pp. 461–464.

Schwind, V., S. Mayer, A. Comeau-Vermeersch, R. Schweigert, and N. Henze (2018). “Up to the Finger Tip: The effect of Avatars on Mid-Air Pointing Accuracy in Virtual Reality”. In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. New York, NY, USA: ACM, pp. 477–488.

Scurry, A. N., T. Vercillo, A. Nicholson, M. Webster, and F. Jiang (2020). “Aging Impairs Temporal Sensitivity, but not Perceptual Synchrony, Across Modalities”. In: *Multisensory Researchu* 32.8, pp. 671–692.

Seifert, A. and A. Schlomann (2021). “The Use of Virtual and Augmented Reality by Older Adults: Potentials and Challenges”. In: *Frontiers in Virtual Reality* 2.April, pp. 1–5.

Seifert, A., A. Schlomann, C. Rietz, and H. R. Schelling (2017). “The use of mobile devices for physical activity tracking in older adults’ everyday life”. In: *DIGITAL HEALTH* 3, p. 2055207617740088.

Setti, A., K. E. Burke, R. A. Kenny, and F. N. Newell (2011). “Is inefficient multisensory processing associated with falls in older people?” In: *Experimental Brain Research* 209.3, pp. 375–384.

Setti, A., J. Stapleton, D. Leahy, C. Walsh, R. A. Kenny, and F. N. Newell (2014). “Improving the efficiency of multisensory integration in older adults: Audiovisual temporal discrimination training reduces susceptibility to the sound-induced flash illusion”. In: *Neuropsychologia* 61.1, pp. 259–268.

Shams, L., Y. Kamitani, and S. Shimojo (2000). “What you see is what you hear”. In: *Nature* 408, pp. 788–788.

Shams, L., Y. Kamitani, and S. Shimojo (2002). "Visual illusion induced by sound". In: *Cognitive Brain Research* 14.1, pp. 147–152.

Shipley, T (1964). "Auditory flutter-driving of visual flicker". eng. In: *Science* 145.363.

Simon, J. R. (1990). "The Effects of an Irrelevant Directional Cue on Human Information Processing". In: *Advances in Psychology* 65, pp. 31–86.

Slutsky, D. A. and G. H. Recanzone (2001). "Temporal and spatial dependency, of the ventriloquism effect". In: *NeuroReport* 12.1, pp. 7–10.

Stauffert, J.-P., F. Niebling, and M. E. Latoschik (2020). "Latency and Cybersickness: Impact, Causes, and Measures. A Review". In: *Frontiers in Virtual Reality* 1, p. 31.

Stawicki, M., P. Majdak, and D. Baskent (2019). "Ventriloquist Illusion Produced with Virtual Acoustic Spatial Cues and Asynchronous Audiovisual Stimuli in Both Young and Older Individuals". In: *Multisensory Research* 32.8, pp. 745–770.

Stein, B. E., M. A. Meredith, W. S. Huneycutt, and L. McDade (1989). "Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli". In: *Journal of Cognitive Neuroscience* 1.1, pp. 12–24.

Stein, B. E., N. London, L. K. Wilkinson, and D. D. Price (1996). "Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis". In: *Journal of Cognitive Neuroscience* 8.6, pp. 497–506.

Stein, B. E. and M. A. Meredith (1993). *The merging of the senses.* MIT Press, xv+211p.

Steinweg, B. and F. W. Mast (2017). "Semantic incongruity influences response caution in audio-visual integration". In: *Experimental Brain Research* 235.1, pp. 349–363.

Stenzel, H. (2017). “Modeling horizontal audio-visual coherence with the psychometric function”. In: *142nd Audio Engineering Society International Convention 2017, Aes 2017*. Audio Engineering Society.

Stenzel, H., J. Francombe, and P. J. Jackson (2019). “Limits of perceived audio-visual spatial coherence as defined by reaction time measurements”. In: *Frontiers in Neuroscience* 13.MAY, pp. 1–17.

Stitt, P., S. Bertet, and M. Van Walstijn (2013). “Perceptual investigation of image placement with ambisonics for non-centred listeners”. In: *Dafx 2013 - 16th International Conference on Digital Audio Effects*.

Stitt, P., S. Bertet, and M. Van Walstijn (2014). “Off-centre localisation performance of ambisonics and HOA for large and small loudspeaker array radii”. In: *Acta Acustica United With Acustica* 100.5, pp. 937–944.

Stropahl, M. and S. Debener (2017). “Auditory cross-modal reorganization in cochlear implant users indicates audio-visual integration”. In: *NeuroImage: Clinical* 16.August, pp. 514–523.

Suárez, A. S., N. Kaplanis, S. Serafin, and S. Bech (2019). “In-virtualis: A study on the impact of congruent virtual reality environments in perceptual audio evaluation of loudspeakers”. In: *Proceedings of the Aes International Conference*. Audio Engineering Society.

Tabry, V., R. J. Zatorre, and P. Voss (2013). “The influence of vision on sound localization abilities in both the horizontal and vertical planes”. In: *Frontiers in Psychology* 4.DEC.

Talsma, D., D. Senkowski, S. Soto-Faraco, and M. G. Woldorff (2010). “The multifaceted interplay between attention and multisensory integration”. In: *Trends in Cognitive Sciences* 14.9, pp. 400–410.

Tannen, R. S., W. T. Nelson, R. S. Bolia, J. S. Warm, and W. N. Dember (2004). "Evaluating adaptive multisensory displays for target localization in a flight task". In: *International Journal of Aviation Psychology* 14.3, pp. 297–312.

Tannen, R. S. et al. (2000). "Adaptive Integration of Head-Coupled Multi-Sensory Displays for Target Localization". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44.13, pp. 77–80.

Tastevin, J (1937). "En partant de l'experience d'Aristote". In: *L'encéphale* 1.57-84, pp. 57–84.

Taylor, K. I., H. E. Moss, E. A. Stamatakis, and L. K. Tyler (2006). "Binding cross-modal object features in perirhinal cortex". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.21, pp. 8239–8244.

Thomas, J. P. and M. Shiffrar (2013). "Meaningful sounds enhance visual sensitivity to human gait regardless of synchrony". In: *Journal of Vision* 13.14, pp. 1–13.

Thresh, L., C. Armstrong, and G. Kearney (2017). "A direct comparison of localisation performance when using first, third and fifth order ambisonics for real loudspeaker and virtual loudspeaker rendering." In: *143rd Audio Engineering Society Convention 2017, Aes 2017* 1, pp. 489–497.

Thurlow, W. R. and C. E. Jack (1973). "Certain Determinants of the "Ventriloquism Effect"". In: *Perceptual and Motor Skills* 36.3, pp. 1171–1184.

Tye-Murray, N., M. S. Sommers, and B. Spehar (2007). "Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing". In: *Ear and Hearing* 28.5, pp. 656–668.

Ulrich, R., J. Miller, and H. Schröter (2007). "Testing the race model inequality: An algorithm and computer programs". In: 39.2, pp. 291–302.

Van Den Bogaert, T., E. Carette, and J. Wouters (2011). “Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna”. In: *International Journal of Audiology* 50.3, pp. 164–176.

Van Eijk, R. L., A. Kohlrausch, J. F. Juola, and S. Van De Par (2008). “Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type”. In: *Perception and Psychophysics* 70.6, pp. 955–968.

Van den Bogaert, T., T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters (2006). “Horizontal localization with bilateral hearing aids: Without is better than with”. In: *The Journal of the Acoustical Society of America.*

Van der Burg, E., C. N. Olivers, A. W. Bronkhorst, and J. Theeuwes (2008). “Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search”. In: *Journal of Experimental Psychology: Human Perception and Performance* 34.5, pp. 1053–1065.

Vatakis, A. and C. Spence (2007). “Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli”. In: *Perception and Psychophysics* 69.5, pp. 744–756.

Vatakis, A. and C. Spence (2008). “Evaluating the influence of the ’unity assumption’ on the temporal perception of realistic audiovisual stimuli”. In: *Acta Psychologica* 127.1, pp. 12–23.

Venskytis, E. J., C. Clayton, C. Montagne, and Y. Zhou (2019). “Audiovisual Interactions in Stereo Sound Localization for Individuals With Unilateral Hearing Loss”. In: *Trends in Hearing* 23, pp. 1–17.

Vroomen, J., P. Bertelson, and B. De Gelder (2001). “The ventriloquist effect does not depend on the direction of automatic visual attention”. In: *Perception and Psychophysics* 63.4, pp. 651–659.

Vroomen, J. and J. J. Stekelenburg (2014). “A bias-free two-alternative forced choice procedure to examine intersensory illusions applied to the ventriloquist effect by flashes and averted eye-gazes”. In: *European Journal of Neuroscience* 39.9, pp. 1491–1498.

Vu, K. P. L., T. Z. Strybel, and R. W. Proctor (2006). “Effects of displacement magnitude and direction of auditory cues on auditory spatial facilitation of visual search”. In: *Human Factors* 48.3, pp. 587–599.

WHO (2021). “World report on hearing”. In: *Geneva: World Health Organization*, Licence: CC BY–NC–SA 3.0 IGO.

Wan, M., C. Narinesingh, H. C. Goltz, M. Chandrakumar, and A. M. F. Wong (2014). “Integrated Visual-Auditory Perception in Amblyopic Adults: A Study Using the McGurk Phenomenon”. In: *Investigative Ophthalmology and Visual Science* 55.13, p. 817.

Wang, B. et al. (2018). “Increased functional brain network efficiency during audiovisual temporal asynchrony integration task in aging”. In: *Frontiers in Aging Neuroscience* 10.OCT, pp. 1–15.

Warren, D. H., R. B. Welch, and T. J. McCarthy (1981). “The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses”. In: *Perception & Psychophysics* 30.6, pp. 557–564.

Wassenhove, V. van, K. W. Grant, and D. Poeppel (2007). “Temporal window of integration in auditory-visual speech perception”. In: *Neuropsychologia* 45.3, pp. 598–607.

Winkel, K. N. de, M. Katliar, and H. H. Buelthoff (2017). “Causal Inference in Multisensory Heading Estimation”. In: *Plos One* 12.1. Ed. by M. Lappe, e0169676.

Wozny, D. R., U. R. Beierholm, and L. Shams (2010). “Probability matching as a computational strategy used in perception”. In: *PLoS Computational Biology* 6.8.

Yang, J., Q. Li, Y. Gao, and J. Wu (2011). “Task-irrelevant auditory stimuli affect audiovisual integration in a visual attention task: Evidence from event-related potentials". In: *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*. IEEE, pp. 248–253.

Zierul, B., J. Tong, P. Bruns, and B. Röder (2019). "Reduced multisensory integration of self-initiated stimuli". In: *Cognition* 182.

Zou, Z., B. K. Chau, K. H. Ting, and C. C. Chan (2017). "Aging effect on audiovisual integrative processing in spatial discrimination task". In: *Frontiers in Aging Neuroscience* 9.NOV, pp. 1–12.

van Opstal, J. (2016). "Chapter 13 - Multisensory Integration". In: *The Auditory System and Human Sound-Localization Behavior*. Ed. by J. van Opstal. San Diego: Academic Press, pp. 361–392.

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse,* Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.
External examiners: Mark Lutman, Stefan Stenfeld

Vol. 2: *Olaf Strelcyk,* Peripheral auditory processing and speech reception in impaired hearing, 2009.
External examiners: Brian Moore, Kathrin Krumbholz

Vol. 3: *Eric R. Thompson,* Characterizing binaural processing of amplitude-modulated sounds, 2009.
External examiners: Michael Akeroyd, Armin Kohlrausch

Vol. 4: *Tobias Piechowiak,* Spectro-temporal analysis of complex sounds in the human auditory system, 2009.
External examiners: Jesko Verhey, Steven van de Par

Vol. 5: *Jens Bo Nielsen,* Assessment of speech intelligibility in background noise and reverberation, 2009.
External examiners: Björn Hagerman, Ejnar Laukli

Vol. 6: *Helen Connor,* Hearing aid amplification at soft input levels, 2010.
External examiners: Inga Holube, Birgitta Larsby

Vol. 7: *Morten Løve Jepsen,* Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.

External examiners: Birger Kollmeier, Ray Meddis

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.
External examiners: David Kemp, Stephen Neely

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.
External examiners: Bernhard Seeber, Michael Vorländer

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.
External examiners: Christopher Plack, Christian Lorenzi

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.
External examiners: Joost Festen, Jürgen Tchorz

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.
External examiners: Bob Burkard, Stephen Neely

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.
External examiners: Stuart Rosen, Christian Lorenzi

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.
External examiners: Michael Stone, Oded Ghitza

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.
External examiners: John Culling, Martin Cooke

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.
External examiners: Lawrence Rosenblum, Matthias Gondan

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.
External examiners: Shihab Shamma, Guy Brown

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.
External examiners: Sascha Spors, Ville Pulkki

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.
External examiners: Bernhard Seeber, Steven van de Par

Vol. 20: *Michał Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
External examiners: Christopher Plack, Enrique Lopez-Poveda

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.
External examiners: Steven van de Par, John Culling

Vol. 22: *Federica Bianchi*, Pitch representations in the impaired auditory system and implications for music perception, 2016.
External examiners: Ingrid Johnsrude, Christian Lorenzi

Vol. 23: *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.
External examiners: Judy Dubno, Martin Cooke

Vol. 24: *Johannes Käsbach*, Characterizing apparent source width perception, 2016.
External examiners: William Whitmer, Jürgen Tchorz

Vol. 25: *Gusztáv Löcsei*, Lateralized speech perception with normal and impaired hearing, 2016.
External examiners: Thomas Brand, Armin Kohlrausch

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.
External examiners: Laurel Carney, Bob Carlyon

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.
External examiners: Volker Hohmann, Piotr Majdak

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.
External examiners: Maria Chait, Christian Lorenzi

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.
External examiners: Ville Pulkki, Pavel Zahorik

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.
External examiners: Roland Schaette, Ian Bruce

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing-impaired listeners, 2018.
External examiners: Enrique Lopez-Poveda, Tim Jürgens

Vol. 32: *Alan Wiinberg*, Perceptual effects of non-linear hearing aid amplification strategies, 2018.
External examiners: Armin Kohlrausch, James Kates

Vol. 33: *Thomas Bentsen*, Computational speech segregation inspired by principles of auditory processing, 2018.
External examiners: Stefan Bleeck, Jürgen Tchorz

Vol. 34: *François Guérit*, Temporal charge interactions in cochlear implant listeners, 2018.
External examiners: Julie Arenberg, Olivier Macherey

Vol. 35: *Andreu Paredes Gallardo*, Behavioral and objective measures of stream segregation in cochlear implant users, 2018.
External examiners: Christophe Micheyl, Monita Chatterjee

Vol. 36: *Søren Fuglsang*, Characterizing neural mechanisms of attention-driven speech processing, 2019.
External examiners: Shihab Shamma, Maarten de Vos

Vol. 37: *Borys Kowalewski*, Assessing the effects of hearing-aid dynamic-range compression on auditory signal processing and perception, 2019.
External examiners: Brian Moore, Graham Naylor

Vol. 38: *Helia Relaño Iborra*, Predicting speech perception of normal-hearing and hearing-impaired listeners, 2019.
External examiners: Ian Bruce, Armin Kohlrausch

Vol. 39: *Axel Ahrens*, Characterizing auditory and audio-visual perception in virtual environments, 2019.
External examiners: Pavel Zahorik, Piotr Majdak

Vol. 40: *Niclas A. Janssen*, Binaural streaming in cochlear implant patients, 2019.
External examiners: Tim Jürgens, Hamish Innes-Brown

Vol. 41: *Wiebke Lamping*, Improving cochlear implant performance through psychophysical measures, 2019.
External examiners: Dan Gnasia, David Landsberger

Vol. 42: *Antoine Favre-Félix*, Controlling a hearing aid with electrically assessed eye gaze, 2020.
External examiners: Jürgen Tchorz, Graham Naylor

Vol. 43: *Raul Sanches-Lopez*, Clinical auditory profiling and profile-based hearing-aid fitting, 2020.
External examiners: Judy R. Dubno, Pamela E. Souza

Vol. 44: *Juan Camilo Gil Carvajal*, Modeling audiovisual speech perception, 2020.
External examiners: Salvador Soto-Faraco, Kaisa Maria Tippana

Vol. 45: *Charlotte Amalie Emdal Navntoft*, Improving cochlear implant performance with new pulse shapes: a multidisciplinary approach, 2020
External examiners: Andrew Kral, Johannes Frijns

Vol. 46: *Naim Mansour*, Assessing hearing device benefit using virtual sound environments, 2021.
External examiners: Virginia Best, Pavel Zahorik

Vol. 47: *Anna Josefine Munch Sørensen*, The effects of noise and hearing loss on conversational dynamics, 2021.
External examiners: William McAllister Whitmer, Martin Cooke

Vol. 48: *Thirsa Huisman*, The influence of vision on spatial localization in normal-hearing and hearing-impaired listeners, 2021.

External examiners: Steven van de Par, Christopher Stecker

The end.

To be continued…

Hearing-impaired people have been shown to have degraded auditory localization abilities. However, it is unclear how they are affected by this in their daily life when they have access to visual information and self-motion cues that can aid localization.

In a series of five experiments, this thesis demonstrated that visual information strongly influences auditory localization in normal-hearing and hearing-impaired listeners. While auditory localization results of the hearing-impaired people were strongly biased towards visual information, the probability of this shift occurring was not higher in hearing-impaired as compared to normal-hearing people in the same age range.

These results can help better understand some of the challenges that hearing-impaired listeners face with regards to spatial localization and may guide future research on audio-visual localization in hearing-impaired and aided-hearing impaired listeners.

DTU Health Tech
Department of Health Technology

Ørsteds Plads

Building 352

DK-2800 Kgs. Lyngby

Denmark

Tel: (+45) 45 25 39 50

www.dtu.dk