Response Accuracy and Response Time in Multisensory Localization

by

Colton Clayton

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2021 by the
Graduate Supervisory Committee:

Yi Zhou, Chair
Tamiko Azuma
Ayoub Daliri

ARIZONA STATE UNIVERSITY

December 2021

ABSTRACT

Spatial awareness (i.e., the sense of the space that we are in) involves the integration of auditory, visual, vestibular, and proprioceptive sensory information of environmental events. Hearing impairment has negative effects on spatial awareness and can result in deficits in communication and the overall aesthetic experience of life, especially in noisy or reverberant environments. This deficit occurs as hearing impairment reduces the signal strength needed for auditory spatial processing and changes how auditory information is combined with other sensory inputs (e.g., vision). The influence of multisensory processing on spatial awareness in listeners with normal, and impaired hearing is not assessed in clinical evaluations, and patients' everyday sensory experiences are currently not directly measurable.

This dissertation investigated the role of vision in auditory localization in listeners with normal, and impaired hearing in a naturalistic stimulus setting, using natural gaze orienting responses. Experiments examined two behavioral outcomes—response accuracy and response time—based on eye movement in response to simultaneously presented auditory and visual stimuli. The first set of experiments examined the effects of stimulus spatial saliency on response accuracy and response time and the extent of visual dominance in both metrics in auditory localization. The results indicate that vision can significantly influence both the speed and accuracy of auditory localization, especially when auditory stimuli are more ambiguous. The influence of vision is shown for both normal hearing- and hearing-impaired listeners. The second set of experiments examined the effect of frontal visual stimulation on localizing an auditory target presented from in front of or behind a listener. The results show domain-specific effects of visual capture on both response time and response accuracy. These results support previous findings that auditory-visual interactions are not limited by the spatial rule of proximity. These results further suggest the strong influence of vision on both the processing and the decision-making stages of sound source localization for both listeners with normal, and impaired hearing.

DEDICATION

This dissertation is dedicated to my son. Your undying curiosity and excitement to learn inspire me every day.

This work is also dedicated to my wife. You pushed me to pursue my dreams and have kept me grounded and focused through all of this.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Yi Zhou, for her guidance in this project and unwavering support throughout my graduate training from Au.D. to Ph.D. I will forever treasure your counsel that was crucial in finishing this dissertation (and Ph.D.) and will be invaluable in my future career. Committee members Dr. Tamiko Azuma and Dr. Ayoub Daliri have also provided invaluable insights to this dissertation, and I deeply appreciate their thoughtful questions and encouragement. To each of my committee members, thank you for helping me expand my horizons. I would like to thank other key individuals who contributed greatly to this project: Drs. William Yost and Torben Pastore for our journal club discussions that increased my understanding of the field and for your eyes in and brains in reviewing draft after draft; Dr. Andrea Pittman who connected me with many of the participants; The incredible ASU Au.D. students who spent hours in the booth looking for sounds. Thank you all for your time, talents, and kind words.

Finally, I would like to thank my friends and family near and far for their continued love and support. To my wife, Allison, thank you for understanding the many hours spend in the lab or in front of the computer and thank you for the personal sacrifices you have made to support my goals and keep our home running.

TABLE OF CONTENTS

CHAPTER

CHAPTER Page

APPENDIX

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Each day our senses are flooded with a variety of information with the primary task of determining what information is relevant to function, how information should be combined, and what, if any, information can be safely discarded. The successful synthesis of this information is crucial, as oftentimes there will be many cues that need to be processed and integrated to make a single decision or action. An example of multiple cues playing a role in a single decision is found in the task of sound source localization. Accurate localization of a sound source is critical to interacting with our environment and avoiding possible dangers. The task of localizing a sound source is a complex process which is made possible through integrated processing of several localization cues by our nervous system (Yin, 2002). Typically, our nervous system has little problem integrating various cues related to the same source and making accurate spatial localization decisions. As a result, sound localization is a task that can be performed very quickly and without much thought. In contrast, those affected by a sensory deficit such as hearing impairment (HI), show an increase in difficulty and a decrease in performance in sound localization tasks (Häusler, Colburn, & Marr, 1983). These handicaps in localization have significant negative effects on listener safety, communication, and the overall aesthetic experience of life (Byrne & Noble, 1998). While the current remedy to this issue, hearing aids, seeks to restore auditory cues limited by the sensory deficit, hearing aid users continue to report difficulty in sound source localization (Van den Bogaert, Klasen, Moonen, Van Deun, & Wouters, 2006) suggesting a difficulty that goes beyond just audibility. To overcome the deficit that remains by remedying the access to sound alone, a more in-depth understanding of the neural mechanisms used to integrate the cues used in localization is necessary. By first understanding the mechanisms of auditory localization cues and their interactions with visual cues in normal hearing listeners, we can better identify the changes in auditory, visual, and combined cue processing that due to HI. This identification will allow for treatments that can better target the source of the problem by extending beyond peripheral aspects of HI. To further improve understanding of these auditory-visual interactions in sound source localization for both NH and HI listeners, my dissertation has incorporated four key components with an aim to better describe how these interactions occur in real-word situations: (1) the influence of vision, (2) naturalistic testing environments, (3) saccadic eye movement as a response measure, and (4) response time.

In the real-world, objects in our environment are often both seen and heard. Vision provides a frame of reference for sound source localization (Boring, 1926; Shelton & Searle, 1980; Stein & Meredith, 1993; Warren, 1970) and can have a significant influence on auditory decision making, influencing both the perceived location of an auditory stimulus (Jack & Thurlow, 1973; Spence & Driver, 2000; Welch & Warren, 1980) and the time to execute a response (Lukas et al., 2010; Schroger & Widmann, 1998; Ragot et al., 1988; Tomko & Proctor, 2017). Appropriate integration of auditory and visual information is essential for spatial attention and may be altered following HI. A previous study in our lab showed that vision can have a greater influence on sound source localization accuracy in listeners with unilateral HI relative to normal hearing (NH) (Venskytis et al., 2019). Most of the body of multisensory research, including our previous work, typically evaluated auditory-visual interactions within the central field of vision (±20°). However, in the real-world, sound sources exist all around us. To characterize the impact of vision on sound source localization in HI listeners under more naturalistic conditions, I have measured saccadic eye movements in response to sound stimuli presented from an array of loudspeakers surrounding the listener. Eye saccades represent a natural orienting behavior and are employed frequently in real-world environments to gather information about the surrounding space (Muñoz, 2002). Additionally, eye saccades can provide a reliable response measure in populations that struggle with button pushing or more complex orienting responses. Using saccades, I have measured both listener response accuracy (RA) and response time (RT). RA, the more commonly used metric, reflects the final perceived location of a sound source. RT, however, reflects the time it takes to process sensory information and to make an appropriate response (Luce, 1991). Thus, analysis of RT may allow for the measurement of processing changes resulting from HI that may not be reflected by RA measures alone. To date, few studies have evaluated both metrics together to understand how vision affects accuracy and speed of sound source localization. Evaluation of RA and RT together in localization tasks will allow us to observe how HI affects both the final behavioral output and processing of sensory data.

In summary, work presented in this dissertation addresses four key components pertaining to real-world auditory tasks. (1) The influence of vision—providing an understanding of auditory-visual interactions that is critical for understanding localization in real world environments (Yost et al., 2015) as well as how localization deficits in listeners with HI could be improved (Experiment 1). (2) Naturalistic testing environments—providing an understanding of the spatial extent of visual influence on auditory inputs by exploring how frontal vision can interact with a wide spatial range of auditory inputs to better define the effects of HI in real world environments (Experiment 2). (3) Saccadic eye movement as a response measure—establishing the utility of a natural orienting behavior in measuring sound source localization in listeners with NH and HI. (4) Response time—demonstrating that RT can provide additional information about stimulus processing over and above RA measures, potentially providing a further understanding of individual differences in participant behavior.

Before discussing each experiment, I have included a review of the literature describing the base and rationale that has guided the development my experimental design and interpretation. The review outlines the basis of sound source localization, proposed models for multisensory integration, the role of attention on attending to multiple sensory cues and the impact of HI on attention, consideration of the speed accuracy trade off, and the neural basis of saccadic eye movements.

CHAPTER 2

LITERATURE REVIEW

Sound Source Localization

Accurate localization of sound sources is a critical task which allows us to better interact with our social environment and to avoid possible environmental dangers. The task of localizing a sound source is a complex process made possible by the combination and processing of several acoustic cues by our nervous system. The interaction of an incoming sound with a listener's external ears and head will alter the temporal and intensity characteristics of the sound to provide the important cues used to determine the sound source's location in space. These cues can be separated into two categories: (1) cues primarily providing information for lateral judgments and (2) cues primarily providing information for vertical and front/back judgements.

Lateral judgments. The nature of lateral cues has been explored for over a century with the British physicist Lord Rayleigh (1907) first describing some of the key features used in lateral judgement of sound source location. His findings, referred to as the "duplex theory", suggest that for higher frequency sounds the interaural level difference (ILD) created by acoustic shadowing or dampening by the head is the primary cue for lateral judgments. However, at lower frequencies where the wavelength of a sound is larger than the diameter of the head and therefore free from the effect of head shadow, interaural arrival time differences (ITD) are used to determine lateral positions. Many following studies have substantiated these claims through both psychophysical and physiological experiments (Macpherson & Middlebrooks, 2002; Makous & Middlebrooks, 1990; Sandel, Teas, Feddersen, & Jeffress, 1955; Stevens & Newman, 1936).

Vertical and front/back judgments. Although ILD and ITD are essential to lateral judgements of sound source position, they are incapable of providing cues for vertical and front/back judgments of sound source location as sound sources in the same sagittal plane will provide constant ILDs and ITDs. Therefore, vertical, and front/back judgements must be explained through a different cue. Batteau (1967) was one of the first to heavily explore the pinna as the generator of the spatial cues used for these judgments. His work suggested that alterations in a sound's spectrum due to reflections within the folds of the pinna could provide additional cues relevant for sound localization. Several studies that followed have confirmed that indeed these "spectral cues" are necessary for both vertical (Butler & Helwig, 1983; Roffler & Butler, 1968) and front/back localization (Musicant & Butler, 1984; Oldfield & Parker, 1984). While it has been suggested that these spectral cues are processed independently of ILD/ITD (Davis, Ramachandran, & May, 2003; Johnson & Hautus, 2010), final appropriate integration of all cues is an important step in creating a perceptual representation of our spatial environment. The study of this cue integration, along with the integration of other non-auditory sensory cues, can provide important insights to the total function of our nervous system's processing capabilities which can in turn help us understand the effects of and how to target sensory deficits.

Audiovisual Integration

While the study of auditory localization alone can provide a degree of insight, limiting our evaluation to one system cannot provide a full picture of real-world function. Objects and events in the natural world are rarely represented by information that is isolated to a single sensory system. Rather, our environment is represented by a series of multimodal features which in turn stimulate a variety of sensory pathways. To create a perceptual representation of the world, we must accurately synthesize and interpret information from these different sensory modalities. In the realm of sound localization, the influence of vision should not be ignored as it is used to provide an important spatial reference of our environment. By studying auditory localization under the influence of visual stimulation we can further probe the integrative pathways of the nervous system and gain a better understanding of the rules of integration. The logical first step in the process is to gain an understanding of the benefits and limitations that come with auditory-visual stimulation on localization tasks.

Benefits of Integration. A primary benefit of integrated auditory-visual stimulation has is a clear reduction in response latency or reaction time during localization tasks (Schroger & Widmann, 1998). In a classic paper, Miller (1982) described this reduction in response latency as the “redundant target effect,” stating that the mean response time to audio-visual information is faster than response times to either visual or auditory targets alone. The current literature has proposed two models to explain the facilitation of these faster response times: race models and coactivation models.

In race models (Raab, 1962) interactions between the neural responses of different modalities are not necessary, in fact, neural responses of the different stimuli directly compete for motor response initiation with the faster of the two modalities mediating reaction time. According to these models, the decrease in reaction time to a multisensory stimulus is simply attributed to probability summation, as the likelihood of either of the two stimuli eliciting a faster reaction time is higher than for one stimulus alone. Alternatively, in coactivation models (Miller, 1982) the interaction of neural responses across modalities is expected. Neural responses from stimulus pairs interact and are pooled prior to motor response initiation. As a result of response pooling, the threshold for response initiation is met earlier for stimulus pairs than for a single stimulus, thus eliciting a faster response. The plausibility of each group of models to explain the decrease in response times has been widely debated. More evidence supports coactivation models in explaining the decrease in response times, as the reaction times seen in multisensory studies are often faster than the race models alone can explain (Schroger & Widmann, 1998; Yang, Altieri, & Little, 2018). This would suggest that the reduction in response time is a result of the auditory and visual signals being processed together rather than separately.

In addition to facilitating faster response times, joint processing of audiovisual signals has also been shown to improve localization accuracy judgements (Bolognini, Leor, Passamonti, Stein, & Làdavas, 2007; Stein, Meredith, Huneycutt, & McDade, 1989). The addition of a coincident auditory target during a visual localization task has elicited improved precision in localization judgements in animal models (Stein et al., 1989) as well as in humans (Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003). The addition of visual cues during an auditory localization task has also been shown to provide important spatial references, resulting in improved localization performance (Stein & Meredith, 1993). In summary, the presentation of two sensory stimuli has elicited improvements in response time and response accuracy relative to the presentation of a single sensory target alone. However, there are limitations to the utility of multisensory presentations.

Limitations of Integration. While the evidence clearly outlines the benefits of audio-visual integration on localization tasks, there are constraints as to how and when the multimodal information is integrated. Temporal and spatial alignment of sensory information are thought to be the primary factors in facilitating decreased response times, as targets need to occur within a certain temporal or spatial boundary to lead to a decrease in response time.

Temporal alignment of multisensory targets is one key in the facilitation of improved response times. There are several models seeking to define the exact time window necessary for optimal integration (Colonius & Diederich, 2004; Schroger & Widmann, 1998) with debate as to the exact windowing process for temporal integration. However, a majority of models suggest that as long as the two sensory stimuli are perceived as either 1) a singular event in the temporal domain, or 2) relevant to the assigned localization task, they are able to facilitate a decrease in response time. According to this hypothesis, a possible window of integration could extend to as much as a 200-millisecond difference between the two signals (Colonius & Diederich, 2010; Frens, Van Opstal, & Van Der Willigen, 1995).

Spatial alignment is also believed to play a primary role in the degree of response time reduction based on the spatial rule of multisensory integration. This rule states that spatial congruency enhances the degree of audiovisual integration while spatial incongruence encourages independent, unisensory processing or in some cases, competition between the two stimuli (Alais & Burr, 2004; Frens et al., 1995; Ragot, Cave, & Fano, 1988; Tomko & Proctor, 2017). Simply stated, multisensory targets that are spatially aligned result in faster response times than multisensory targets that are spatially separated. This pattern has been shown most prominently in selective attention tasks in which participants were asked to attend to one select modality and ignore an accessory input. Literature has evaluated both visually influenced auditory tasks and auditory influenced visual tasks, revealing results consistent to the proposed spatial rule of integration. It is also interesting to note that vision has shown a predominance over audition, eliciting a greater delaying effect on incongruent auditory signals. Due to this visual dominance, many studies that evaluate multisensory interactions limit their study to targets that fall within a narrow region of the visual field (typically +/- 20^{o} from center).

Currently, the extent of visions influences on response times to auditory targets presented from the rear has yet to be evaluated.

Selective Attention

In many cases, our awareness of the world is based on what we choose to attend, not simply driven by the information that enters our senses. In busy, real world environments, attention plays a critical role in formulating both perception and behavioral output by focusing cognitive resources on relevant information in our surrounding space (Alain & Bernstein, 2008; Jon Driver, 2001). Selective attention, specifically, describes our ability to focus our cognitive resources on the processing of one thing rather than another. In describing models of selective attention, two significant questions are important to consider: 1) How do individuals segregate what they wish to attend from competing signals? and 2) What is the difference in processing for attended stimuli vs. unattended stimuli?

Early researchers in selective attention quickly identified that to selectively attend one target vs. another, there needs to be clear physical differences between the stimuli. Described as object based attention, it is proposed that stimuli originating from the same source may be grouped together as one perceptual object and attention is focused toward one target object while ignoring others (Driver & Baylis, 1998; Driver, 2001; Shinn-Cunningham, 2008). Researchers also identified that on selective attention tasks fairly little, if anything, is recalled about the object that is ignored when questioned after the fact (Driver, 2001; Moray, 1959).

Based on the ideas presented above, Donald Broadbent (1958), published his simple yet clear “filter theory” integrating information across multiple experiments. In general terms, Broadbent’s filter theory proposes two primary stages in perceptual processing. In the first stage, basic physical properties of all incoming stimuli are extracted in a parallel manner. In the second stage, more complex features of stimuli are processed. Broadbent proposed that this second stage of processing was more limited in its capacity than the first and could therefore only deal with a fixed number of stimuli at once. To prevent the second processing stage from overload, he proposed the idea of a selective filter. This filter would only pass through stimuli with a particular physical property, as identified by the first stage, for further processing. (Broadbent, 1954). Broadbent's initial filter theory proved to be both influential and controversial motivating two primary theories of thought surrounding the selection process of attention. The theory of *early selection,* which followed Broadbent's theory that perception is limited, requiring selective attention to proceed, and the theory of *late selection* (championed by Deutsch and Deutsch, (1963)), which assumes that perception is in fact unlimited with selection only occurring after perception to provide a relevant response.

Early selection. Early selectionist approaches propose the idea that attended and unattended stimuli are treated differently early on in processing. A common mechanism used to describe early selectionist theory is a bottleneck through which only a narrow stream of stimuli can pass with all other competing stimuli being filtered out completely prior to in depth processing. This theory would suggest that only basic physical characteristics about the unattended stimuli are noted (e.g., the pitch of an auditory stimuli). While this theory gained much traction in early studies of attention, later exploration revealed some significant gaps in the narrative (Driver, 2001; Duncan, 1980; Lavie & Tsal, 1994). One of the most significant being the finding that participants could identify key stimuli, such as their own name, or respond differently to primed stimuli in the unattended object (Moray, 1959; Neill, 1977; Tipper, 1985).

Late selection. These gaps in the story left by the early selectionist theories gave rise to a different school of thought which became the focus of many attention theories.

This approach, termed the late selection approach, assumed that perception is an unlimited process that can automatically be performed for multiple streams of stimuli without the need for selection (Deutsch et al., 1963; Duncan, 1980). According to this position, selection of stimuli comes late in the process, only to provide a relevant response. Late selectionists would propose that some of the limitations in recall of the unattended stimuli seen in previous experiments are due to limitations on memory rather than limitations in perceptual processing (Driver, 2001).

Attenuation. While both early and late selectionist theories come with strengths and weaknesses, a more nuanced approach that considers key features of both theories, combined with an understanding of bottom up and top-down processing, and cognitive load may provide a more valuable model of selective attention. Building off the initial filter model proposed by Broadbent, Treisman (1960) proposed an adapted model capable of accounting for many of the limitations noted in both early and late selection theories. Treisman proposed that unattended stimuli are simply attenuated relative to the attended stimuli rather than completely filtered out. Based on the two stages of processing proposed by Broadbent earlier, this would suggest that inputs from all stimuli reach the second stage of perceptual processing however, unattended stimuli are attenuated (rather than completely filtered out as suggested by Broadbent) relative to the attended stimuli (Treisman, 1960, 1969). This would explain early demonstrations in the literature that stimuli that have a stronger sense of importance (i.e. a participants name or a primed stimuli) can still have an effect on perception, even when unattended, by overcoming the attenuation (Driver, 2001; Moray, 1959). Treisman provided support for this attenuation theory in a classic shadowing experiment (Treisman, 1960). Participants were asked to attend and repeat a stimulus (sentence passages) played to one ear while ignoring a competing stimulus in the opposite ear. Treisman showed that if at some point the passages were switched between the two ears, participants would transfer to the opposite ear for one or two words before switching back to the target ear. This pattern of behavior would suggest that initially the stimuli in the unattended ear were too weak to break through the attenuation however, due to the increase in context after the switch in passages, the weight of the stimuli in the attended ear became sufficient enough to capture attention (Treisman, 1960). From Treisman’s work we may conclude that a greater degree of sensory information is processed than initially suggested by the filter theory, with selectivity and the importance of the stimulus itself playing a role in what is recalled. Based on early evidence supporting Treisman’s model along with how it fits into later research findings described below (Driver, 2001), this will serve as the basis of the attentional models that guide my future hypothesis. Within this model I believe it will also be key to consider aspects of perceptual load, multisensory interactions in attention and that role that both top down and bottom-up processing play in guiding attention.

Perceptual load. Later research in selective attention identified perceptual load, or the amount of information involved in processing a stimulus, as a major contributor to how sensory information is processed. Lavie and Tsal (1994) proposed that selective processing is under capacity limitations, suggesting that processing of irrelevant stimuli is only prevented when the perceptual load of relevant stimuli demands all available attentional resources (Lavie & Tsal, 1994). In an extensive literature review of studies of selective attention, Lavie and Tsal demonstrated that in tasks with low perceptual load, findings seemed to mirror a late selectionist theory, with both relevant and irrelevant stimuli being processed. In contrast, for studies with high perceptual, findings mirror an early selectionist theory, with only the relevant stimuli being processed (Lavie, 2000; Lavie, 2005; Lavie & Tsal, 1994). According to these findings, we can assume that our processing system uses the full extent of resources that it has available at the time.

Processing of what the system deems as relevant and irrelevant stimuli then becomes a supply and demand issue. The system will supply enough resources to meet the demand of high priority stimuli with any additional resources being used to processes lower priority or irrelevant stimuli. Thus, when considering a model of attention in developing a hypothesis for a task using distractors, we must consider the perceptual load of the task.

Top-down and bottom-up control. Much of the preceding discussion has focused on how and when processing capacity is allocated however, it will also be important to consider to *what* processing capacity is allocated. The most widely accepted models describe attentional control as a contrast between top down (or endogenous) and bottom up (or exogenous) control (Driver, 2001). Top-down control describes attentional control that is directed by factors that are internal to the observer such as current goals on a task. In contrast, bottom up control describes attentional control that is directed by factors that are external to the observer such as a particularly salient stimulus capturing attention (Awh, Belopolsky, & Theeuwes, 2012; Theeuwes, 1991). While many studies suggest that top down control will dominate what will be attended by the observer in many selective attention tasks, it is possible for bottom up control to capture perception when a salient stimulus is presented (Sinnett, Spence, & Soto-Faraco, 2007). Thus, in developing a hypothesis regarding visual influence on sound source localization, we must consider the interaction between top-down control on the directed task and the salience of the distractor stimuli. Additionally, given that we are examining interactions across modalities, it is also necessary to consider how selective attention is allocated across sensory modalities.

Multisensory processing. Attention has shown to have a significant effect on the interactions between stimuli of different modalities (Driver & Spence, 1998; Driver &

Spencet, 2004; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). Literature suggests that individuals are capable of deploying selective attention across sensory modalities, influencing outcomes of either multisensory integration or sensory separation based on factors of top down control (Eimer, Cockburn, Smedley, & Driver, 2001; Martin Eimer & Schröger, 1998; Spence & Driver, 1997; Talsma et al., 2010) While top down control appears to be the major factor in multisensory interactions, salient exogenous stimuli have also been shown to be capable of capturing attentional resources (Talsma et al., 2010). In a study of spread of attention across sensory modalities, Busse et al. (2005) looked to determine if attention to one sensory modality can spread to include simultaneous yet spatially disparate and task irrelevant stimuli from a different modality. Under the assumption that attending to a stimulus is known to enhance the neural response to that stimulus, their study used event related potentials (ERPs) and fMRI to evaluate the brain's response to task irrelevant auditory stimuli when participants were attending a visual stimulus. Results showed enhanced processing for the relevant visual stimulus and the spatially separated irrelevant auditory stimulus suggesting that attention to one sensory modality can spread to simultaneous signals from another modality effectively pulling that stimulus into the 'attentional spotlight' (Busse, Roberts, Crist, Weissman, & Woldorff, 2005). This would imply that while observers are capable of directing their attention to a particular sensory modality with top down control (Spence & Driver, 1997; Van Ee, Van Boxtel, Parker, & Alais, 2009), salient sensory stimuli are capable of capturing attention and biasing perception (Talsma et al., 2010; Theeuwes, 1991; Welch et al., 1980).

Hearing Impairment and Attention. One of the most common complaints of listeners with hearing impairment is difficulty communicating in noisy environments. In evaluating these concerns, it is important to not only consider the implications of hearing impairment on peripheral processing but to also evaluate how peripheral changes may impact more central cognitive processes such as selective attention. In evaluating the limitations of hearing impairment on selective attention it will be necessary to distinguish between limits in the quality of the data provided to the perceptual system (data limits) and limits in perceptual processing itself (resource limits) (Lavie & De Fockert, 2003; Norman & Bobrow, 1975). Current literature suggests data limits due to hearing impairment may be the primary reason for altered selective attention rather than changes in available processing resources (Dai, Best, & Shinn-Cunningham, 2018; Shinn-Cunningham & Best, 2008). Two major data limitations caused by hearing impairment are described below, specifically degradation in auditory object formation and cross-modal reorganization.

Auditory object formation. A key component of navigating busy acoustic environments is the ability to extract pertinent information and ignore unwanted information (Kaya & Elhilali, 2017). Much of the work on attention builds on the concept of perceptual objects, described as a perceptual estimate of the sensory inputs that are coming from a distinct physical item in the external world (Alain & Bernstein, 2008; Shinn-Cunningham, 2008). Which object receives the focus of attention will be based on interactions between the salience of the objects in the environment along with the goals of the observer (see top down vs. bottom up control discussed earlier) (Awh et al., 2012; Shinn-Cunningham, 2008; Shinn-Cunningham & Best, 2008). Therefore, when trying to selectively attend to a distinct target, the observer's ability to form distinct perceptual objects and segregate the desired object from competing objects is essential.

The formation of auditory objects is dependent on several factors ranging from simple stimulus features to observer expectations (Shinn-Cunningham, 2008; Shinn-Cunningham & Best, 2008; Talsma et al., 2010). One of the strongest influences on object formation is extraction of spectrotemporal structures of auditory inputs, specifically, sounds with common onset/offset, amplitude modulation, and frequency content tend to be perceived as originating from the same source (Bregman, 1994). Hearing impairment commonly results in a reduction in both temporal and spectral acuity relative to normal hearing listeners (Gatehouse, Naylor, & Elberling, 2003; Lorenzi, Gatehouse, & Lever, 1999). Comparison of spectrotemporal processing in listeners with normal and impairment hearing using both neural and psychoacoustic studies has shown that processing becomes significantly slower and noisier following hearing impairment (Bernstein et al., 2013; Trujillo & Razak, 2013). These reductions in spectrotemporal acuity will result in a significant deficit in the ability of listeners with hearing impairment to form distinct auditory objects. This failure in auditory object formation will, in turn lead, to a reduced ability to focus attention on a particular target while suppressing others. (Shinn-Cunningham & Best, 2008). The implications of these changes were shown in a 2018 study by Dai and colleagues who documented that listeners with hearing impairment showed significantly poorer performance on tasks of auditory attention and less robust attentional modulation of cortical responses compared to normal hearing listeners. The overall reduction in attention seen as a result in hearing impairment can be significantly impactful in busy environments when attending to a target feature while ignoring others is necessary.

Cross-modal reorganization*.* Changes in overall attention in listeners with hearing impairment may also be due to cross-modal reorganization in response to hearing loss. This line of thought is based on the general principle within neuroplasticity that neural pathways will re-organize following sensory deprivation. For example, in case of hearing impairment, it is often suggested that changes in the visual system will seek to compensate for auditory deficits. Early studies sought to provide evidence for this theory by focusing on changes in visual attention of pre-lingually deafened adults and children (Horn, Davis, Pisoni, & Miyamoto, 2005; Proksch & Bavelier, 2002). However, recent studies have also begun to evaluate changes in visual attention in adults with various degrees of post-lingual hearing loss (Campbell & Sharma, 2014; Glick & Sharma, 2017). Findings from these studies have demonstrated that even in early stages of hearing impairment there is some recruitment of auditory cortical areas for visual processing (Proksch & Bavelier, 2002). Based on these findings, one may hypothesize that re-organization of sensory processing may lead to an increase weighting of visual stimuli over auditory stimuli, even when they are task irrelevant however, this conclusion still needs to be evaluated.

Speed Accuracy Tradeoff

The speed accuracy tradeoff (SAT) has been a major area of interest in the field of cognitive science for years with many researchers seeing the tradeoff as a hallmark of the decision process itself (Heitz, 2014). Much of our understanding of the SAT is built off the sequential sampling models, which suggests that individuals make decisions based on a sequential accumulation of sensory information (for a review see Forstmann & Ratcliff, 2016). Evidence has shown that the decision process is adaptable, with actions being dictated not only by the nature of sensory input but also environmental constraints, internal goals, and biases (Heitz, 2014; Ratcliff & Smith, 2004). A reflection of this adaptability, the SAT arises due to the opposition between response speed and decision accuracy. Based on the sequential processing models, faster responses entail less accumulated evidence, and thus less informed decisions. In many cases, it is the individual that imposes the timeline for evidence accumulation via top-down control. Observers set a decision criterion (or the amount of evidence required to commit to a choice) based on current task demands and internal goals. It is key to understand what decision criteria participants employ (one favoring speed or accuracy) and how this may change cognitive processing.

In one of the earliest discussions on the SAT, Lange (1888) suggested that participants in reaction time tasks adopt either a sensory or a motor priority, depending on whether their attention is focused on the stimulus or on the response. This would suggest that they may strategically emphasize either their sensory or motor processes. A sensory emphasis would result in more accurate but slow performance while a motor emphasis would result in faster but less accurate performance (Rinkenauer, Osman, Ulrich, Müler-Gethmann, & Mattes, 2004). Over the past several years, many researchers have tried to characterize the locus of control of the SAT with findings indicating that it is not as clear cut as initially proposed by Lange. Evidence across various studies has suggested that the SAT is a multifaceted phenomenon, resulting in changes at multiple processing stages ranging from the sensory processing stage to the motor response stage.

Sensory processing. In recent literature there has been evidence indicating that SAT emphasis may affect perceptual processing (Heitz, 2014; Rae et al., 2014). Current findings support the hypothesis that significant changes can take place in sensory processing due to SAT manipulations. One current theory is based on the proposition that information extracted in early processing stages can change in quality during the course of a single decision (Rae et al., 2014; Ratcliff & McKoon, 2008). These studies have shown across several different tasks that, when decision times are shortened (i.e., emphasizing response speed rather than accuracy), the type and amount of information being accumulated changes. These ideas rely on the assumption that the quality of information extracted from the environment accrues and improves during a decision, such that faster decisions will necessarily be based on less and poorer information than slow decisions. Additional studies have also shown via single cell-recordings that SAT manipulations can alter baseline neural activity in sensory receptors, resulting in increased sensitivity to certain inputs or increased gain in sensory neurons tuned to particular stimulus features (Heitz, 2014; Rae et al., 2014; Scolari & Serences, 2009). These findings would suggest that firing rates for different populations of neurons are increased depending on whether speed or accuracy is the primary goal. Additionally, observers might base their evidence accumulation on different neural inputs under speed versus accuracy emphasis, with accuracy emphasis resulting in slower, but more detailed information accrual than speed emphasis.

Decision making. As evidence is accumulated, observers may also alter their decision-making criteria based on goals of either speed or accuracy. A large body of work has been dedicated to understanding how the SAT may affect this decision-making stage of processing with much of our understanding coming from the application of sequential processing models. As described earlier, the structure of the sequential processing models assumes that upon stimulus presentation, activity starts to increase from baseline; when a threshold is reached, a decision is made (Ratcliff & Smith, 2004). This model would then suggest that the SAT can be accounted for by a change in the distance between the baseline and decision threshold (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2009; Ratcliff & Smith, 2004). Shorter baseline–threshold distances would result in faster reaction times with potential of more errors as noise or irrelevant stimuli might lead the incorrect response to reach the threshold first. In contrast, larger baseline–threshold distance would result in slower reaction times with a smaller likelihood of error. The primary goal of research within this area is to provide evidence for these baseline-threshold shifts in conditions emphasizing either response speed or accuracy. The dominating conclusion in the literature has shown that depending on the goals of the observer, rather than altering the threshold response criteria, instead the baseline is altered as evidenced by increased initial neural activity in network areas related to decision making (Bogacz et al., 2009; Rinkenauer et al., 2004; Van Veen, Krug, & Carter, 2008). In these studies, data suggests that when speed is emphasized, an increase in baseline firing activity is seen, resulting in a reduction in the amount of neural evidence needed to reach a decision threshold. An absence of this activity is seen when accuracy is emphasized. In summary, these findings support the hypothesis that the SAT in the decision-making stage is manipulated by changes in the baseline-threshold distance. Specifically, demonstrating that the emphasis to pursue either speed or accuracy in tasks can result in changes to the starting point of decision making thus altering the amount of evidence that needs to be accrued to make a final decision.

Motor response. In addition to showing changes in perceptual processing and decision-making stages, evidence has shown changes in the motor response planning stages with SAT. Studies that have revealed changes in baseline neural activity in decision-making networks have also shown changes in cortical areas related to motor response preparation (Rinkenauer et al., 2004). For example, this increased activity is evident in neuroimaging data of frontal eye fields (FEFs), a cortical area responsible for saccade preparation and initiation. This preparatory activity in FEFs has been shown to correlates inversely with saccade latencies, indicating that greater activation of FEFs can facilitate more rapid saccade triggering (Klein & Ettinger, 2019). Additionally, event-related potential evidence has shown changes in response stage activity in response to SAT manipulations with a speed focus resulting in more baseline activity than an accuracy focus (Rinkenauer et al., 2004). Overall, changes in the response stage are like those shown in the decision-making stage, with areas responsible for motor response generation showing higher baseline activity when speed is emphasized over accuracy.

The coupling of SAT methodology and neuroscience techniques has the potential to offer insight into the cognitive processes of decision-making. The consensus emerging suggests that SAT is a multifaceted phenomenon, influencing several components of the decision process and is accompanied by distinct changes in brain activity. In interpreting studies that evaluate both response speed and response accuracy it will be important to consider changes in the quality and quantity of sensory information processing, the goal of the participant within a particular study and the influence that this top-down processing may have on baseline activity of neurons involved in both decision making and response execution. Accounting for these changes may help explain individual differences in behavior.

In conclusion, reaction time has been shown to reflect the result of multiple stages of cognitive processing including perception and refinement of sensory stimuli, decision-making, and motor response planning and activation. Each of these stages may be influenced by the nature of stimuli in our environment, internal goals, and the activation of additional cognitive processes such as attention. Changes within these stages will lead to alterations in how information is processed, and the final reaction time measured. Experimenters can use this understanding of typical processing to better identify changes in cognitive processing stages due to impairment including limitations in processing of perception information, changes in the decision-making process including time and energy costs, and alterations in the motor response circuitry. This line of work can be particularly relevant in the evaluation of those with hearing impairment as these individuals may show normal performance on many auditory tasks however, the additional strain in processing to reach this level of performance may be overlooked.

Eye Saccades

Saccadic eye movements are characterized by a rapid shift of gaze from a fixed location to a location of interest. The primary purpose of these movements is to bring images from the environment to focus on the fovea, the region of greatest visual acuity. This allows the visual system to extract detailed information about objects in the environment, build a mental representation of space, and guide behavior (Foulsham, 2015). Generation of saccades typically requires little thought and occurs often in daily living (2–3 times per second), however, proper execution of saccades requires a complex series of neural processes (Klein & Ettinger, 2019). Extensive research of the ocular motor system in addition to the cognitive mechanisms behind saccade generation has revealed a distinction between two classes of saccadic eye movements: visually guided or reflexive saccades and volitional or endogenous saccades. Reflexive saccades are stimulus-driven and occur rapidly and frequently in everyday situations. Volitional saccades are internally driven and rely more upon cognitive processes such as attention, inhibition, and working memory (Klein & Ettinger, 2019). The generation of simpler visually guided saccades and more complex volitional saccades require similar basic neural circuitry, with additional neural regions supporting higher level processes when required (Mcdowell, Dyckman, Austin, & Clementz, 2008). The neural circuitry involved in both types of saccade generation has been well characterized in humans. The primary regions identified to control saccades include visual cortex, posterior parietal cortex, frontal and supplementary eye fields, prefrontal cortex, anterior cingulate cortex, cerebellum, thalamus, basal ganglia, and superior colliculus. These areas are involved in multiple stages of saccade planning and generation with different regions activated to a greater or lesser strength or extent depending on the type of saccade task and the extent cognitive demands required. The summary of the neural circuitry of saccade generation in humans will begin with a focus on the neural circuitries identified to support reflexive saccade responses and then progress to describe the additional brain regions involved in supporting more complex, volitional saccades.

Reflexive saccades. A reflexive saccade is a response that involves the rapid redirection of gaze to a stimulus and is typically generated to align the fovea with visual targets of interest (Mcdowell et al., 2008). In the literature, reflexive saccades are evaluated primarily in response to visual stimulation, however additional sensory stimuli, such as auditory targets, can be used. Functional neuroimaging studies in humans have contributed significantly to characterization of regions involved in reflexive saccade generation. The general pathway discussed will include subcortical (the striatum, thalamus, and superior colliculus) and cortical (visual and extrastriate cortex, regions of posterior parietal cortex, and frontal and supplementary eye fields in frontal cortex) structures.

For saccadic responses to visual stimulation, input from the target stimulus travels from the retina through the thalamus to primary visual cortex in the occipital lobe contralateral to the location of the stimulus in the visual field (Klein & Ettinger, 2019). From primary visual cortex, information is passed to extrastriate cortical regions to extract general stimulus features. From here, information is then passed along the ventral “what” stream in the temporal lobe for stimulus identification and to the dorsal “where” stream in parietal cortex to determine location of the stimulus in space (Klein & Ettinger, 2019; Mcdowell et al., 2008). Additionally, visual and extrastriate cortex have been shown to have access to saccade generators in the brainstem through the superior colliculus (Mcdowell et al., 2008). Subsequent studies have shown that the activity of neurons in these cortical areas can predict subsequent saccadic behavior (Mcdowell et al., 2008), especially for reflexive saccades as the occipital lobe has been identified as one of the only regions to show increased activity for reflexive relative to volitional saccades (Dyckman, Camchong, Clementz, & McDowell, 2007).

As visual information progress along the “where” stream it reaches the parietal cortex. Several regions within the parietal cortex have been shown to have direct connections to the superior colliculus and the frontal and supplementary eye fields (Mcdowell et al., 2008; Munoz & Everling, 2004). These projections would suggest a role in saccade triggering (Munoz, 2002; Munoz & Everling, 2004). Additionally, parietal cortex is known to support deployment of visual attention (Klein & Ettinger, 2019) including spatial awareness and sensorimotor transformation (Colby & Goldberg, 1999). Accordingly, most imaging studies have suggested that parietal cortex plays a greater role in volitional saccade generation (discussed more later) however, some activation is present in parietal cortex during reflexive saccade tasks (Mcdowell et al., 2008).

Following processing in the parietal cortex, the saccade generation pathway will progress to the motor regions within frontal cortex, specifically the frontal and supplementary eye fields (FEF/SEF). These regions are critical for preparation and initiation of saccades (Klein & Ettinger, 2019). and have been shown to be one of the most active regions in the saccade circuitry (Anderson et al., 1994; Connolly, Goodale, Goltz, & Munoz, 2005). Significantly, evaluation has shown that connections exist between the FEFs, the superior colliculus and saccade generators within the brainstem. These projections, along with studies showing an inverse correlation between saccade latency and FEF activation would suggest that build-up of activity within the FEF region contributes directly to saccadic initiation (Hanes & Wurtz, 2001). The SEFs have also been shown to be involved in saccadic initiation however, activation in this area is more typically observed for volitional compared to reflexive saccades (Mcdowell et al., 2008) and will therefore be discussed more later.

In addition to the cortical areas involved in reflexive saccade pathway, several subcortical areas have been shown to play a significant role in saccade generation. While the depth of these structures results in certain measurement challenges, the role of several areas has been established including that of the cerebellum and superior colliculus (Mcdowell et al., 2008). The cerebellum, primarily the oculomotor vermis, is critical for saccade adaptation, motor planning and timing, and amplitude control (Klein & Ettinger, 2019; Mcdowell et al., 2008). Additionally, the superior colliculus (SC) has been shown to play a significant role in reflexive saccade generation, controlling both visual target selection and saccade initiation (Munoz, 2002). At its most superficial layers, the SC has been shown to receive direct input from multiple cortical areas as well as from the retina (Klein & Ettinger, 2019; Munoz, 2002). Intermediate layers of the SC contain a retinotopically organized motor map corresponding to saccade vectors within the visual field (Munoz, 2002). Evidence also suggests that this retinotopic motor map may be the primary site of multisensory integration or competition, as multiple locations may be activated by various stimuli in the environment (Munoz, 2002; Stein & Meredith, 1993). For example, processing of multiple visual or accessory auditory stimuli at the level of the SC may prove to impact final saccade generation resulting in changes in saccade timing and/or spatial accuracy. However, in the end, processing within the SC must result in final selection of a single target and activation of the oculomotor circuit within the brainstem (Munoz, 2002).

The pathway described here, starting in the visual cortex, and progressing through several cortical structures to the superior colliculus is generally accepted as the neural circuitry for the generation of reflexive saccades in response to an external stimulus. Additionally, in some cases of express reflexive saccades, saccade generation may occur via a highly direct pathway from visual input to motor output in superior colliculus that, bypassing many of the cortical structures described earlier (Klein & Ettinger, 2019). However, this pattern can be rare in many real-world situations as our environment may contain multiple competing stimuli which may prevent express reflexive responses toward each stimulus. Selection of relevant stimuli therefore must occur to determine which stimuli are worth the direction of gaze, at the expense of other potential targets. This selection process will incorporate additional cortical structures in addition to the basic reflexive pathway.

Volitional saccades. To selectively attend to a particular aspect of the environment, responses to other parts of the environment often must be suppressed. Saccadic movements requiring inhibition, working memory and higher order processes are considered volitional saccades (Klein & Ettinger, 2019). Our understanding of the generation of volitional saccades comes primarily though studies which require the inhibition of responses (i.e., anti-saccade or delayed response tasks. Neuroimaging data during these tasks has identified increased activity in the parietal cortex and frontal and supplementary eye fields relative to reflexive saccade generation, in addition to activity in prefrontal cortex, anterior cingulate cortex, thalamus and basal ganglia.

In addition to the function presented earlier, parietal cortex has also been suggested to play a significant role in initial suppression of saccades toward external stimuli along with working memory maintenance and attention associated functions (Klein & Ettinger, 2019; Mcdowell et al., 2008). Increased activity has been shown in select regions of parietal cortex during anti-saccade tasks, specifically during the inhibitory period prior to anti-saccade generation (Ettinger et al., 2008). This may suggest that these regions of parietal cortex play a significant role in inhibition of responses during volitional saccades (Mcdowell et al., 2008).

In conjunction with increased activity in parietal cortex, volitional saccade tasks have also revealed increased activity in FEFs and SEFs. In FEFs, neuroimaging studies have shown heightened neural activity prior to anti-saccade generation, suggestive of an inhibitory effect, in addition to heightened activity through the selection stage of a saccade response, suggesting involvement in the final selection of a saccade target (Mcdowell et al., 2008). SEFs have also shown significant activation prior to saccade generation during volitional saccade tasks with researchers suggesting that the SEFs are heavily involved in the suppression of a movement toward a peripheral stimulus (Dyckman et al., 2007). Overall, these significant increases activation in would suggest the involvement of FEFs and SEFs regions in inhibition unwanted responses and identifying target response locations during volitional saccade tasks.

In addition to enhancements in levels of activity of reflexive saccade circuitry, there is evidence that additional regions of the brain are recruited during more cognitively complex saccades. One of these areas is the prefrontal cortex (PFC) which has shown increased activation in response to the greater cognitive demands associated with the generation of volitional saccades. The PFC has been implicated in maintaining appropriate stimulus-response mappings, selecting the proper response based on the specific task context, and biasing sensory and motor regions to generate the desired action(Klein & Ettinger, 2019). Specifically, dorsolateral PFC (DLPFC) plays a role in the suppression mechanisms necessary for anti-saccades and other tasks that require withholding a saccade at the appearance of a new stimulus (DeSouza, Menon, & Everling, 2003). This top-down suppression can result in control changes to early visual processing of peripheral stimuli, dampening the visual response or delaying the activation of a reflexive saccade (Klein & Ettinger, 2019; Mcdowell et al., 2008).

Another region commonly recruited to aid in the execution of volitional saccades is the anterior cingulate cortex (ACC). The ACC has shown increased activation in activities requiring conflict and error monitoring as well as preparatory performance optimization (Brown, Goltz, Vilis, Ford, & Everling, 2006). Activation patterns of the ACC during saccade generation tasks suggest that it is primary responsible for suppressing of and monitoring for error (Mcdowell et al., 2008).

Finally, two subcortical structures, the thalamus, and basal ganglia, have shown increased activation for volitional saccades relative to reflexive saccades. This activation is likely due to the in relation of these areas to attention and motor learning (Aglioti, 1997). The thalamus, particularly the pulvinar (Klein & Ettinger, 2019), has been shown to play a role in visual attention shifts. The basal ganglia receive motor signals from frontal cortex and provide inhibitory input to superior colliculus and thalamus (Munoz & Everling, 2004) and have also been associated motor learning and reward processing (Hikosaka, Nakamura, & Nakahara, 2006). Overall, the changes in the generation of volitional saccades involve heightened activation in or recruitment of additional areas responsible for mediating attention, inhibition, and working memory to allow for finer control over direction of gaze toward a desired target.

In short, the neural control of saccadic eye movements is made up of a network of brain areas beginning in visual cortex, encompassing both parietal and frontal cortices, and extending through cerebellum, thalamus, basal ganglia, and superior colliculus. Through coordinated activity within this network, saccades are generated efficiently as the brain responds to changing cognitive demands. While both reflexive and volitional saccades may use the same general pathway, they are primarily differentiated by the level of activation of structures within the pathway based on type of saccade generated and the cognitive demand of the task at hand.

Present Study

While an important framework for the mechanisms of audiovisual integration in sound source localization has been laid, there are still significant questions in terms of the capabilities and limitations in the integration of these cues both NH and HI listeners. One important area of focus is the pattern and extent of interaction, if any, when cues do not align, as often happens in real-world environments, and how misaligned cues may influence both the speed and accuracy of decision making. The current study will outline the extent of vision's influence on auditory cues of different degrees of spatial salience in both NH and HI listeners (Experiment 1). It will also describe vision's influence on auditory cues presented from in front of and behind a listener (Experiment 2). Results show that vision has a significant influence on the localization of auditory targets, influencing both the accuracy and the variability of responses, especially for less salient auditory stimuli, even when the auditory and visual stimuli are separated in space.

CHAPTER 3

EXPERIMENT 1: EFFECTS OF STIMULUS SPATIAL SALIENCE ON RESPONSE ACCURACY AND RESPONSE TIME IN MULTISENSORY LOCALIZATION TASKS

Introduction

Perceptual experience in everyday environment is often built upon interactions between multiple different senses to guide action and decision-making. An excellent example of this multisensory strategy is sound source localization. Knowing where sounds come from helps guide attention towards a target location and improve listeners' awareness of their physical surroundings. Humans use primarily interaural differences of time and intensity (ITDs and ILDs) (Makous & Middlebrooks, 1990; Middlebrooks & Green, 1991), spectral cues (Butler & Helwig, 1983; Roffler & Butler, 1968) and their combinations to realize the directional information of a sound source. Hearing, however, is not the only sense that contributes to target localization and spatial awareness. Objects in the everyday environment are often both seen and heard. Vision provides a frame of reference for sound source localization (Warren, 1970; Stein & Meredith, 1993) and can influence the perceived location of an auditory stimulus and the time to execute a response (Lukas et al., 2010; Schroger & Widmann, 1998; Ragot et al., 1988; Tomko & Proctor, 2017). Many studies have shown that auditory localization can be biased by a simultaneously presented, spatially disparate visual stimulus (Jackson, 1953; Pick et al., 1969; Thurlow and Jack, 1973; Choe et al., 1975; Slutsky & Recanzone, 2001). A well-known example of visual bias is the `ventriloquism effect', in which a visual stimulus causes a change in the perceived location of a sound source (Bertelson & Radeau 1981; Hairston et al 2003a; Wallace et al 2004). This effect of visual capture reflects the dominant effect of vision over audition and is generally the case (Welch and Warren 1980). Additionally, studies have shown that a simultaneously presented, spatially incongruent visual stimulus may also delay the reaction time to an auditory stimulus (Lukas et al., 2010; Ragot et al., 1988; Tomko & Proctor, 2017). Researchers suggested that this visual dominance in processing time is attention based, hypothesizing that because visual stimuli are not as automatically attention-capturing such as auditory stimuli (Posner et al., 1976), individuals must actively focus their attention toward a visual stimuli. This active attention focusing requires additional cognitive resources, thus fewer resources were left to attend to (and to process) stimuli in other modalities (Lukas et al., 2008; Posner et al., 1976).

In addition to visual stimuli influencing sound source localization, the visual environment itself has also been shown to affect auditory localization (Mastroianni, 1982; Majdak et al., 2009; Warren, 1970; Wightman & Kistler, 1993). Several studies have shown that visual information can enhance auditory function in humans. For example, for sound source localization tasks, Warren (1970) and Shelton and Searle (1980) showed that individuals were better at localizing an auditory target with the eyes open (or in a lighted environment) than with them closed (or in a dark environment). Furthermore, it has been demonstrated that a textured visual environment can further improve the accuracy of sound source localization judgments relative to a blank visual field (Majdak et al., 2009). This visual facilitation in sound source localization has been suggested to be a result of the improved frame of reference that vision provides to the task, allowing for more precise execution of a motor response (Mastroianni, 1982; Warren, 1970).

In previous studies of sound source localization and how it is influence by vision, several behavioral response methods have been used including: verbal responses (Mason, Ford, Rumsey, & De Bruyn, 2001; Wightman & Kistler, 1989); rotating a dial or drawing (Haber, Haber, Penningroth, Novak, & Radgowski, 1993); pointing with the head/nose (Bronkhorst, 1995; Middlebrooks, 1999; Pinek & Brouchon, 1992); pointing with a finger (Haber et al., 1993); pointing with a laser pointer (Lewald & Ehrenstein, 1998; Seeber, 2002); using a computer interface with a preset layout (Begault, Wenzel, & Anderson, 2001; Montagne & Zhou 2016); and pointing with the eyes (Frens & Van Opstal; 1994). Many of these methods may include time consuming instrumentation and analysis procedures or necessitate unnatural responses of the participants which require extensive training. Additionally, as the different types of motor responses engage different sensorimotor systems and different spatial reference frames, they may be susceptible to different extents of multisensory influence. This may lead to difficulty in the interpretation of results across studies due to differences in motor variability or other confounds introduced by the task.

Previous studies have demonstrated that as spatial uncertainty in one sensory domain increases, a relatively salient sensory stimulus in another sensory domain would assume dominance in multisensory localization (Battaglia et al., 2003; Alais & Burr, 2004; Ernst & Bülthoff, 2004; Heron, Whitaker & McGraw, 2004). In a series of studies, we have shown that the strength of vision's influence on sound source localization is mediated by the perceived reliability of the auditory signals. More specifically, visual bias observed in auditory-visual localization is correlated with response uncertainty revealed in auditory-alone localization for both normal-hearing and hearing-impaired individuals (Montagne & Zhou, 2016; Venskytis et al. 2019). In these experiments, we used timing- and level-based stereophony to manipulate the left and right direction of an apparent sound source. We showed that these stereophonic manipulations introduce binaural ambiguity between ITDs and ILDs, and visual bias are greater in response to timing-based manipulations than to level-based manipulations (Zhou et al. 2018).

These previous results were obtained by having participants reported their perceived location of a sound source using a touch screen user interface with a pre-set layout (Montagne & Zhou 2016, 2018; Zhou et al., 2018). While participants were reliably able to indicate their perceived sound source location and influences of vision on audition consistent with the literature were measured, we have also seen significant within- and between-subject variability in our results. We have hypothesized that this variability may be due to the sensorimotor transformation that is required for our participants to generate a localization response that matches onto the preset layout. To evaluate this hypothesis and further detail audio-visual integration in free-field sound source localization, we measured the influence of visual stimuli and the visual environment on participant eye saccades in response to auditory signals of varying degrees of binaural ambiguity and compared these results to data from our button pushing experiments. Eye saccades are natural orienting response to sound and have been shown to enhance auditory spatial resolution (Braga et al., 2016; Collins et al., 2010; Vliegan et al., 2004). We observed that results obtained using eye saccades reveal similar patterns of audio-visual interaction as those seen in our previous studies and that saccade responses show decreased within- and between-subject variability than button pushing responses.

While most of the literature has been dedicated to the study of multisensory interactions on the accuracy of participant’s response, it will also be important to consider how multisensory interactions may influence the timing of participant responses. Response accuracy, which is the more commonly used metric, reflects the results of sensory processing. Response time, however, reflects the time it takes to process sensory information and to make an appropriate response (Luce, 1991). Thus, analysis of response times would allow for the measurement of changes in processing that may not be reflected in measures of accuracy alone. To date, few studies have evaluated both metrics together to understand how vision affects accuracy and speed of sound source localization. Evaluation of the effects of vision on both response accuracy and time in auditory localization tasks would allow for observation of these effects on both final behavioral output and processing of sensory data. This analysis would be essentially meaningful when considering listeners with sensory impairment such as hearing loss.

Clinically, it is not uncommon to observe variability in treatment outcomes when addressing hearing impairment. Additionally, while treatments such as hearing aids may provide benefit in quiet, single talker environments, individuals with hearing loss often report difficulty in dynamic, real-world environments. Focusing on the auditory system alone often fails to fully address these difficulties. Improvement of treatment outcomes will require understanding auditory perception in real-world environments, and that in turn requires an understanding of how cognitive processes and other sensory systems interact with audition. Considering sensory interactions on both the accuracy and timing of these sensory interactions could provide additional insight into the variable outcomes seen between individuals and across different sensory environments. The following three studies were designed to provide insight regarding the influence of vision on auditory localization in listeners with normal hearing and hearing impairment through the measure of both response accuracy and response time.

EXPERIMENT 1A: EFFECTS OF STIMULUS SPATIAL SALIENCE ON RESPONSE ACCURACY IN MULTISENSORY LOCALIZATION TASKS

Methods

Participants

A total of twenty-two normal hearing participants (ages 20-34, 2 males) completed this study. Hearing sensitivity was verified using standard audiometric techniques with insert earphones. All participants were right-handed and reported normal or corrected-to-normal vision. Participants provided written informed consent and received financial compensation for their participation. The experiments were conducted in accordance with procedures approved by the Arizona State University's *Institutional Review Board.*

Apparatus and Stimuli

The free-field experiment was conducted in a dimly lit (30 lumen/m2), double-walled, sound-attenuated chamber (Acoustic Systems RE-243, 2.1 m x 2.1 m x 1.9 m). The participants were seated in the center of the chamber with their head fixed using a high precision head positioner (HeadLockTM, Arrington Research), which was securely attached to a small table in front of the participant. Participants confirmed the start and the end of a trial using a modified gaming console (Xbox 360 Arcade Joycon) during the experiment. Stereophonic presentations were used to render the spatial perception of a sound in the horizontal plane as described in our previous studies (Montagne & Zhou. 2016, 2018; Zhou et al., 2018) and in the later section. Briefly, for delivering stereo stimuli, two loudspeakers (full-range monitor, Adam F5), hidden behind a black, acoustically transparent curtain were positioned at +/- 35 ° from center at a distance of 1.1 m from the participant's head. Auditory stimuli were broadband noise bursts gated on and off with a 15-ms rectangular window so that the frequency range of the noise burst was bounded by the frequency response range of the loudspeakers (50 to 50 kHz). Identical noise signals were presented from the two loudspeakers and the relative delay and amplitude ratio were manipulated to create a phantom source location based on stereophony techniques (Fig. 1A). The average intensity for all auditory stimuli was maintained at 65 dB SPL (dBA) as verified using a sound level meter (Brüel & Kjær 2250-L) positioned at the location of the listener's head. The stimuli were generated using custom-designed software written in MATLAB using stereophonic techniques described below.

For delivering visual stimuli, a light emitting diode (LED) bar spanning +/- 60° in the azimuth plane was suspended from the ceiling, directly in front of the acoustically transparent curtain. Only a subset of LEDs (+/- 20°) was used for this experiment. The LED bar was positioned to be at participant eye level when they were seated with their head fixed by the head positioner at a distance of 1 m from the participant's head. Visual stimuli were 50-ms duration blue light flashes generated by a single LED (8 candela) at three different locations (0° (center) and +/- 8° (left and right). The color, intensity, and position of LEDs were controlled by an Arduino Uno board. The sound and light stimulation were controlled by MATLAB commands. Synchronicity of A and V stimuli was verified using a digital oscilloscope (RIGOL, DS1052E), revealing a small time gap (7.88 ms, +/- 2.96 ms) between the onsets of A and V stimuli.

Two visual environments ("no reference" and "with reference") were tested in the experiments. For the "no reference" condition, no visual landmarks were present in front of the participants during saccades. For the "with reference" condition, seven 1"x1" white stickers were placed above the LED stripe bar, spanning +/- 15° at 5° apart (Fig. 1B). Among twenty participants, twelve participated in the "no reference" environment and ten in the "with reference" environment, six completed the study in both environments.

Time-based and level-based stereophony

Time-based and level-based stereophony were used to create phantom sound source locations for this experiment (Fig. 1A). In stereophony, to pan a perceived sound location to a left/right direction, the left-speaker signal is set to lead/lag the right-speaker signal (timing-based) or to amplify/attenuate its intensity relative to that of the right-speaker signal (level-based). When the two loudspeakers emit identical sounds with small timing or level delay, a phantom sound source is perceived at a location in between them (Leakey, 1959). This procedure has been successfully implemented in our laboratory to investigate audiovisual interactions in listeners with normal hearing and hearing impairment (Zhou et al., 2018; Venskytis et al. 2019). To vary the horizontal position of a stereo sound from left to right, we employed four inter-channel delays (ICDs) at -0.66, -0.33, +0.33 and +0.66 ms and four inter-channel levels (ICLs) at -10, -5, +5, +10 dB . Negative ICD/ICL values lead to a perceived sound from the left direction.

Eye Tracking and Calibration

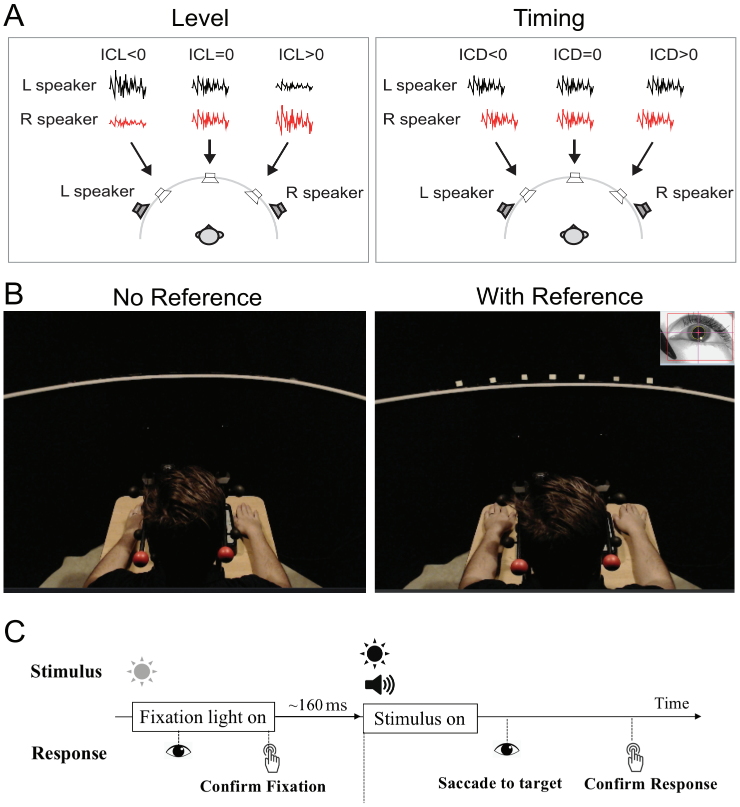

Figure 1. Apparatus. (A) Stereophonic setup. A single perceived sound source was created by varying the ICL (Level) or the ICD (Timing) difference between the two loudspeaker signals, as shown in (A) and (B), respectively. The stereo sound was perceived toward the loudspeaker emitting either earlier or louder signals (e.g., negative ICD and ICL for leftward responses). (B) View of the visual field showing the two visual environment texture effects used (No Reference and With Reference) with an inlaid example of pupil capture. (C) Stimulus-response sequences for each trial in the auditory localization task.

Participant eye movements were recorded monocularly using an infra-red eye tracking camera (Arrington Research, 90-Hz Vi*ewpoint EyeTracker ®).* The camera was mounted on an arm fixed to the left post of the head stabilizer and directed at the left eye. The camera was positioned approximately 15 cm from the eye at 0° azimuth and approximately 40° below participant line of sight. An infrared light source illuminated the left eye to increase the contrast between the iris and the pupil to allow for the Dark Pupil method of eye tracking. The eye tracking system was calibrated prior to each experimental session and following each break in testing. The eye image was adjusted so that the pupil was centered in the camera frame (See the inset image of a captured pupil in Fig. 1B). The brightness of the eye image was then manually adjusted to maximize the contrast between the dark pupil and the light corneal glint. Pupil threshold was adjusted as the participant gazed in five different directions (center, far right, far left, up, and down) to ensure there was no slippage in pupil tracking for the most extreme eye movements. Once an adequate eye image and pupil lock was obtained, the participant's relative eye position was mapped to their gaze point in space across +/- 20° in the azimuth plane at 2 ° resolution. The calibration procedure was completed using a set of 21 LEDs (10 candela) distributed evenly across the LED bar. Only one LED was illuminated at a time in sequential order, moving from right to left. Saccadic eye movements have a typical duration of 30 – 100 ms and a latency of 100 – 300 ms (Ramanauskas, 2006). Therefore, LEDs were illuminated for a duration of three seconds, to obtain sufficient calibration data without causing participant fatigue. Visual cue position and gaze position were compared to ensure a calibration error, calculated as the mean difference between gaze position and target position, was below 0.5 degrees. If the calibration error was above the cut-off, the participant was re-positioned to repeat calibration. Fig. 2A shows the calibration check for two participants, L56 and

L36. The calibration error for these participants was 0.2° and 0.3° respectively. Once the calibration was complete, participants were directed to look at randomly illuminated LEDs within the calibration field to ensure that the calibrated eye position matched the actual gaze position. Each participant's eye and gaze locations were viewed by experimenters on a computer monitor during data collection to ensure that adequate eye tracking was maintained throughout the duration of the experiment.

Procedures

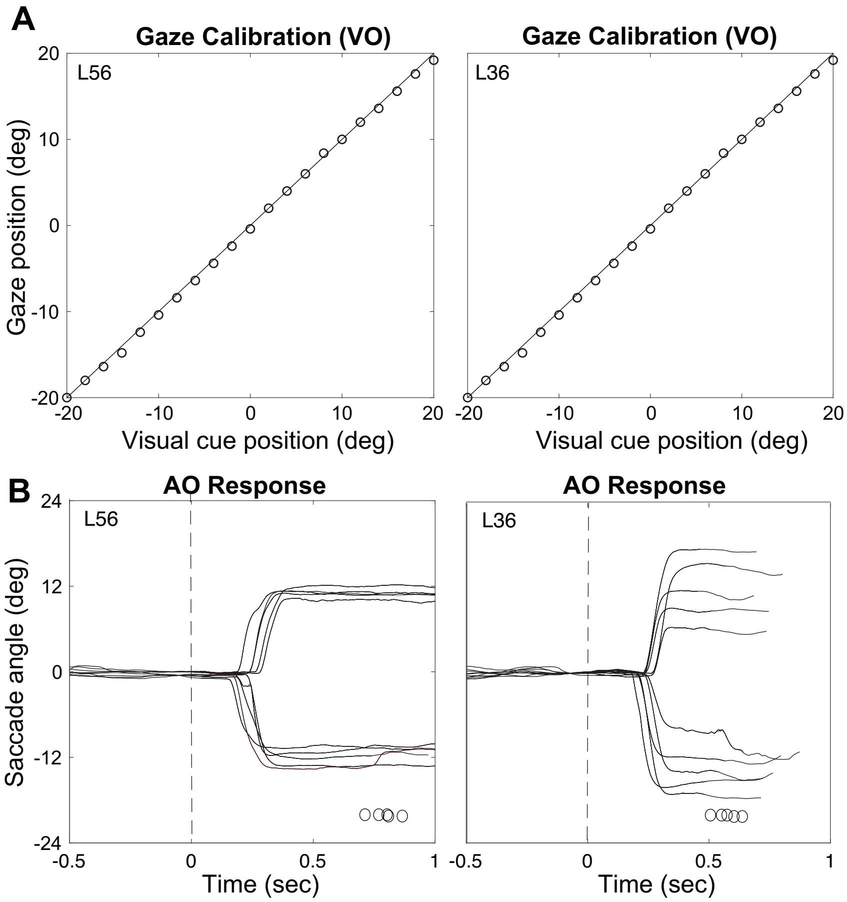

Figure 2. Gaze calibration and saccade tracings. (A) Gaze calibration results for two example participants (L56 and L36). (B) Saccade response to left and right AO stimuli for the same two participants with one (L56) showing little response variability and the other showing larger response variability (L36).

Two experimental conditions (auditory localization and visual localization) were conducted over two visits on two separate days for a total of 2 hours of testing. Participants who completed the study in both visual environments were tested over four visits on four separate days. In the auditory localization condition, participants were instructed to ignore any visual cue and to make a gaze shift as rapidly and accurately as possible to the perceived location of the auditory target. Sound stimuli with or without lights were presented in randomized blocks, denoted as audio-visual (AV) and auditory-only (AO) blocks, respectively. The AO block contained eight stimuli total (four time-based and four level-based manipulations). In the AV block, each of the eight auditory stimuli was paired with one of the three LEDs, resulting in a total of 24 stimuli. The order of blocks and stimuli was randomized. This contrasts with our previous studies where timing-based and level-based stimuli were presented in separate blocks (Montagne & Zhou, 2016; Zhou et al., 2018). In our early studies, we observed that the level of auditory ambiguity affects the overall strength of visual bias on auditory performance within a block. For example, a greater degree of visual bias was found for the single-speaker control stimuli when they were presented within timing-based stimuli block than the level-based signals stimuli block. To minimize this contextual effect, we tested the timing-based and level-based signals in the same AO or AV block in the present study.

In the visual localization condition, participants were instructed to make a gaze shift as rapidly and accurately as possible to the visual target. Visual stimuli only (VO) were presented without sounds. The VO block contained three stimuli total (0^{o} (center) and +/- 8^{o}). The presentation order of the visual stimuli was randomized. Each participant completed an average of six of each auditory block and two of each visual block resulting in a total of 30 presentations of each of the 32 auditory stimuli and a total of 20 presentations for each of the 3 visual stimuli. Altogether, participants completed 1020 trials (auditory: 32*30 =960 + visual: 3*20+60) over a total of 2 hours of testing.

Figure 1C shows the stimulus/response sequence within an AO, AV, or VO trial. At the start of each trial, a red fixation LED was illuminated at 0^o. Following fixation, the participant pressed a button on the gaming console to confirm fixation and ready for the target stimulus. While the participant maintained center fixation, the fixation light turned off. After 160 ms (+/- 50 ms), the sound stimulus (AO block), sound and light stimuli (AV block), or light stimulus (VO block) was presented. The participant indicated the perceived location of the target stimulus by first directing his/her gaze as rapidly and as accurately as possible toward the perceived direction of the target source on the azimuth plane. After the gaze shift, the participant pressed a button to confirm the response and end the trial. Participants were not provided any feedback or knowledge of their results during or after the experiment. They were also unaware of the total number of loudspeakers and spatial location of each speaker. During testing, participants were encouraged to take a brief break outside the sound booth every 15 minutes. Following each break, participants were reminded of task instructions and consistent head position and eye calibration was verified.

Before the experiment, to help participants understand the nature of the experiment and the response tasks expected a training block for auditory localization was presented. Participants were presented with stimuli from the AO block of the experiment and were instructed to respond by 1) directing their gaze to the lateral direction of a sound and 2) pressing a button to finalize the response. The training block continued until participants demonstrated the ability to properly execute the sequential response (~10 minutes).

Data analysis

Figure 2B shows typical saccade responses for two listeners (L56 and L36) in response to left and right AO stimuli. Central fixation can be seen prior to the initiation of each trial. At the onset of the sound (time=0 sec), rapid saccadic eye movements can be seen in the direction of the auditory target. While maintaining target fixation, the participant confirmed the response with a button push (black circles). The target fixation angle was calculated as the averaged gaze position after the peak saccade amplitude (i.e., saccade endpoint) over a period of 100 msec. The averaged gaze position was used as the perceived lateral position of the sound source for each trial. The average gaze position and its variability was calculated as the mean and standard deviation of responses across all trials, respectively. Participants showed various degrees of variability in their saccade behavior in auditory localization. For example, while showing similar performances in visualization (Fig. 2A), L56 showed little variability in AO response, whereas L36 showed a high degree of response variability from trial to trial. The mean and standard deviation (SD) of each listener's eye fixation position in response to each stimulus were extracted. These responses were used to analyze the effects of stereo signals on response variability. Eye fixation position was also used to determine visual bias, which was calculated as the unsigned magnitude of bias ($|\Delta AV| = |AV - AO|$) between AV and AO responses on trial-by-trial basis for the same auditory stimulus. Data were further grouped and averaged across conditions for which light and sound were presented from the same side of the midline (same) and when light and sound were from opposing sides of the midline (opposite). For group analysis, the mean and standard deviation (SD) of each listener's responses for each stimulus were pooled.

EXPERIMENT 1A: EFFECTS OF STIMULUS SPATIAL SALIENCE ON RESPONSE ACCURACY IN MULTISENSORY LOCALIZATION TASKS

Results

The results of eye fixation responses to an auditory or visual target were analyzed to address three questions: (1) Can a listener successfully indicate the perceived location of a stereo sound source using fixation? (2) Do less salient auditory stimuli increase visual bias in eye localization similar to those found with our previous button-pushing experiments? (3) Do the patterns of eye localization change between the two visual environments (“no reference” and “with reference”)?

Eye fixation response reveals stereo perception

A previous study from lab our showed that signals generated using timing-based and level-based stereophony have different degrees of saliency or ambiguity, with timing-based manipulations resulting in increased ambiguity and a wider auditory image than that created with level-based manipulations (Montagne & Zhou, 2016). These observations were based on a button pushing experiment in which the listener was instructed to indicate the perceived location of a stereo sound by between 7 potential azimuth locations by pressing a button on a touch screen GUI with a preset layout.

Figure 3 shows the target fixation positions (mean +/- SD) for all participants in response to level-based and timing-based AO stimuli in the “no reference” (Fig. 3A) and “with reference” (Fig. 3B) visual environments. For all listeners, changing level-based (ICL) and timing-based (ICD) stereo cues elicited responses in the left and right directions, as shown by the mean fixation positions associated with the four target stimuli. In our previous studies, it was revealed that timing-based stereo localization is primarily based on ITDs, whereas level-based stereo localization is based on combined ITDs and ILDs (Montagne and Zhou, 2016). Figure 3 shows that eye fixation can be used in stereo localization and its magnitude is scaled with the size of binaural signals. This is an encouraging result because localization responses obtained using eye tracking can reveal spatial perception with much refined precision (at least ~2 degrees, Fig. 2A) and less motor errors than handing/head pointing. More importantly, eye

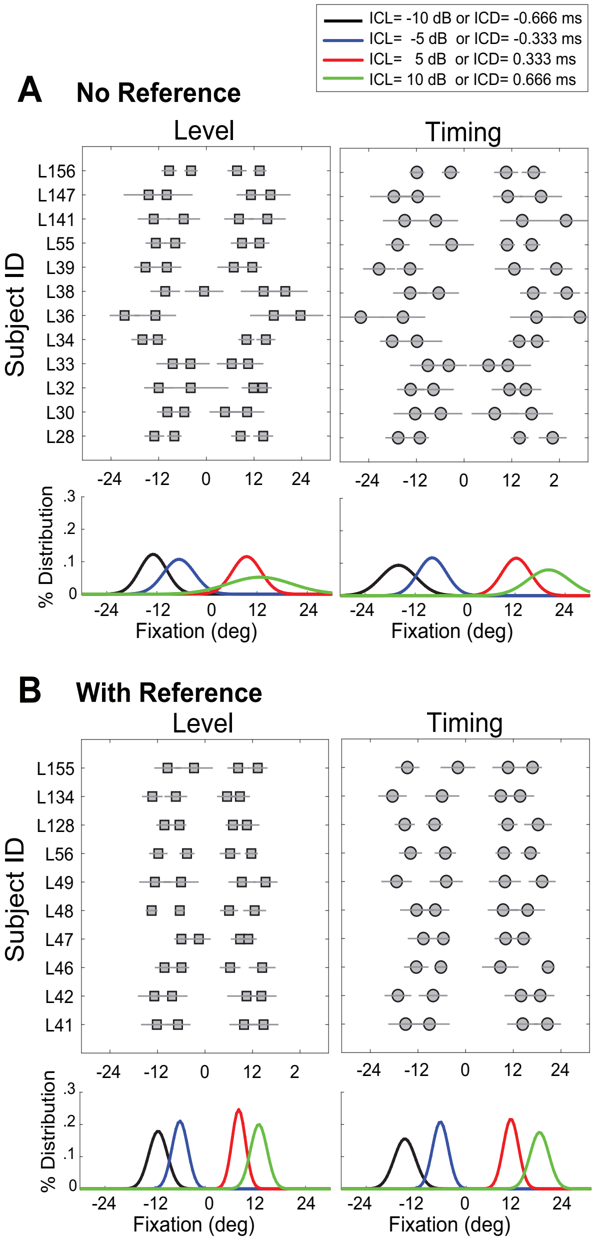

Figure 3. Auditory only response position. Mean response position and distribution of population responses for all participants in response to AO Level (Left) and AO Timing (Right) stimuli in the No Reference (A) and Reference (B) environments.

saccade is a more natural gaze orientation response that is guided by perceived target location, not the preset response button layout as used in our previous work.

Comparing the four experimental conditions in Fig. 3, we found that the perceived sound location appear to be different among different listeners for same stereo signal. Across the listeners, the spread of their mean fixation positions is wider for the timing signals than the level signals in both visual environments, suggesting more inter-subject variability in localizing the more binaurally ambiguous timing-signals. For the same signal types, the spread of mean fixations is narrower in the referenced (Fig. 3B) than no-reference environments (Fig. 3A), suggesting that the visual reference helps normalize the response range of fixation in localization a sound. These differences between stereo signals and between visual environments can be visually inspected based on the Gaussian fit of the population distributions shown in the bottom panels in Fig. 3 and the statistics of these distributions (mean and SD) are summarized in Table I.

Table I.I: Population Distribution of AO responses in No Reference and Referenced visual environments.

	Level (ICL in dB)	**-10**	**-5**	**5**	**10**	**Timing (ICD in ms)**	**-0.66**	**-0.33**	**0.33**	**0.66**
No Reference	mean	-12.9	-6.7	9.5	12.7	mean	-16.0	-7.9	12.3	20.0
	SD	3.3	3.7	3.4	7.6	SD	4.3	3.4	3.5	5.0
With Reference	mean	-11.3	-6.1	8.0	12.8	mean	-14.7	-6.2	10.7	17.6
	SD	2.2	1.9	1.6	2.0	SD	2.6	1.9	1.8	2.3

When additional visual signals were presented (AV condition), we observed the typical pattern of visual capture. However, the degree of visual bias varied across listeners. To demonstrate this inter-subject variability, Figure 4 shows response to all stimulus conditions (VO, AO, and AV) for four listeners who completed the experiments in both the “no reference” and “with reference” environments. Plots allow for comparison between level- and timing-based manipulations (top to bottom) and visual environment textures (left to right) within each listener.

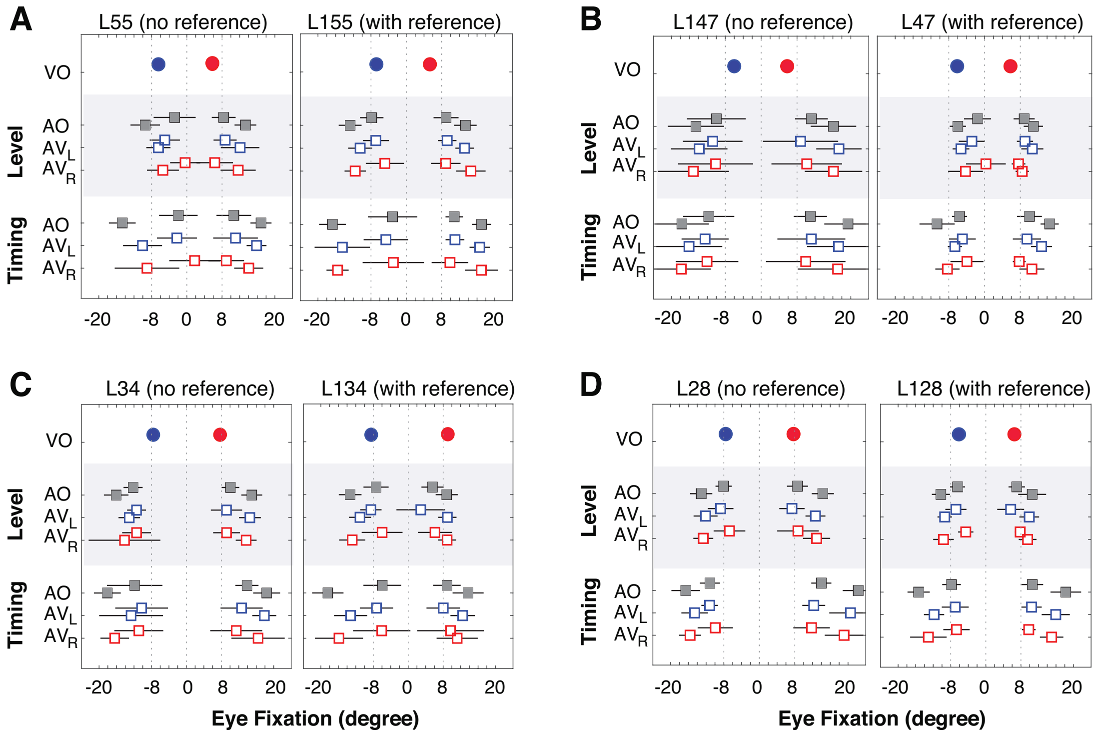

Figure 4. Individual response examples. Example responses for all stimulus conditions for four participants that completed the study in both the No Reference and Reference visual environments. Key: AO: Auditory Only, AVL: Auditory-Visual left light, AVR: Auditory-Visual right light

Again, for all listeners, changing level-based (ICL) and timing-based (ICD) stereo cues elicited responses in the left and right directions in both AO and AV conditions for both visual environments, as shown by the mean fixation positions associated with the four target stimuli. The effect of visual stimuli can be observed as the change in mean response position between AO and AV presentations. While the effect of visual stimuli does vary across stimuli, a general pattern may be observed. Overall, plots show that the “with reference” environment shows less visual bias as evidenced by the smaller observed change in mean response position between AO and AV presentations relative to the “no markers” environment. For these listener’s the degree of visual bias does not appear to change between timing and level signals.

Eye-fixation variability reveals the effects of binaural ambiguity and visual environment on stereo perception

Despite the similarity in the general patterns of mean responses across acoustic and visual conditions, response variability (SD), which is shown by the error bars in individual responses, is evident in fixation responses for all listeners in Figs. 3 and 4. To quantify the patterns of response variability, we analyzed the pooled results of SD in the AO (Fig. 5A) and AV condition (Fig. 5B). Two pairwise comparisons were made for both AO and AV results: 1) level-based vs. timing-based responses in the no reference environment (non-boldface boxplots) and 2) level-based vs. timing-based responses in the referenced environment (boldface boxplots). In the AO condition (Fig. 5A), there is a greater response variability for timing-based cues than level-based cues in both the no reference environment (t(11)= 2.62, p <0.05) and the reference environment (t(9)= 2.11, p<0.05). Similar observations were made for the AV results (Fig. 5B) for the no reference environment (t(11)= 3.82, p <0.01) and the reference environment (t(9)= 6.41, p<0.001). Between the two visual environments, the referenced environment led to an overall less response variability than the no reference environment in the AO condition. The two-sample comparisons showed a significant greater response variability in the no reference environment than the reference environment in response to both level signals (t(19)= 1.86, p <0.05) and timing signals (t(9)= 2.55, p<0.05). There were no significant differences in response variability between the visual environments in the AV condition (Fig. 5B).

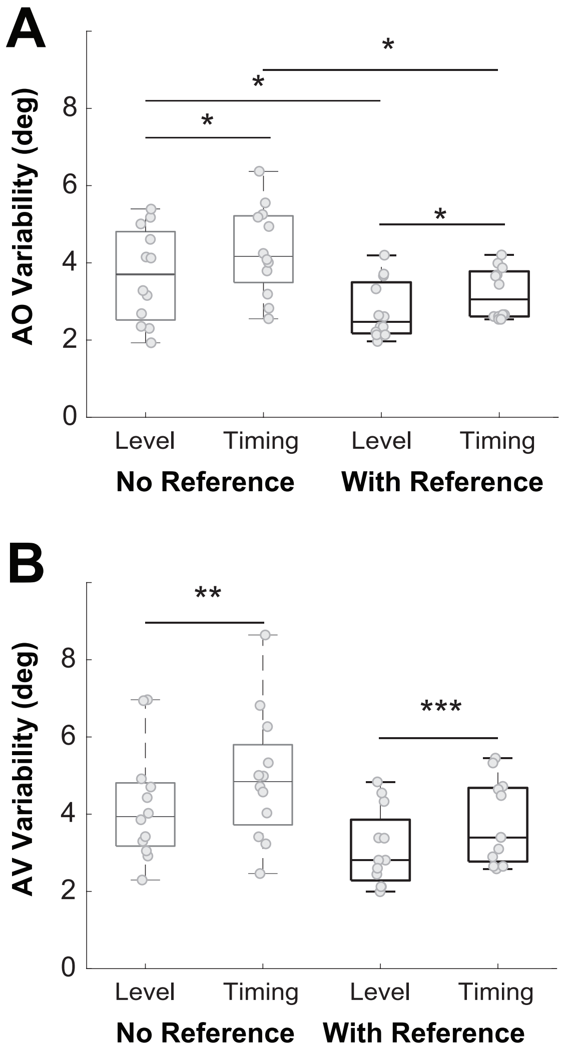

Figure 5. Response variability. Comparison of response variability between Level and Timing stimuli and No Reference and Reference visual environments for auditory only stimuli (A) and auditory-visual stimuli (B).

Eye fixation response reveals visual bias in stereo perception

Previous research has shown that spatially incongruent visual cues cause more errors in auditory localization based on button-pushing responses (Tomko and Proctor, 2017). Next, we analyzed whether eye fixation also reveals visual bias in auditory localization in this form (i.e., “spatial congruency rule”) and if so, to what extent the observed visual bias correlates with the response variability seen in the AO condition (i.e., “cue saliency rule”). In our analysis, the visual bias was measured as the average change in eye fixation position between AO and AV presentations. Figure 6 shows the distribution of visual bias of all participants when sound and light were presented on the same side of midline (6A) and opposing sides of midline (6B). Data is partitioned so that comparisons can be made between level and timing stereo stimuli and between the two visual reference environments. Our previously published button-pushing results are also shown, and these data were based on twenty-eight normal-hearing listeners pooled from three separate studies (Montagne and Zhou, 2016; Zhou et al. 2018; Venskytis et al. 2019).

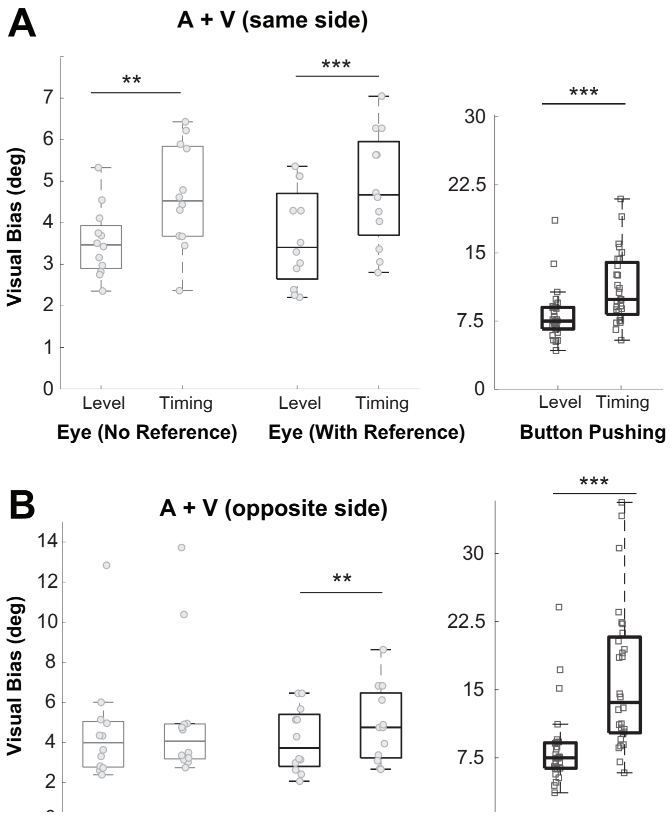

Figure 6. Visual bias. Comparison of visual bias between Level and Timing stimuli and No Reference and Reference visual environments when light and sound were on the same side of midline (A) and opposing sides of midline (B). Visual bias for button pushing responses from our previous experiments are also shown.

As we have reported in these early studies and shown here in the aggregated button-pushing results, timing signals lead to a stronger visual bias than level signals (same side: t(27)= 6.58, $p<0.001$; opposing side: t(27)= 9.87, $p<0.001$). Results of eye localization show the similar effects in all configurations with exception of the no reference opposing sides condition (same side No Reference, t(11)= 4.17, $p<0.01$; same side With Reference: t(9)= 9.18, $p<0.001$; opposing side No Reference: t(11)= 0.98, $p=0.35$; opposing side With Reference: t(9)= 2.89, $p<0.05$). In the Eye Saccade conditions, comparisons were also made between the two visual environment conditions (no reference and with reference), however none of the difference reached significance. Additionally, comparison was made between the AV alignment conditions (same side vs. opposing side) with only the button press timing condition showing a significant difference (t(27)= 6.42, $p<0.001$; final column top row compared to final column bottom row).

In our previous work, we found that timing signals render a greater binaural ambiguity between ITDs and ILDs. We attributed this binaural ambiguity as the underlying factor for the higher variability in AO responses and greater visual bias (Zhou et al. 2018). To evaluate the applicability of this “cue-saliency rule” in eye localization, we conducted the regression analysis between the magnitude of visual bias and AO response variability (SD). Results are summarized in Fig.7. Similar to those in Fig. 6, we partitioned the results into the AV same (top) and AV opposing (bottom) configurations based on “No Reference” (7A), and “With Reference” visual environment (7B) along with the aggregated button-pushing data (7C). A strong positive correlation is seen between AO variability and visual bias in response to both level-based stimuli in all configurations except the No Reference opposing side condition; See regression results in the figure legend. The regression analysis indicated that the uncertainly in auditory cues lead to a greater extent of visual capture in eye localization.

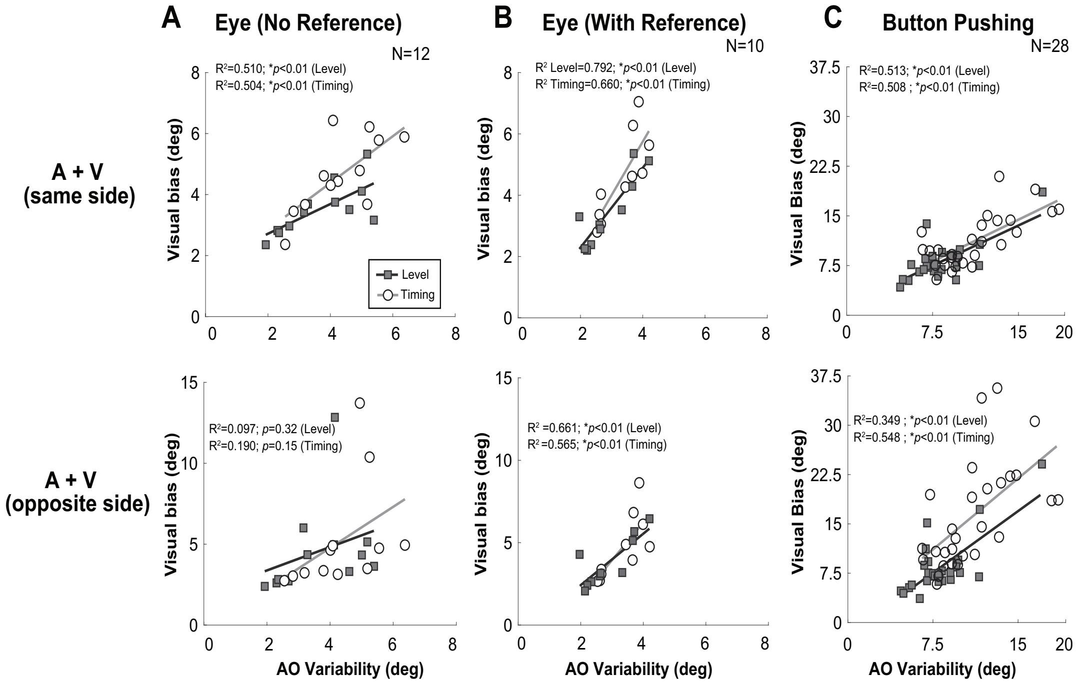

Figure 7. Between subjects: variability and visual bias. Correlation between the AO standard deviation and visual bias in responses of individual listeners for level-based (filled square) and timing-based (open-circles) when light and sound were on the same side of midline (top) and opposing sides of midline (bottom) in the no reference (A) and reference (B) visual environments. Results are also plotted for our previous button pushing experiments (C).

For both eye localization and button-pushing localization, there is a notable size of between-subject variability in AO response variability and visual bias. Thus, the estimate of a correlation between these two metrics could be influenced by the spread of data from individual subjects. Thus, we further analyzed for each participant, how the *change* in SD of AO responses is correlated with the *change* in |∆AV| between timing and timing stereo signals. From this with-subject measure, a positive correlation would indicate that a participant relies more on visual cues when auditory perception becomes uncertain. All change measures are obtained by subtracting the level results from the timing results on a trial-by-trial basis to capture the potential time-dependent changes, such as fatigue, during experiments.

Figure 8 shows this analysis for the No Reference (7A) and With Reference (7B) visual environments as well as in our previous button-pushing data (7C) in both AV same and AV opposite configurations. While both button pushing configurations showed a significant positive correlation between change in AO variability and change in visual bias (same side: $R^2=0.310$, $p<0.05$; opposing side: $R^2=0.300$, $p<0.05$) only the No Reference same side configuration was significant for the eye saccade responses (No Reference same side timing: $R^2=0.415$, $p<0.05$). One distinction between eye localization and button-pushing results is that in our early procedures, level-based and timing-based stimuli were tested in different experimental sessions, whereas in the present study both were interleaved within the same session. The weak correlation revealed in Figs. 8 A and B suggests that a listener's response variability in eye fixation is not calibrated on a stimulus-by-stimulus basis within a trial that contains stimuli with mixed saliency.

Inter and intra subject variability varies with response type

As demonstrated by others and our previous and current work, multiple motor response types can be used in conducting sound source localization experiments including eye saccades, finger pointing, and button pushing. Significant individual variability has been reported in these behavior measures. The different motor pathways involved in the execution of motor responses may contribute to the individual variability in results. We looked to determine if the motor responses used in our studies (eye fixation and button pushing) would result in different degrees of intra-subject variability when the two types of behavioral responses were elicited by similar stimuli and in the same experimental setting. Figure 8D shows the two-dimensional Gaussian fit of data shown in Figs 8A, 8B, and 8C. Three ellipses captured the confidence interval of 80% of the data distributions for the No Reference (red), Referenced (blue) and button pushing (black) data sets. The size of the ellipse indicates the spread of the intra-subject variability in both AO response variability and visual bias. A large amount of intra-subject variability is seen in the button pushing data set. This contrasts with the relatively small variability in the eye fixation data sets, as evidenced by much smaller changes in visual bias for the similar range of changes in AO response reliability, and vice versa. One potential explanation is that the variability seen in the button pushing may be due to the rescaling errors or different internal mapping between perceived sound direction and the button layout. The variability could also be due to the activation of additional motor pathways to execute a button press response.

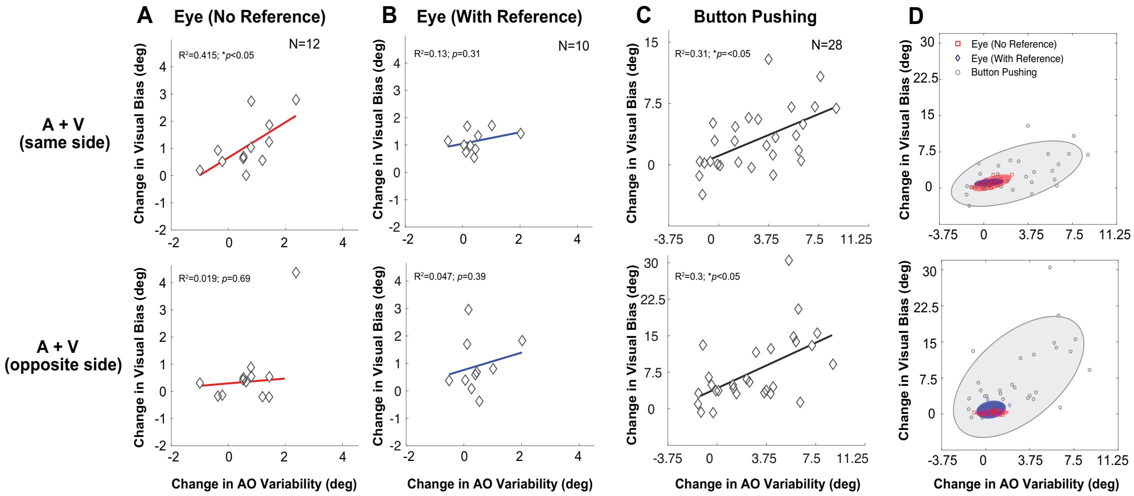

Figure 8. Within subjects: variability and visual bias. Correlation between the change in AO standard deviation and change in visual bias between level-based and timing-based manipulations within each individual listener's responses when light and sound were on the same side of midline (top) and opposing sides of midline (bottom) in the no reference (A) and reference (B) visual environments. Results are also plotted for our previous button pushing experiments (C). Figure 8D shows the two-dimensional Gaussian fit of data shown in A, B, and C.

Summary

Vision has a significant influence on the accuracy of auditory localization judgements. Visual influence varies between participants. The extent of this variability differs between level-based and timing-based manipulations and between the two visual environments. In general, for the same environment, the level-based stereo signals lead to less variability and visual bias than the timing-based signals. For the same signals, the referenced environment led to less variability than the no-reference environment.

EXPERIMENT 1B: EFFECTS OF STIMULUS SPATIAL SALIENCE ON RESPONSE TIME IN MULTISENSORY LOCALIZATION TASKS

Methods

The participants, apparatus, stimuli, and procedures used in this experiment were identical to those used in Experiment 1 (see Experiment 1A Fig. 1). Rather than examining the fixation position (response accuracy) of participants, this experiment focused on the timing of participant responses.

Data Analysis

Figure 1 shows typical saccade responses for one listener (L56) in response to left and right AO stimuli. Central fixation can be seen prior to the initiation of each trial. At the onset of the sound (time=0 sec), rapid saccadic eye movements can be seen in the direction of the auditory target. While maintaining target fixation, the participant confirmed the response with a button push (black circles). Two measures of saccade response time (RT) were calculated from the eye tracings: saccade start RT (red, open box) and saccade fixation RT (red, closed box). Saccade start RT is measured as the time from stimulus onset to initiation of the saccadic eye movement when the saccade amplitude has reach 20% of fixation amplitude. Saccade fixation RT is measured as the time from stimulus onset to the end of the saccadic eye movement when the saccade amplitude has reach 90% of fixation amplitude. If multi-stage saccades were generated in one trial (i.e., a participant initially responds in one direction but changes direction before making the final response) the endpoint of the final saccade was used in reporting saccade fixation RT. Analysis of saccade start RT and saccade fixation RT was conducted separately.

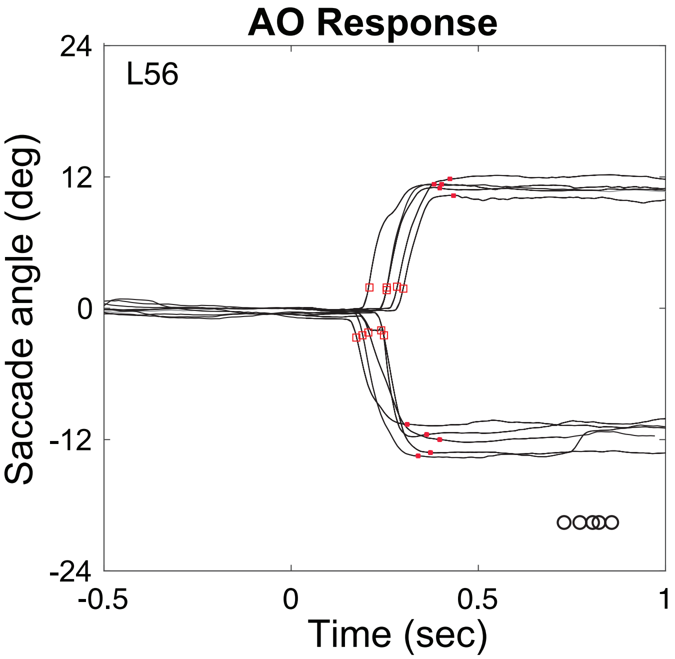

Figure 1. Saccade tracings. Saccade response to left and right AO stimuli for one participant (L56) showing saccade start RT (open box) and saccade fixation RT (filled box).

The mean and standard deviation (SD) of each listener's saccade start and saccade fixation RTs in response to each stimulus were extracted. These responses were used to analyze the effects of stereo signals on response time. RTs were also used to determine visual bias (reported as change in response time), which was calculated as the change between AV and AO response times (ΔAV= AV - AO) on a trial-by-trial basis for the same auditory stimulus. A positive change in RT would indicate that AV RT is greater (slower) than AO RT whereas a negative change in RT would indicate AV RT is less (faster) than AO RT. Data were further grouped and averaged across conditions for which light and sound were presented from the same side of the midline (same) and when light and sound were from opposing sides of the midline (opposite). For group analysis, the mean and standard deviation (SD) of each listener's responses for each stimulus were pooled. Again, 12 participants completed the task in the no reference visual environment and 10 participants completed the task in the with reference visual environment.

Results

The results of response times to an auditory or visual target were analyzed to address five questions: (1) Do less salient auditory stimuli result in longer absolute response times relative to more salient auditory stimuli? (2) Do the different visual environments ("no reference" and "with reference") result in different absolute response times to auditory or visual stimuli? (3) Do less salient auditory stimuli increase change in response time in eye localization similar to findings in our previous button-pushing experiments? (4) Does the spatial alignment of auditory and visual stimuli affect the change in response time in eye localization similar to findings in our previous button-pushing experiments? (5) Do the patterns of response times change between the two visual environments?

Auditory stimulus saliency and the visual environment do not affect absolute response times

Experiment 1A showed that the spread of participant mean fixation positions was wider for timing-based signals than level-based signal, suggesting more inter-subject variability in localizing the more binaurally ambiguous timing-signals. The analysis in Figure 2 looks to identify if this same pattern is true for response times (RTs) to auditory-only (AO) signals. Figure 2 plots both saccades start (2A) and saccade fixation (2B) RTs in the level condition against the timing condition for all listeners in the no reference (open, red circle) and with reference (filled, gray square) environments. Results reveal that there is no significant difference in either type of RT between level and timing presentations (grouping of data on the line). tested in either visual environment.

Comparison of the two visual environments in Experiment 1A showed that the with reference visual environment led to a decrease in response variability relative to the no reference condition. To further explore if this trend holds true for participant RTs,

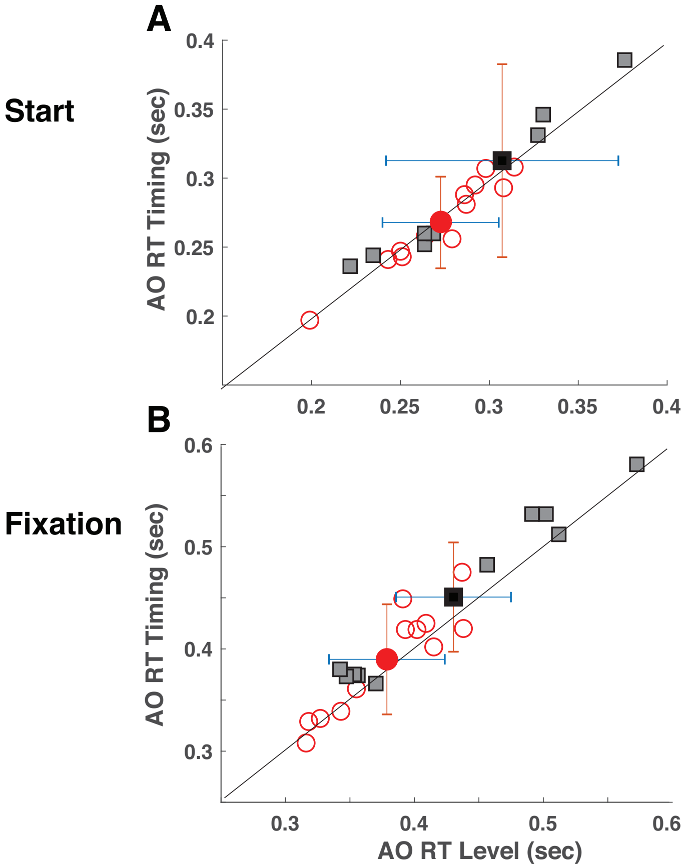

Figure 2. Absolute RT. Comparison of saccade start (A) and saccade fixation (B) AO RT in response to Level and Timing-based signals for the no reference (circle) and reference (square) environments.

group mean AO RTs were compared between the two environments. Figure 2A shows mean and SD of RTs for the no reference condition (filled, red circle) and with reference condition (filled, black square). While visual inspection suggests that the with reference condition may result in an increase in absolute response time, this minor difference does not reach to a statistically significant level.

Additional comparisons were made between AO saccade fixation RT and visual-only (VO) saccade fixation RT. Figure 3 shows the AO, VO comparison for AO level stimuli (3A) and AO timing stimuli (3B), again for both reference environments. Results demonstrate that for nearly all participants VO presentations resulted in faster RTs than AO presentations (grouping of data above the line). Analysis of group mean t-test

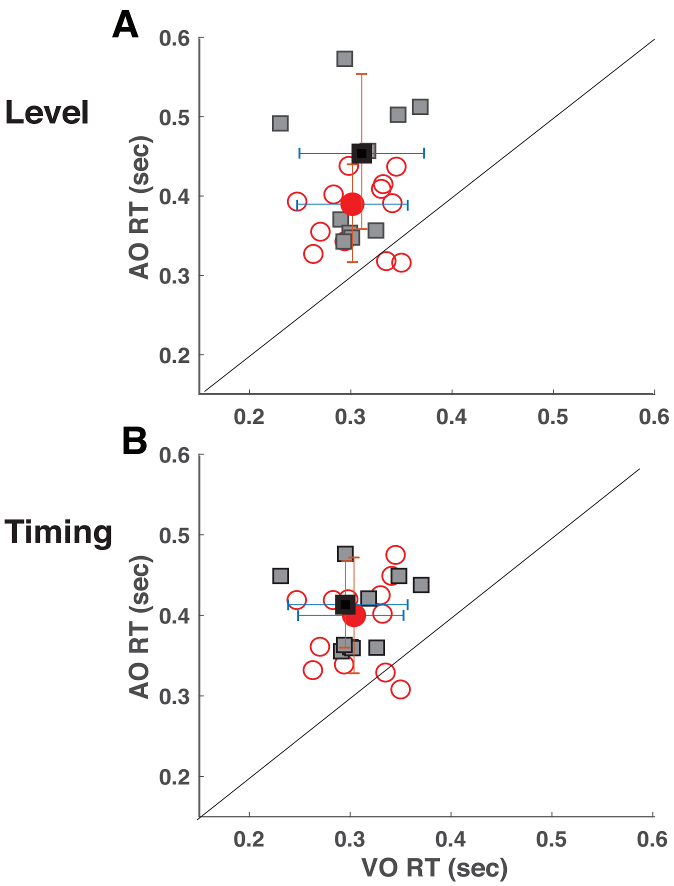

Figure 3. AO and VO RT. Comparison of saccade fixation RT between VO and Level AO (A) and VO and Timing (B) presentations for the no reference (circle) and reference (square) environments.

confirmed this significant difference in all conditions (level no reference: (t(11)= 2.58, $p<0.05$); level with reference: (t(9)= 2.91, $p<0.05$); timing no reference: (t(11)= 2.32, $p<0.05$); timing with reference: (t(9)= 2.61, $p<0.05$)).

Eye response time reveals the effects of stimulus spatial alignment on visual bias

Previous research has shown that spatially incongruent visual cues cause an increase in RT to an auditory target based on button-pushing responses (Tomko and Proctor, 2017). Next, I analyzed whether eye saccade RT also reveals a change in RT in auditory localization in this form (i.e., “spatial congruency rule”) and if so, if the observed change in RT change with binaural ambiguity in the auditory signal. Figure 4 shows the pooled results of change in RT (calculated as AV RT- AO RT) for saccade start (4A) and saccade fixation RT (4B). Each panel provides 3 primary comparisons: 1) level vs. timing stimuli, 2) conditions in which A and V stimuli were presented on the same side of midline (Same) vs. opposing side of midline (Opposite), and 3) no reference visual environment vs. with reference visual environment.

In comparing change in RT between timing and level stimuli, no significant pattern is observed. This is in contrast to RT data shown form our previous button pushing experiment which shows a significant increase in change in RT in response to timing signals stimuli to level stimuli (t(27)= 2.32, $p<0.05$; t(27)= 3.25, $p<0.05$). As discussed previously, one distinction between eye localization and button-pushing results is that in our early procedures, level-based and timing-based stimuli were tested in different experimental sessions, whereas in the present study both were interleaved within the same session. The lack of influence of stimulus salience on change in RT in the eye-saccade data may be due to these differences in procedure.

In contrast to the level vs. timing analysis, comparison of change in RT between same and opposite AV alignments did reveal a significant pattern of change in our results. For saccade start RT, A and V stimuli presented from opposing sides of midline resulted in a greater change in RT relative to those presented from the same side of midline in all conditions. For saccade fixation RT, the same pattern was seen but only in the reference visual environment. Table II.I shows the results of the statistical analysis. This pattern of results was consistent with what was previously measured in the button pushing experiment and would suggest that a greater spatial disparity between the two stimuli result in an increased cost in processing time.

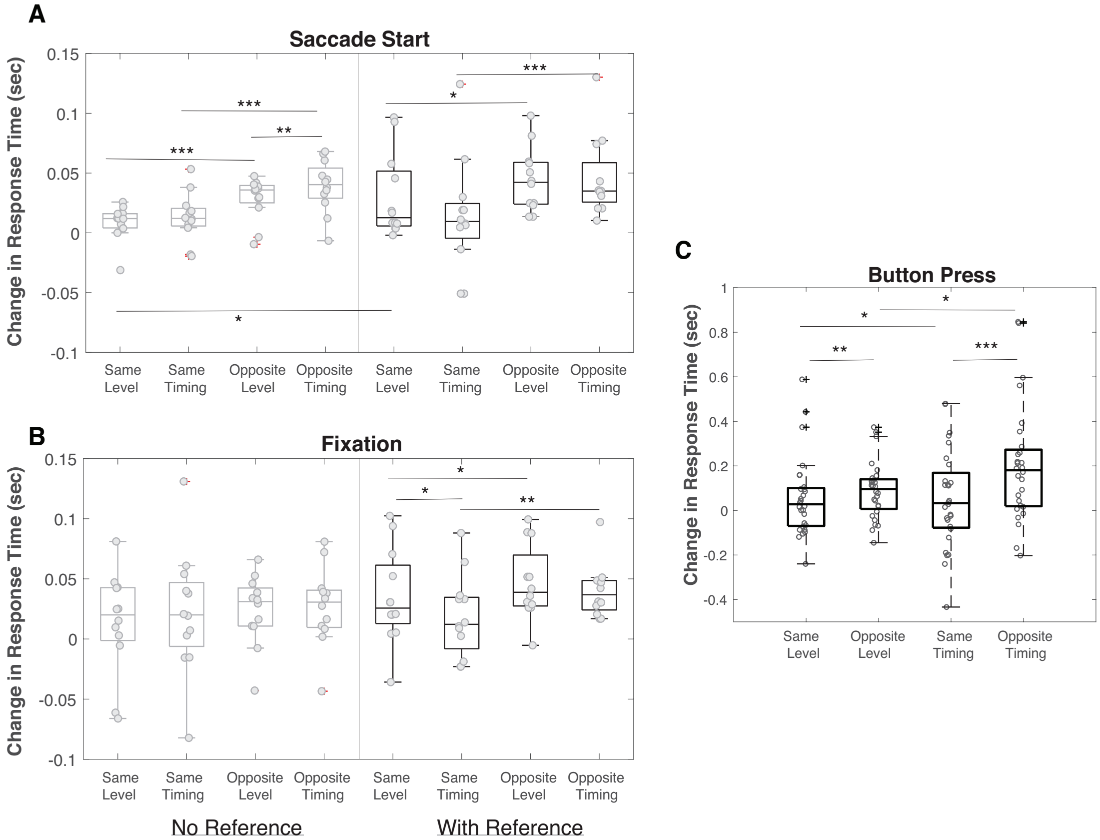

Figure 4. Change in RT. Comparison of change in RT between Level and Timing stimuli, Same and Opposite alignment, and No Reference and Reference visual environments for saccade start (A) and saccade fixation (B) RT. Comparisons of change in RT between Level and Timing stimuli and Same and Opposite alignment are also shown for previous button pushing data (C).

Table II.I: Same vs. Opposite Change in RT Statistics

	Saccade start		**Saccade fixation**		**Button Press**	
	Same vs. Opposite		Same vs. Opposite		Same Vs. Opposite	
	Level	Timing	Level	Timing	Level	Timing
No Reference	$t(11)=6.22$; $p<0.001$	$t(11)=6.61$; $p<0.001$	$t(11)=1.21$; $p=0.25$	$t(11)=0.50$; $p=0.62$	$t(27)=3.33$; $p<0.01$	$t(27)=7.54$; $p<0.001$
Reference	$t(9)=1.92$; $p<0.05$	$t(9)=4.32$; $p<0.001$	$t(9)=2.42$; $p<0.05$	$t(9)=2.89$; $p<0.01$		

Finally, change in RT between the two visual environments was evaluated. Overall, no significant pattern of difference in change in RT was shown between no reference and with reference environments. Suggesting that overall, the visual environment did not affect the pattern of influence of an additional visual stimulus on the timing of an auditory localization response.

Change in response time does not consistently vary with auditory variability

Our previous studies and Experiment 1A, found that timing signals render a greater binaural ambiguity between ITDs and ILDs. This binaural ambiguity was attributed as the underlying factor for the higher variability in AO responses and greater visual bias (Zhou et al. 2018). Further analysis showed that variability in AO responses was significantly correlate with visual bias in accuracy responses. To further evaluate the applicability of this “cue-saliency rule” in eye saccade response times, a regression analysis was conducted between change in RT and AO response variability (SD). Results are summarized in Fig.5. Like Fig. 4, results were partitioned between saccade start RT (A) and saccade fixation RT (B). In contrast to what was shown in the accuracy experiment, there was no significant correlation between AO variability and change in RT in any condition (see regression results in the figure legend).

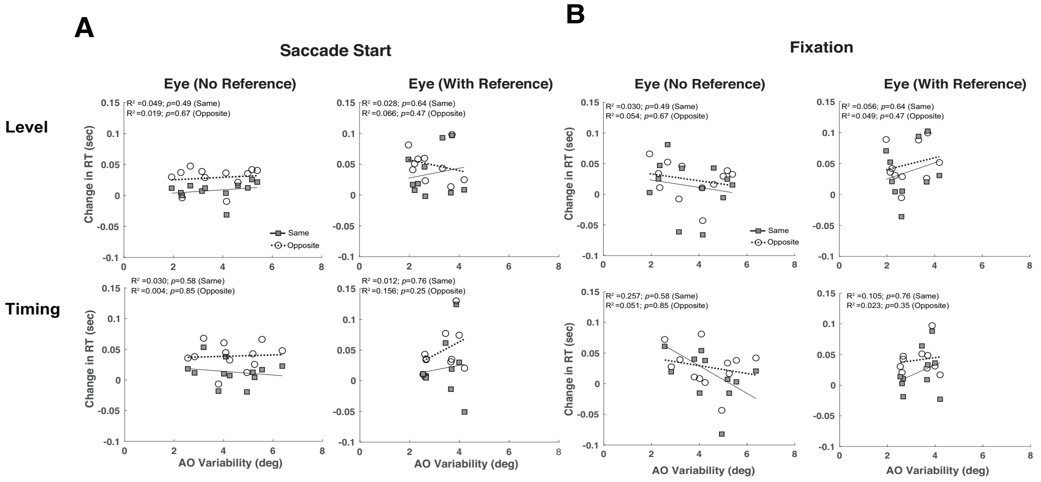

Figure 5. Between subjects: variability and change in RT. Correlation between AO standard deviation and change in RT of individual listeners for AV same side (filled square) and AV opposite side (open circles) conditions for level (top) and timing stimuli (bottom) in the no reference and with reference environments. Saccade start (A) and fixation (B) RTs were analyzed separately.

As previously noted, for both eye localization and button-pushing localization, there is a notable size of between-subject variability in AO response variability and change in RT. Thus, the estimate of a correlation between these two metrics could be influenced by the spread of data from individual subjects. Thus, we further analyzed for each listener, how the *change* in SD of AO responses is correlated with the *change* in |∆AV| between timing and timing stereo signals. From this with-subject measure, a positive correlation would indicate that a listener relies more on visual cues when auditory perception becomes uncertain. All change measures are obtained by subtracting the level results from the timing results on a trial-by-trial basis to capture the potential time-dependent changes, such as fatigue, during experiments.

Figure 6 shows this analysis for the No Reference (saccade start (6A); saccade fixation (6B)) and the With Reference (saccade start (6C); saccade fixation (6D)) visual environments as well as our previous button-pushing data (6E) in both AV same and AV opposite configurations. While both button pushing configurations showed a significant negative correlation between change in AO variability and change in RT, no such significant correlation was found in the eye RT data.

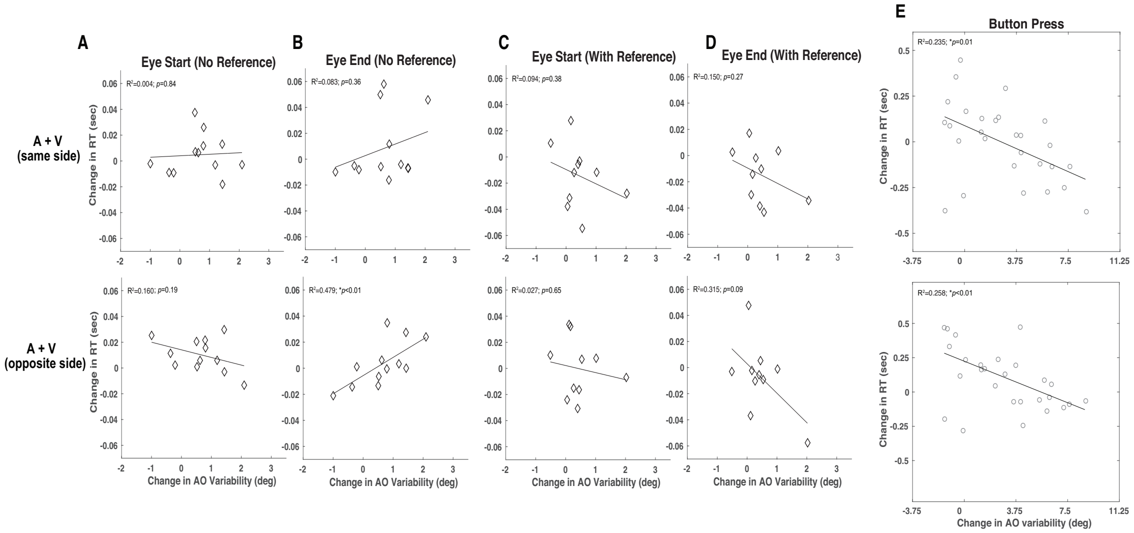

Figure 6. Within subjects: variability and change in RT. Correlation between change in AO standard deviation and change in RT (Timing-Level) of individual listeners for AV same side (top) and AV opposite side (bottom) conditions for level (top) in the no reference and with reference environments. Saccade start (A) and fixation (B) RTs were analyzed separately. Button pushing data from our previous experiments is also shown.

Response times reveal idiosyncratic differences in behavior

A consistent trend in auditory research is significant individual variability in behavior measures. As demonstrated by others' and our previous and current work, multiple motor response types may be used in conducting sound source localization experiments including eye saccades, finger pointing, and button pushing. The different motor pathways involved in the execution of these motor responses may contribute to the individual variability in results. This variability was shown to a degree in the accuracy data presented in Experiment 1A and is again seen in the RT data presented here. Figure 8 shows an example of this variability in change in RT for two participants. While the addition of visual stimulation results in speeded response times relative to AO stimuli for one participant, the opposite is true for the other, showing a delay in RT in response to AV stimuli. These conflicting patterns of results may increase the difficulty in observation of group trends and highlight the importance of considering how AV interactions occur at the individual level.

Summary

Spatial alignment of AV cues has a significant impact on the speed of auditory judgments with spatially separated signals resulting in an overall delay in response time. Significant variability/idiosyncratic differences in participant behavior were also observed in response to AV signals with some listeners showing speeded RT in response to AV presentations and others showing delayed RT in response to AV presentations making it difficult to quantify conclusive patterns in group behavior.

EXPERIMENT 1C: EFFECTS OF STIMULUS SPATIAL SALIENCE ON RESPONSE ACCURACY AND RESPONSE TIME IN MULTISENSORY LOCALIZATION TASKS IN LISTENERS WITH NORMAL HEARING AND HEARING IMPAIRMENT

Methods

The apparatus, stimuli, procedures, and analysis used in this experiment were identical to those used in Experiment 1 and Experiment 2 (see Figure A1, A2, and B2). Six additional participants with bilateral, mild to moderate sensorineural hearing loss were tested in the no reference and with reference environments. Their results were compared to the six normal hearing participants who also completed the study in both visual environments.

Participants

Six participants with hearing impairment completed this study (age 25-32; 2 males). Their demographic information can be seen in Table III.I. All participants were long time hearing aid users (>4 years). Audiometric thresholds were verified using standard audiometric techniques under insert earphones prior to experiment participation. Figure 1 shows the audiometric results for all participants.

Table III.I: Hearing loss population demographics.

ID	Gender	Age (years)	Length	HL Type
L50	M	32	Congenital	Sensorineural
L53	F	25	7 years	Sensorineural
L59	F	28	5 years	Sensorineural
L60	F	27	Congenital	Sensorineural
L61	M	25	~15 years	Sensorineural
L62	F	31	Congenital	Sensorineural

Compensation for hearing loss

All participants with hearing loss completed the study without their hearing aids in two stimulus conditions, uncompensated and compensated. In the uncompensated stimulus condition, no gain adjustment was made to t test stimulus to compensate for the hearing loss of the participant. The stimulus used was set to a level of 65 dB SPL, the same stimulus level used when testing normal hearing listeners. In the compensated stimulus condition, the gain of the stimulus was adjusted to account for the degree of hearing loss for each participant. The gain adjustment was calculated based on recommendations from the National Acoustics Laboratory (NAL) NL2 fitting formula commonly used in hearing aids, which seeks to amplify gain at individual frequencies to maximize speech perception. The recommended NAL-NL2 gain compensation was calculated for each participant individually and applied to the test stimulus.

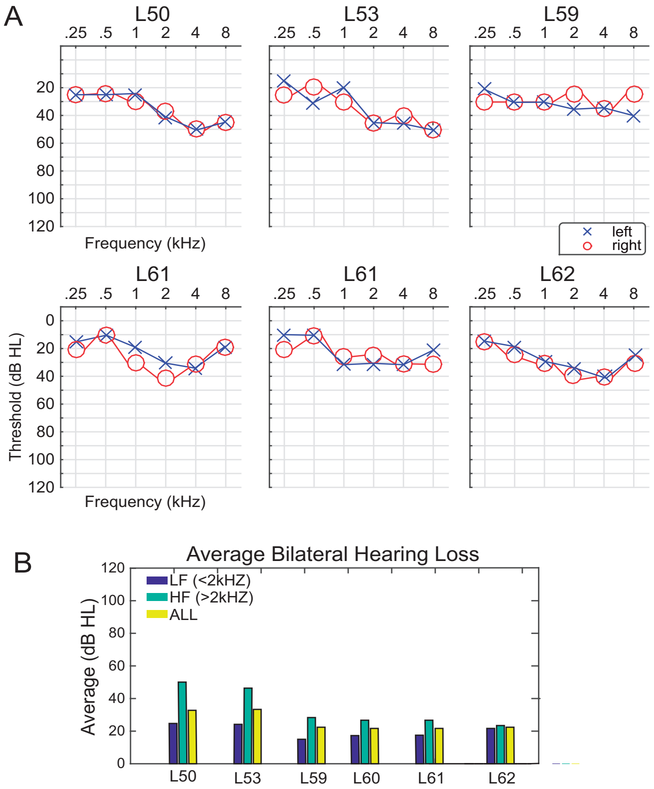

Figure 1. Audiograms HI listeners. (A) Hearing levels, at six frequencies, in the left and right ears for all HI listeners. (B) Average dB HL at LF, HF, and ALL ranges. (LF=low frequency; HF=high frequency; ALL=all-frequency; dB HL=decibels in hearing level)

Results

The purpose of analysis was to compare localization response accuracy and response timing between participants with normal hearing (NH) and hearing impairment (HI). Results address 3 primary questions: (1) Does hearing impairment result in increased variability in eye fixation position and does this variability change with gain compensation? (2) Does hearing impairment result in increased visual bias in eye fixation position and does the degree of visual bias change with gain compensation? (3) Does hearing impairment result in increased absolute and/or change in response time and do response time patterns change with gain compensation? Current analysis only shows results from the no reference visual environment.

Eye fixation response reveals stereo perception

Similar to the NH participants in Experiment 1A, participants with HI showed the ability to complete the stereo localization task in both the uncompensated and compensated stimulus conditions. Figure 2 shows responses to all stimulus conditions (VO, AO, and AV) for one listener in the uncompensated (left) and compensated (condition). Overall, participants showed a consistent ability to indicate the perceived direction of a sound source using eye saccades and were able to distinguish between the 4 different level and 4 different timing signals in both the AO and AV conditions, although, vary degrees of visual bias were observed.

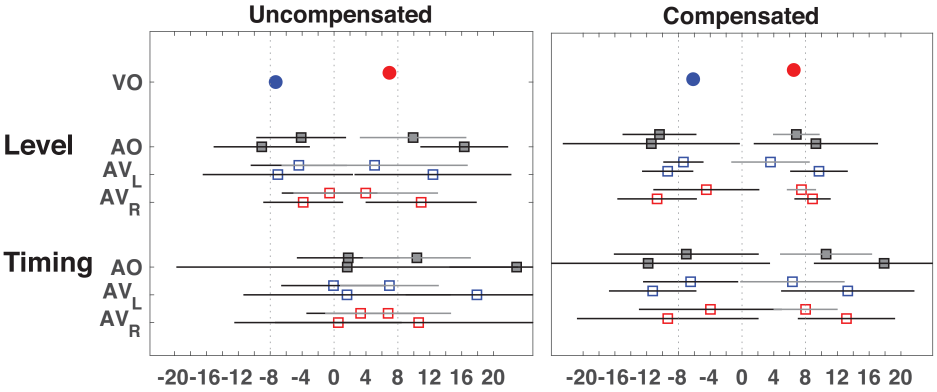

Figure 2. Individual response examples. Example responses for VO, AO, and AV conditions for one participant in the uncompensated and compensated stimulus conditions. Key: AO: Auditory Only, AVL: Auditory-Visual left light, AVR: Auditory-Visual right light

Hearing impairment results in increased response variability and visual bias

While participants with HI showed the ability to localize the target sound sources in all conditions, significant variability can be observed in individual responses (see Fig. 2 error bars for example). Analysis looked to determine if response variability in the HI population was significantly increased relative to the NH population and if gain compensation had a significant effect on variability. Figure 3 shows AO variability for NH participants (black diamond), HI participants in the compensated condition (blue square), and HI participants in the uncompensated condition (red circle). Individual responses are shown as open symbols and group mean responses are shown as filled symbols. A significant difference in AO variability is seen between all three groups with the uncompensated HI condition resulting in the largest degree of AO variability, the NH group showing the smallest AO variability and the compensated HI condition falling between the two (see Table III.II in Appendix for ANOVA and Tukey multiple comparison results). This would suggest that HI resulted in increased uncertainty in the auditory localization task and that while compensating for the HI in the gain of the signal did decrease response variability, performance with the compensated signal did not reach the level of performance of NH participants.

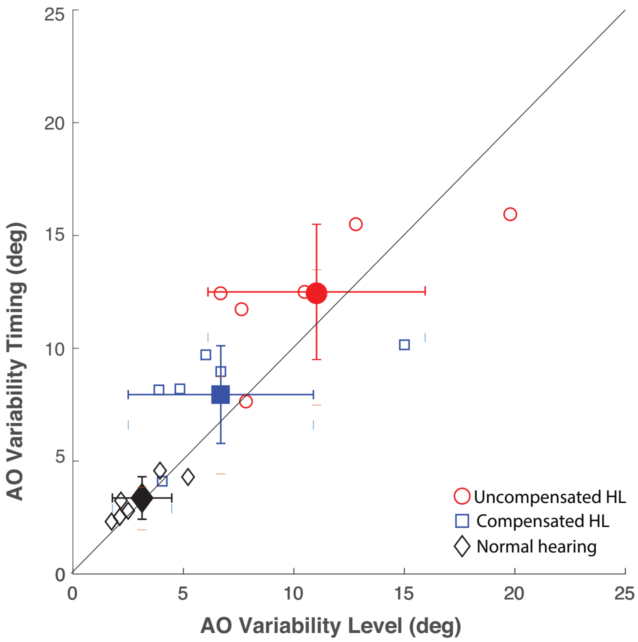

Figure 3. Response position variability. Comparison of response variability between Level and Timing stimuli for NH listeners and HI listeners (uncompensated and compensated stimuli).

The same pattern of performance is also seen in analysis of visual bias in Figure 4. Results show magnitude of visual bias for participants with NH and HI in the compensated and uncompensated conditions. Results were separated between AV presentations on the same side of midline (4A) and opposing sides of midline (4B). Again, the uncompensated condition resulted in the largest degree of visual bias, followed by the gain compensation condition, and finally the NH group. As with AO variability, a significant decrease in visual bias was seen following gain compensation, however, visual bias did not drop to the level of the normal hearing group. (see Table III.III in Appendix for ANOVA and Tukey multiple comparison results).

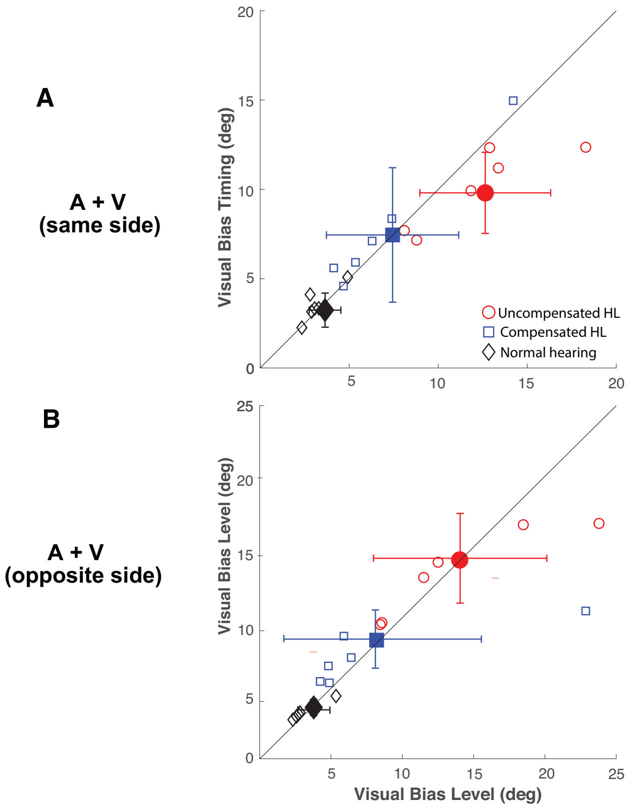

Figure 4. Visual Bias. Comparison of visual bias between Level and Timing stimuli for NH listeners and HI listeners (uncompensated and compensated stimuli).

Hearing impairment results in increased absolute response time

In addition, showing increases in response variability and visual bias relative to NH participants, HI participants also showed a significant increase in overall response time (RT). Figure 5(A,B) shows absolute saccade start (A) and saccade fixation (B) RTs to AO stimuli for all three groups. Overall, a significant increase in RT is seen in both the HI uncompensated and HI compensated conditions relative to the NH group. However, unlike with the response accuracy measures, gain compensation for the HI did not result in a significant reduction in RT relative to the uncompensated condition. Figure 5 (C,D) shows the comparisons of AO saccade fixation RTs for Level (C) and Timing (D) stimuli

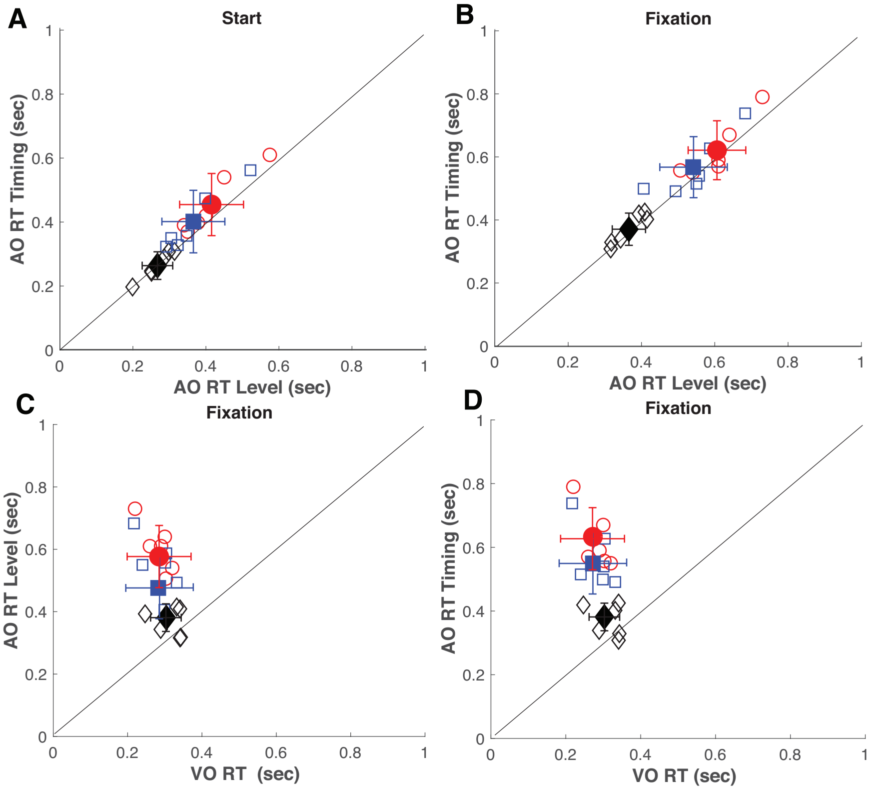

Figure 5. Absolute RT. Comparison of absolute RT between Level and Timing stimuli for NH listeners and HI listeners (uncompensated and compensated stimuli) (top) and AO and VO absolute responses times for the same groups (bottom).

to VO saccade fixation RTs to evaluate for difference between NH and HI groups. I initially hypothesized that HI participants may show a decrease VO RTs relative to NH participants due to central compensation. While VO RT was significantly faster than AO RT for all groups, analysis did not show a significant difference in VO RT between the three groups.

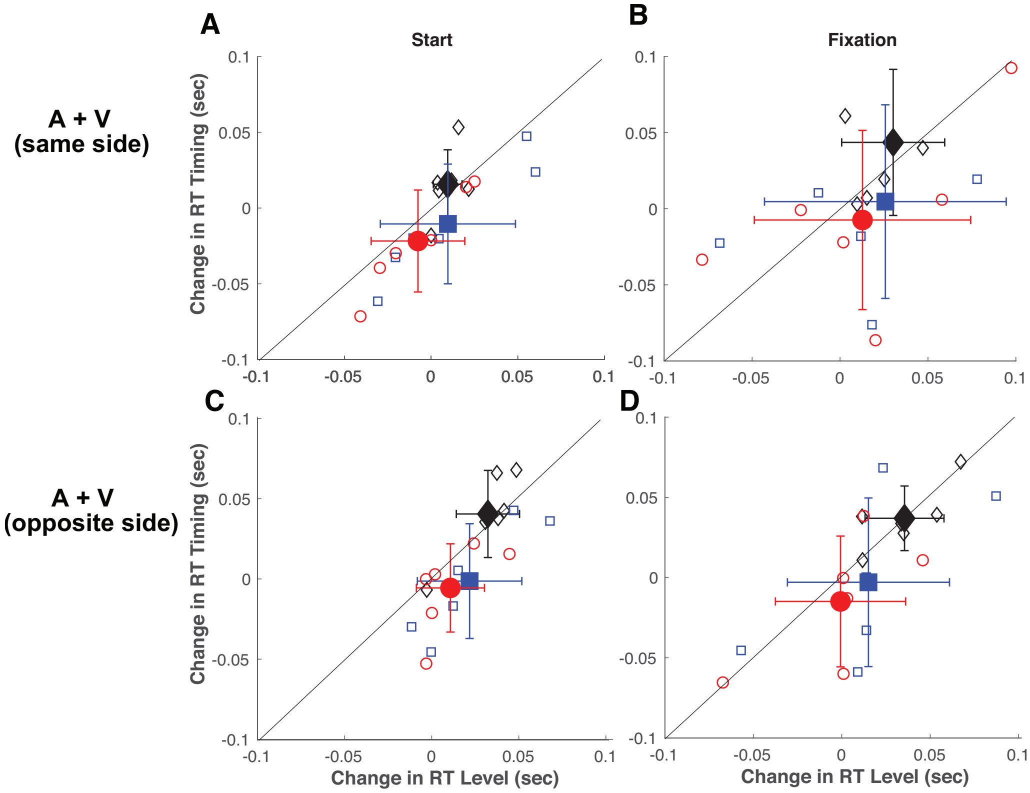

Figure 6. Change RT. Comparison of change in RT between Level and Timing stimuli for NH listeners and HI listeners (uncompensated and compensated stimuli). Analyses were separated between trials for with AV stimuli were on the same side of midline (top) and opposite sides of midline (bottom). Saccade start RTs are shown on the left and saccade fixation RTs are shown on the right.

To further evaluate for difference in the effects of vision on response times between the NH and HI participants, change in RT between AV and AO presentations was analyzed. Comparisons are shown in Figure 6 (recall that change in RT is calculated as (AV RT-AO RT, therefore a positive residual would suggest the addition of a visual cue delays RT and a negative residual would suggest that the addition of a visual cue speeds RT relative to AO). While visual observation suggests that the addition of a visual cue would speed up responses in the HI population (see small or negative residuals) and delay responses in the NH population (see larger residuals), there is significant variability seen between individual participants. Further analysis of difference between group means failed to reach significance.

Idiosyncratic differences in behavior are seen in both NH and HI participants

As demonstrated in the previous two experiments for NH participants, a large degree of individual variability is observed in the response patterns of the HI group, especially for measures of RT (Fig. 7). While the addition of visual stimulation results in speeded response times relative to AO stimuli for one participant, the opposite is true for the other, showing a delay in RT in response to AV stimuli. I would again suggest that these conflicting patterns of results may increase the difficulty in observation of group trends and highlight the importance of considering how AV interactions occur at the individual level.

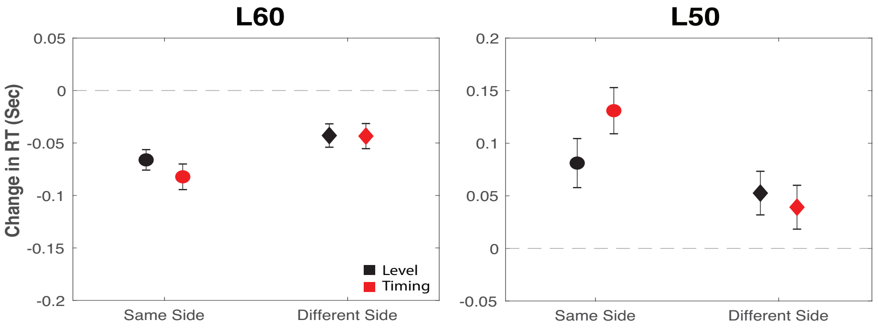

Figure 7. Individual differences. Comparison of change in RT (AV-AO) between two participants (L60 and L50). While one participant shows

Summary

Hearing impairment results in an increase in response variability, visual bias, and response time in auditory localization tasks relative to NH participants' results. While a slight decreases in variability, bias and response times are seen when additional gain is added to the target stimulus to compensate for the hearing loss, performance with the gain compensated signal does not reach the level of performance of NH participants. Significant variability/idiosyncratic differences in participant behavior were again observed in response to AV signals, especially for in RT results, with some participants showing speeded RT in response to AV presentations and others showing delayed RT in response to AV presentations making it difficult to quantify conclusive patterns in group behavior.

CHAPTER 4

EXPERIMENT 2: EFFECTS OF STIMULUS SPATIAL ALIGNMENT ON RESPONSE ACCURACY AND TIME IN MULTISENSORY LOCALIZATION TASKS

Introduction

Sensory experience builds upon a multisensory analysis of many cues that all describe properties of the same objects in an environment (Stein & Meredith, 1993). This multisensory strategy helps build an internal map of the physical environment to enhance perceptual analysis (e.g., who is talking, where is (s)he?) and guide actions (e.g., escape or attack). Sound source localization is an excellent example of this strategy. In everyday interactions with the environment, our initial reaction to the sudden onset of an unexpected sound (e.g., a quickly-passing vehicle) is to estimate its source location to calibrate a reaction better. The speed and accuracy of this reaction depend on the coordination of auditory and visual spatial functions. Vision's effects on auditory localization are manifested in both the spatial and temporal response domains. When sound and light are presented simultaneously, but from different locations, (1) the direction of a light biases the perceived location of a sound source (Jackson, 1953; Pick, Warren & Hay, 1969; Thurlow & Jack, 1973), i.e., the "Ventriloquist Effect", and (2) a flash of light delays auditory localization to a greater extent than sound delays visual localization (Rago, Cave & Fano, 1988; Lukas, Phillip & Koch, 2010). In a simple detection task, visual percepts can overwhelm audition to such an extent that some subjects do not even notice the presence of a sound, i.e., the "Colavita Effect" (Colavita, 1974).

While it is commonly accepted that multisensory integration becomes stronger when sensory inputs always align. For foveal species like humans and monkeys, the visual field is restricted to the frontal space, whereas the panoramic auditory field covers the entire frontal and rear space. The rear sensitivity provided by spatial hearing is critical for avoiding unseen danger coming from behind. Rear space, however, has been largely overlooked in multisensory research. Our recent study shows that in humans, vision's influence on auditory perception can extend to unseen rear space and interacts closely with auditory front-back confusions (FBCs) in auditory localization (Montagne & Zhou, 2018). The FBC errors are responses to the approximately correct angular displacement (relative to the midline) but in the wrong front-back hemifield (Carlile, Leong & Hyam, 1997; Macpherson & Middlebrooks, 2000). While interaural time and level differences (ITDs and ILDs) offer the primary information about the horizontal angle of a sound source (Middlebrooks & Green, 1991), front-back errors are rooted in the fact that these binaural difference cues do not, on their own, correspond with only one sound source location, but a host of them along the "cones of confusion" (Blauert, 1997). We found that frontal visual cues can reduce FBC errors for frontal auditory targets and increase FBC errors for rear auditory targets due to visual capture to the frontal field (Montagne & Zhou, 2018). What remains unknown, however, is whether vision also affects response time (RT) in panoramic auditory localization. This study addresses this question. The test involves a dual-response task, where a listener was asked to localize the lateral, left-right direction of a free-field sound using eye movement and immediately after, decide the front-back direction of the sound by pressing one of two buttons arranged vertically on a game console. The results show that change in RT is correlated with the perceived modality of stimuli. For participants following the sound direction, a visual LED stimulus delayed their RT, whereas for participants following the light direction, light speeded up their RT, relative to auditory-only responses. The changes in RT were more significant for rear sounds.

Methods

Participants

Eight participants completed all sessions (3 male, age range from 19-34 years old, median 21.5 yrs). All were reported right hand dominant. All participants had normal hearing sensitivity as verified by standard audiometric techniques under insert earphones and reported normal or corrected-to-normal vision. Participants provided written, informed consent, and received financial compensation for their participation. The experiment was conducted in accordance with procedures approved by the Arizona State University's Institutional Review Board.

Apparatus and Stimuli

The free-field sound localization task was conducted in a double-walled, sound-attenuated chamber (Acoustic Systems RE-243, 2.1 m x 2.1 m x 1.9 m) lined with 3" acoustic foam. The participant was seated in the center of the sound chamber with his/her head stabilized using a high-precision head positioner (HeadLockTM, Arrington research). Sound stimuli were presented from four loudspeakers (full-range monitor, Adams F5, positioned at +/− 35° and +/− 145° at a distance of 1.1 m) hidden behind a black, acoustically transparent curtain. Light stimuli were delivered via a high-resolution LED bar attached to the acoustic curtain spanning +60° to −60°. For this experiment, only three active LED locations were used for testing (0°, +/− 8°). All sound and light stimuli originated from 0° elevation relative to eye level.

Custom-designed software written in MATLAB generated and controlled the auditory stimuli and recorded participant responses. Auditory stimuli were broadband noise bursts gated on and off with a 15-ms rectangular window. Identical signals were presented from the two front or two rear loudspeakers. The perceived sound source location was manipulated by changing the level difference (panning stereophony) between either the front-left (LF) and front-right (RF) or between the back-left (BL) and back-right (BR) loudspeaker signals. Level differences of −5, 0, and 5 dB were used, with negative intensity ratios indicating that the left speaker signal was more intense than the right. This panning-based stereophonic technique generated a phantom sound source at ~ +/− 10 degrees and 0 degrees from both front and back. The average intensity for all auditory stimuli was maintained at 65 dBA, as verified using a sound level meter (Brüel & Kjær 2250-L) positioned at the location of a listener's head.

Custom-designed software written in Arduino controlled the visual stimuli. Visual stimuli were 15-ms duration blue LED light flashes (13.5 cd/m2, measured at the LED surface) generated from one of three lighting positions. The onset of the light stimulus was synchronized with the onset of the sound stimulus at the leading loudspeaker as verified by oscilloscope measures.

Procedure

Two experimental conditions (auditory localization and visual localization) were conducted over five visits over five separate days. On days one and two, participants completed two auditory conditions followed by one visual condition. On days three through five, participants completed three auditory conditions followed by one visual condition. Each auditory and visual condition lasted approximately ten minutes and two minutes, respectively. Participants were given a three- to five-minute break period between conditions. Before beginning the experiment each day, participant eye tracking was calibrated using 13 known LED locations (from -12^{o} to $+12^{o}$ at 2^{o} increments). This report focuses on the effects of vision on auditory localization; results for visual localization are not reported.

Before the experiment, to help participants understand the nature of the experiment and the response tasks expected a training block for auditory localization was presented. Participants were presented with a sound from one of six locations (three in front and three in back) and were instructed to respond by 1) directing their gaze to the lateral direction of a sound and 2) pressing a button to indicate front or rear origin of the sound source. Participants were provided with alert feedback, following each erroneous front-back judgment, to aid in the proper performance of the front/back decision-making portion of the task. The training block continued until participants achieved >75% accuracy and felt ready to start an experiment.

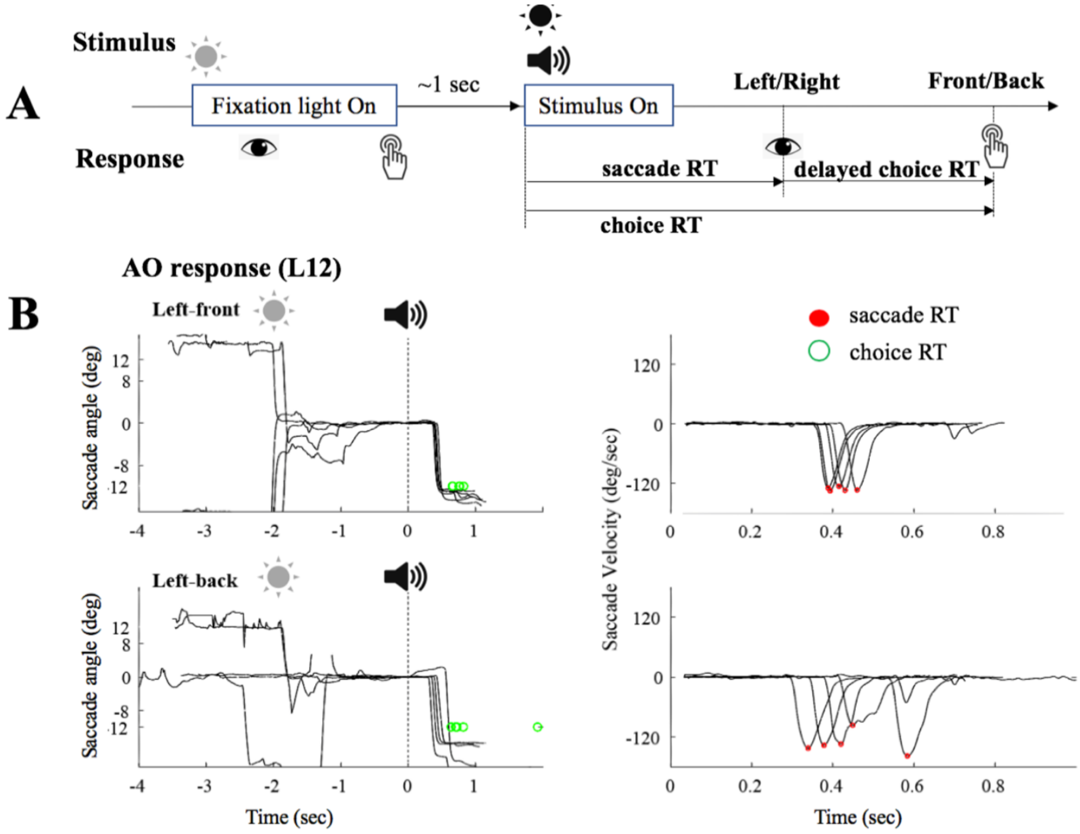

Figure 1. Experimental procedure and reaction time analysis. (A) Stimulus-response sequences for each trial in the auditory localization task. (B) Example saccade and choice responses in response to auditory-only stimuli. Left, saccade angles to left-front and left-back sounds. Right, saccade velocity to these two stimuli. Light gray indicates the start time of center fixation light, the stimulus onset is marked as time zero.

Auditory Localization Task

Sound stimuli with or without lights were presented in randomized blocks, denoted as audio-visual (AV) and auditory-only (AO) blocks, respectively. The AO block contained six stimuli total (three in front and three in back). For the AV block, each of the six auditory stimuli was paired with one of the three LEDs, resulting in a total of 18 stimuli. The order of blocks and stimuli was randomized. Five repeats were administered for each stimulus, resulting in a total of 120 trials. For all trials, participants were instructed to indicate only the direction of the sound they heard, not the light they saw. Figure 1A shows the stimulus and response sequences within a trial. To initiate a trial, a participant moved her/his gaze towards a red LED cue presented straight ahead at 0o. The participant then pressed a button to indicate fixation and initiate stimulus presentation. While the participant maintained center fixation, the fixation light turned off. After ~1 second, a sound stimulus (AO block) or a pair of sound and light stimuli (AV block) were presented. The participant indicated the perceived location of the sound stimulus in two sequential steps: 1) rapidly directing his/her gaze toward the perceived direction of the sound source on the azimuth plane (saccade response) and 2) after the gaze shift, indicating the perceived front/back location of the sound by pressing a front/back button (choice response). Eye movements were recorded using an eye tracking system (90-Hz sampling rate, Arrington Research) and button presses were recorded using a modified gaming console, two buttons for front and back choices were arranged vertically on the game console. Participants were not provided any feedback or knowledge of their results during or after the experiments. They were unaware of the total number of loudspeakers and spatial location of each speaker.

Data Analysis

Figure 1B shows typical saccade and choice responses. Rapid saccade movements can be seen upon hearing a sound. The saccade endpoints (with a duration of ~15 ms) before a participant made a front/back choice response (green dot shows the choice RT) were averaged and used as the perceived lateral position of the sound source. The saccade velocity of eye movement after sound onset was then calculated. The time for achieving the peak velocity (red dot) was used as the saccade RT. Since the responses to lateral and front-back dimensions were made sequentially, the data analysis extracted the delayed choice RT (i.e., the difference between choice RT and saccade RT) to analyze the separate effects, if any, of vision on front-back decisions. The FBC errors were calculated by the percent of incorrect button-pressing choices a participant made when a sound was from the front, but responses were to the back or vice-versa.

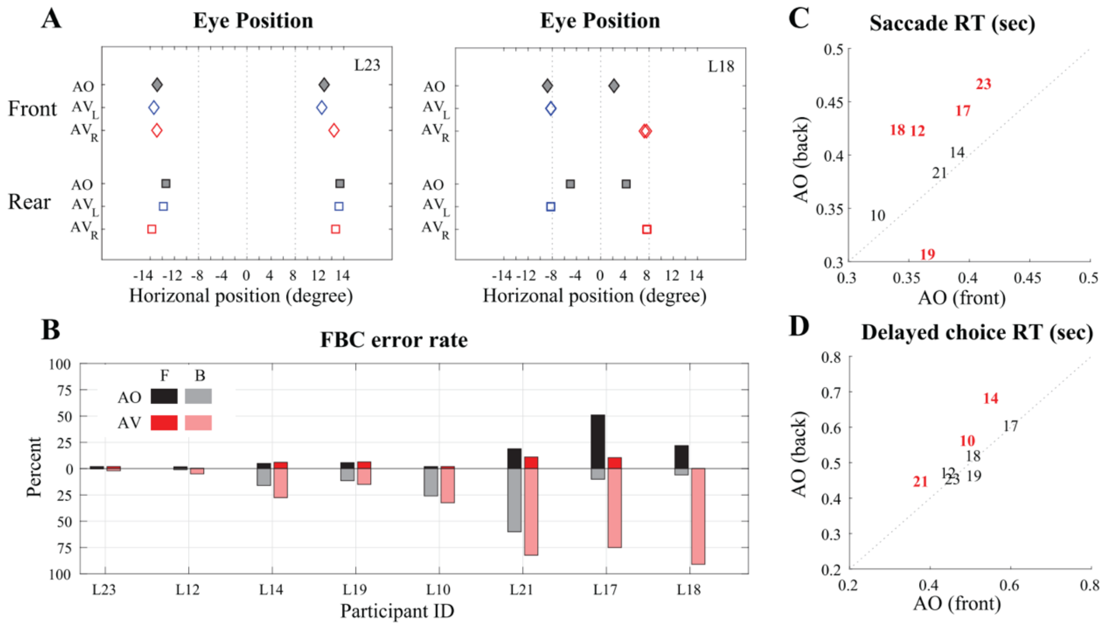

Figure 2. Auditory localization responses. (A) Saccade-based lateral judgment of sound direction with and without light. Left and right responses are associated with the panned directions of left and right sound sources, respectively. (B) Averaged FBC error rate of all participants in AO and AV conditions. (C) Saccade RT and (D) Delayed choice RT for all participants in the AO condition. Red symbols (Participant ID) indicate significant differences between RT to front and rear sound (two-tailed t-test, p<0.05)

Results

Response Accuracy

Results show that saccade eye movement can reliably indicate the perceived lateral direction of a sound source in both frontal and rear space. Even with the presence of light stimuli, the performances of five participants revealed little change in their saccade-based lateral judgment of the sound source direction. Figure 2A (left panel) shows one of these participants. On the other hand, two participants showed complete visual capture by light (Fig. 2A, right panel); their responses in the AV condition followed the light direction, dramatically shifted from their responses in the AO condition. This occurred despite repeated instructions to localize the sound and not the light. The remaining participant showed incomplete visual capture (not shown).

Individual variability is also observed in front-back judgments. Figure 2B shows the FBC rate for each of four stimulation types (sounds in front/back with or without light, 2×2 = 4) for all participants. The five participants showing limited visual influences on lateral judgments also reported limited errors for FBCs in both AO and AV conditions. Their data is shown in the left five columns. By contrast, the three participants showing strong visual capture for lateral judgments reported elevated incidences of FBC errors. Light stimulation caused strong back-to-front response reversals for rear auditory targets, resulting in large FBC error rates (pink).

Response Time

All but one participant showed longer choice RT to sound stimuli presented from the rear than from the front. However, since front-back button pressing responses were made after saccades, this delay could have occurred during the saccade. Thus, we analyzed separately the saccade RT (Fig. 2C) and the delayed choice RT (Fig. 2D); see Fig. 1A for Methods. The results reveal that four participants showed longer saccade RT to rear sounds (red in Fig. 2C), but no difference in their delayed choice RT between front and back sounds (black in Fig. 2D).

On the other hand, three participants showed longer delayed choice RT to rear sounds (red in Fig. 2D), but no difference in their saccade RT to front and rear sounds (black in Fig. 2C). For the participant who reacted faster to rear sound (L19), it is the saccade RT, not delayed choice RT, that showed a significant difference between front and rear. These results suggest that the participants we tested made their front/back decision either before or after their saccade responses. For the former group, they likely needed to realize the actual position of a sound source before making a saccade to its lateral direction, whereas for the latter group, the left-right and front-back decisions were likely made sequentially, in part, due to experimental instruction. This difference might reflect a location-based vs. hemifield-based localization strategy among participants.

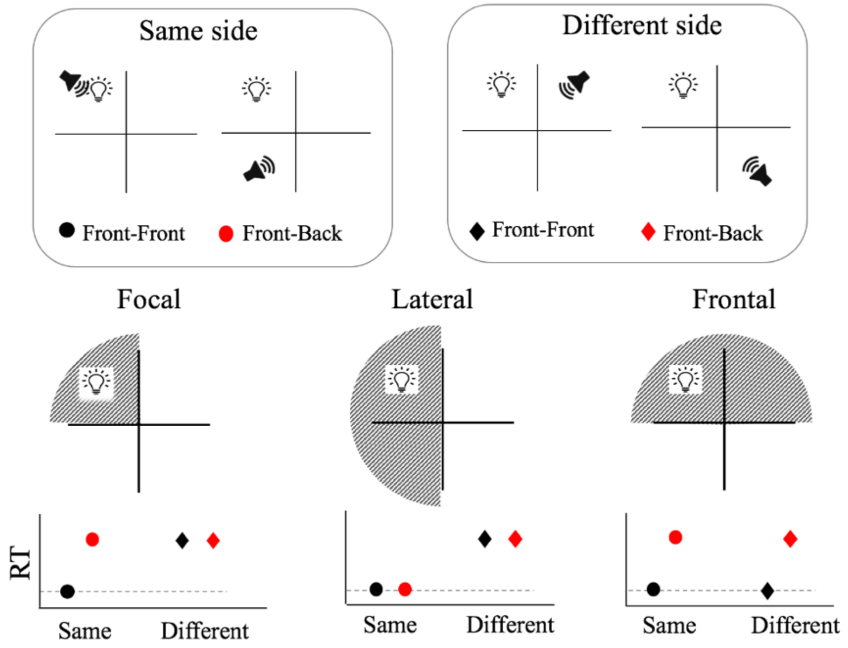

Figure 3. Hypothesis for change in RT with A-V congruent and incongruent stimulus conditions.

The existing literature shows that when sound and light stimuli are both in the frontal field, AV incongruent stimuli (on different sides) have longer RTs than AV congruent stimuli (on the same side) for auditory localization tasks (8, 9). However, little is known about the spatial extent of visual influence, or more likely, the extent of visual spatial attention, when a sound may come from either frontal or rear space. We hypothesize that three scenarios may be possible for panoramic auditory localization, as shown in Figure 3. (1) Visual-spatial attention is focal: any AV incongruity between left-right and/or front-back dimensions delays auditory response. (2) Visual-spatial attention is lateral: AV incongruity in the left-right dimension delays auditory response, independent of perceived front-back directions. (3) Visual spatial attention is frontal: AV incongruity in the front-back dimension delays auditory response, independent of perceived left-right direction.

The results show that although visual stimulation did not affect left-right auditory localization accuracy in some participants (N=4), their saccade RT to rear sound sources in AV response were significantly longer than those in AO responses (one-tailed t-test, $p<0.05$) and visually induced saccade delay was more apparent for left-right incongruent than congruent AV stimuli (Scenario 2). Figure 4A and 4B show the changes in RT between AV and AO conditions of two participants. Interestingly, delayed choice RT either showed no change or did not reveal this asymmetry. These results suggest that, when visual stimuli do not affect auditory localization responses (in terms of accuracy), visual spatial attention is lateral and acts fast for a brief period time before the saccade. On the other hand, for participants (N=3) that experienced a great extent of visual capture of response accuracy, the change in their RT follows the pattern for the frontal visual spatial attention (Scenario 3) but in the opposite direction. That is, adding visual stimulation sped up their responses for rear stimuli. Figure 4C and 4D show the changes in RT between AV and AO conditions of two participants. The asymmetry between congruent and incongruent conditions also applied; incongruent front-back AV stimuli resulted in a significant change in RT. Interestingly, this asymmetrical change is reflected in either saccade RT (L18) or delayed choice RT (L21); one-tailed t-test, $p<0.05$. This suggests that when operating in the capturing mode, visual spatial attention is frontal and acts for a longer period time extending beyond lateral judgment through a saccade.

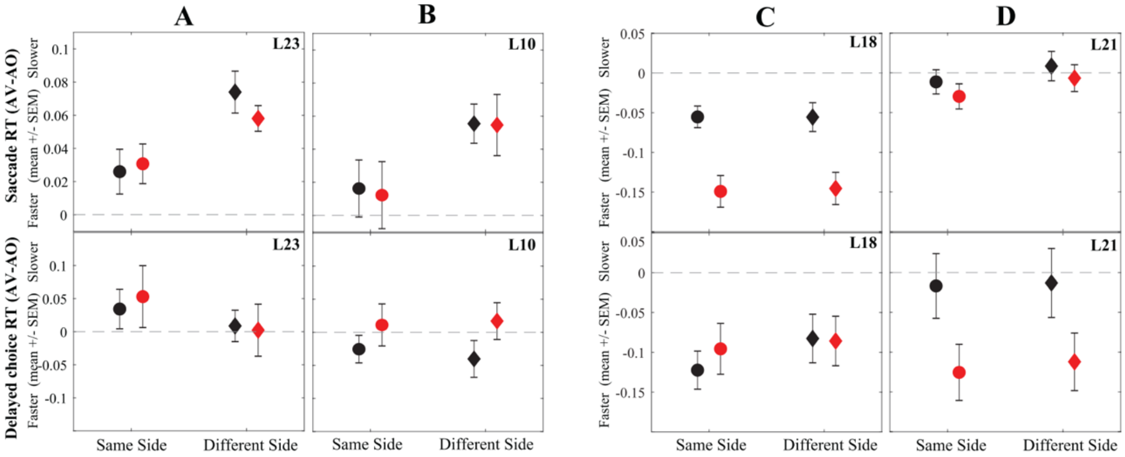

Figure 4. Changes in saccade RT and delayed choice RT between AV and AO conditions for four participants. Black circle, sound is in front and on the same side of light; red circle, sound is in back and on the same side of light; Black diamond, sound is in front and from the opposite side of light; red diamond, sound is in back and from the opposite side of light. See Figure 3 for illustration of symbols.

Summary

The results of this study show that observed visual effects do not adhere to the spatial rule of multisensory interaction regarding the physical proximity of cues. First, the influence of visual cues interacted closely with front-back confusions in auditory localization (Fig. 2B). Second, visual dominance in response time could extend to spatially incongruent auditory stimuli from the rear field, even in the absence of changes in response accuracy, e.g., L23 in Fig. 2A and Fig. 4A. When visual influences were weak and brief, the visual spatial attention channel appeared to operate in the lateral domain, delaying the auditory response from the other side. When visual influences were strong and sustained, the visual spatial attention channel appeared to operate in the front domain, attracting auditory responses from the back. These two different operating modes suggest a domain-specific difference in visual spatial attention between auditory and visual localization.

CHAPTER 5

EXPERIMENT 3: INTERAURAL TIME THRESHOLDS AND THEIR CORRELATION WITH STEREO SOUND SOURCE LOCALIZATION VARIABILITY

Introduction

Sound source localization in the horizontal plane depends heavily on the evaluation of two basic binaural cues: 1) the interaural time difference (ITD), and 2) the interaural level difference (ILD). The duplex theory, presented first by Lord Rayleigh, 1907 and substantiated in several follow-up studies, presents that ITDs are dominant for signals with frequencies below ~1000 Hz, and ILDs are dominant for frequencies above ~1500 Hz (Blauert, 1997; Macpherson & Middlebrooks, 2002; Makous & Middlebrooks, 1990; Sandel, Teas, Feddersen, & Jeffress, 1955; Stevens & Newman, 1936). ITDs represent the difference in arrival time of an acoustic signal between the two ears (created by the travel time from the sound source) while ILDs represent the difference in the level of the acoustic signal between the two ears (created by the acoustic shadowing or dampening of the sound by the head). Studies have evaluated the detection threshold for both ITDs and ILDS in humans and have shown that listeners are capable of surprisingly precise evaluation of ITD and ILD however, significant between subject variability has been seen in the results (Koehnke et al., 1995; Spencer, Hawley & Colburn, 2016). Researchers have looked to correlate these differences in ITD and ILD sensitivity with hearing sensitivity (as measured audiometrically) however, results have been inconclusive. While no clear link has been demonstrated between hearing sensitivity and sensitivity to binaural cues, individual differences threshold measures of binaural cues may still be able to provide information regarding the dependence of individuals on auditory stimuli during multisensory localization tasks.

My previous studies (Experiment 1 and Experiment 2) have demonstrated significant individual variability in terms of sound source localization in both auditory-alone and auditory-visual conditions with some subjects showing greater susceptibility to visual capture. One possible explanation to this pattern of behavior is that some participants are more reliant on auditory cues while others are more reliant on visual cues. Therefore, the purpose of this study was to investigate ITD and ILD thresholds in eight participants who had already completed the auditory-visual localization task and examine whether sensitivity to binaural cues could predict variability on the localization task and/or visual bias. The primary hypothesis was that as individual ITD and ILD threshold increased, so to would localization variability and visual bias.

Methods

Participants

Eight normal hearing participants (ages 20-38, 2 males) completed this study. All participants had previously participated in at least one of the multisensory localization studies detailed in Experiment 1 and Experiment 2. Prior to participation, hearing sensitivity was verified using standard audiometric techniques with insert earphones. All participants were right-handed and reported normal or corrected-to-normal vision. Participants provided written informed consent and received financial compensation for their participation. The experiments were conducted in accordance with procedures approved by the Arizona State University's *Institutional Review Board.*

Apparatus and Stimuli

Testing was completed under headphones (Sennheiser Pro Audio HD280PR) in a sound dampened room. The test stimulus was a 200-ms broadband noise burst with a 10 ms $\cos^2$ ramp. A professional RME Hammerfall MultiFace II sound card was used for presenting the stimuli and for the precise intensity and timing adjustment. The test stimulus was presented at a level of 65 dB SPL.

Procedure

The psychophysical method used was similar to that of Grantham et. al, 2008. Briefly, an adaptive two-interval forced-choice procedure was used, in which the stimulus was presented bilaterally in each of the two intervals, separated by 400 ms. In the first interval, the ITD (or ILD) was presented favoring one side, and in the second interval an ITD (or ILD) of the same magnitude favored the other side, with the order randomized across trials. Participants responded by pressing a button on a user interface indicating a lateral position change for which the sequence of two sound images moved (left-to-right or right-to-left). Correct answer feedback was given on pop-up text box. In the cases in which a listener could perceive no difference between the stimuli, they were instructed to guess.

A two-down one-up tracking rule was used to track the 70.7% correct performance level (Levitt, 1971). The initial ITD or ILD was set to a large value with the intention that the participant could clearly hear the lateral position changes. After a few practice runs, step-sizes were selected individually for each subject to allow an efficient tracking of threshold. The initial step-size was used for the first two reversals of level or time and could optionally be reduced to a smaller value for subsequent turnarounds. Typical values used in the ILD task were 3, 1, and 0.5 dB for the starting value, the initial step-size, and the final step-size, respectively. Typical values used in the ITD task were 500, 50 and 25 μsec, respectively. Threshold runs terminated after eight turnarounds, and the threshold ILD or ITD was computed as the average of the levels on the last six reversals.

Results

Individual and mean ITD and ILD thresholds for all participants along with the response variability and visual bias from the previous experiments are shown in Table 1. The mean ITD and ILD thresholds are consistent with the average threshold range for normal hearing listeners.

Table IIII.I: Population ITD and ILD thresholds, AO Variability, and Visual Bias

	Threshold		AO Variability		Visual bias (A+V Same)		Visual Bias (A+V Opposite	
	ILD (dB)	ITD (µsec)	Level	Timing	Level	Timing	Level	Timing
28	2	9	2.30	2.82	2.82	3.45	2.60	3.02
32	2.5	10	4.60	4.24	3.50	4.43	3.31	3.13
34	2	12	2.35	3.79	2.75	4.61	2.81	3.36
36	5	21	5.17	6.37	5.32	5.89	5.14	4.94
30	4	19	3.15	5.25	3.42	6.22	6.01	10.38
38	3	17	5.01	4.01	4.11	4.30	4.33	4.63
33	6	47	3.75	4.78	3.75	4.78	12.83	13.72
41	4.5	31	3.66	4.20	4.29	5.63	5.13	4.75
Mean	3.625	20.75	3.75	4.43	3.74	4.91	5.27	5.99

Additionally, the between participant variability observed in previous studies is also present. I hypothesized that participants who demonstrated larger ITD and ILD thresholds would show both an increase in variability in their auditory-only (AO) localization responses in the multisensory localization task as well as an increase in visual bias relative to participants with lower thresholds. Figure 1A shows the regression analysis for the ILD threshold results against AO variability, and visual bias when sound and light were on the same side of midline and opposing sides of midline for level-based stimuli. Figure 1B shows the same comparisons but for ITD threshold against timing-based stimuli. While a positive trend is seen for all comparisons, the only significant correlation is seen between threshold and visual bias in the auditory-visual opposite condition (far right column), suggesting that as ITD or ILD thresholds increase, visual bias also increases when auditory and visual information are on opposing sides of midline.

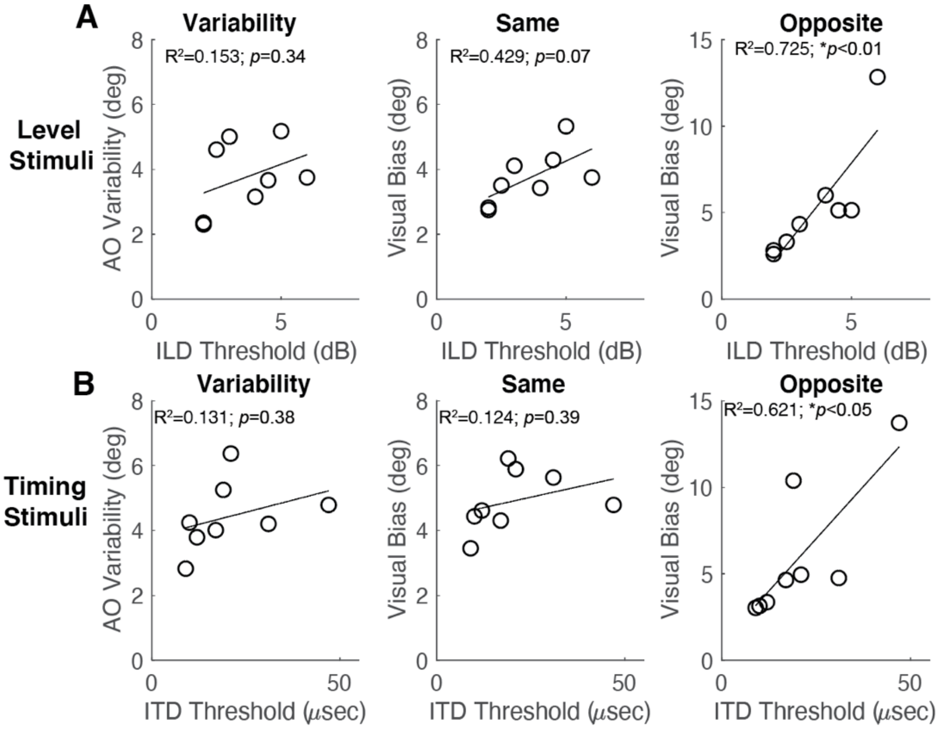

Figure 1. Between-subjects: threshold, variability, and bias. Correlation between ILD threshold (A) and ITD threshold (B) and auditory only variability (left), visual bias when light and sound are on the same side of midline (center) and visual bias when light and sounds were on opposing sides of midline (right). Each point represents one participant (n=8).

Previous work in our lab has shown that it is not ILD or ITD individually that causes response variability, but both cues together (Zhou et al., 2018). Therefore, a multiple regression was carried out to investigate whether ILD and ITD could significantly predict participants' AO variability in response to level stimuli and/or timing stimuli. The results of the regression for level variability indicated revealed that that the model explained was not significant in its prediction (R^2=0.01; p=0.34). The results of the regression for timing variability indicated revealed that that the model explained 82.6% of the variance and that the model was a significant predictor of AO variability in response to timing stimuli, F(3,4) = 17.612, p <0.01. Both ILD (B= 1.28; p<0.01) and ITD (B= 0.51; p<0.01) contributed significantly to the model.

Summary

Large degree of between-subject variability was observed in the auditory-visual localization tasks detailed in Experiments 1 and 2 with some participants showing a greater variability in localization and increased susceptibility to the influence of vision than others. I hypothesized that a potential reason for these differences in behavior was the participants ability to distinguish the basic auditory cues used for localization. ILD and ITD thresholds were measured for eight participants and the results were compared to the participants localization variability and visual bias from the previous experiments. While ILD and ITD thresholds were correlated with degree in visual bias in one condition, comparison of just ILD and ITD alone did not fully explain the variability seen in the data. However, consistent with our previous findings that it is both ILD and ITD together than influence response variability, a model incorporating both ILD and ITD thresholds was a significant predictor of response variability to timing stimuli, showing that as both ILD and ITD thresholds increased, so too did AO response variability.

CHAPTER 6

GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The purpose of this dissertation was threefold: (1) determine the effects of vision (both stimulus effects and environmental effects) on response accuracy and time of sound source localization for auditory stimuli with different degrees of ambiguity, (2) determine how stimulus spatial alignment effects vision's influence on sound source localization, and (3) identify how patterns of auditory visual interactions change between participants with normal hearing and participants with hearing impairment. In addressing these purposes experiments measured eye saccades, a natural orienting behavior in response to auditory and visual stimuli, as the primary response metric.

Experiment 1A and 1B measured eye saccade responses to evaluate the effects of visual stimuli and visual environment on the response accuracy (1A) and response time (1B) of sound source localization for stimuli with differing degrees of spatial ambiguity. Results from Experiment 1A showed that simultaneously presented visual stimuli and the visual environment both have a significant influence on the accuracy and the variability of saccade responses, especially for less salient auditory stimuli, similar to hand pointing responses reported in previous work in our lab. These results suggest that visual influences are integrated at the stage of sensory processing of auditory spatial information, and the effects of cue reliability can be revealed in two different types of motor responses (eye saccade and button pushing) at the stage of decision making. Results from Experiment 1B demonstrated that simultaneously presented visual stimuli have a significant influence on the timing of saccade responses to an auditory target, especially when the auditory and visual stimuli are not aligned in space. Collectively, the results from Experiment 1A and 1B suggest that vision has a strong influence on auditory perception, affecting both the speed and accuracy of sound source localization.

Experiment 2 further investigated the effects of visual stimulation on response accuracy and response times when listeners localized an auditory target presented from the front or rear. The results showed domain-specific effects of visual capture on both response time and response accuracy. The overall pattern of results supports previous findings that audio-visual interactions are not limited by the spatial rule of proximity.

Experiment 1C repeated the measures from Experiments 1A,B in listeners with mild to moderate hearing impairment and compared response accuracy and time results between normal hearing and hearing loss groups. Results showed that hearing impairment resulted in a significant increase in localization variability, localization response time, and visual bias relative to the normal hearing group. While compensating for the degree of hearing loss by increasing the gain of the stimulus did decrease variability, response time, and visual bias significantly compared to the uncompensated condition, results obtained under gain compensation still differed significantly from the performance of normal hearing listeners.

Future directions

Results have shown that vision has a strong influence on auditory perception. The influence of vision on the accuracy and timing of listener behavior is further magnified by hearing impairment. Further exploration of how auditory and visual stimuli interact in more complex visual and auditory environments will be critical. Listeners with hearing impairment frequently report difficulty understanding speech in busy environments (e.g., restaurant environment) even with the use of hearing aids. While improvements in hearing aid processing strategies have helped, the issue remains. Taking a multisensory approach in addressing this common concern may provide additional tools that patients and clinicians can use to improve performance in dynamic environments. Now that a basic framework has been developed to study auditory-visual interactions using natural-orienting responses, I would expand this body of work to include stimuli of complexity, with an end goal of understanding visions influence on speech understanding in multitasker environments for listeners with hearing impairment. The next steps in this pursuit will include:

1) The measure of visual behaviors in normal hearing- and hearing-impaired participants in dynamic and/or degraded visual environments to further address how vision changes following hearing impairment
2) Exploration of how to leverage vision to improve both the speed and accuracy of auditory performance

My dissertation work has given me confidence in my hypothesis that focusing on the auditory system alone will not fully address these difficulties listeners with hearing impairment experience in real-world settings. Improvement of treatment outcomes will require an understanding of how cognitive processes and other sensory systems interact with audition. My future research and training completed will continue to focus on the design and execution of hypothesis-driven research on multisensory interactions so I can better address the real-world challenges faced by listeners with hearing impairment.

REFERENCES

Aglioti, S. (1997). The role of the thalamus and basal ganglia in human cognition. Journal of Neurolinguistics, 10(4), 255–265. https://doi.org/10.1016/S0911-6044(97)00020-1

Alain, C., & Bernstein, L. J. (2008). From sounds to meaning: The role of attention during auditory scene analysis. *Current Opinion in Otolaryngology and Head and Neck Surgery*, *16*(5), 485–489. https://doi.org/10.1097/MOO.0b013e32830e2096

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration Results for the various unimodal location discriminations for naive observer. Current Biology, 14, 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Anderson, T. J., Jenkins, I. H., Brooks, D. J., Hawken, M. B., Frackowiak, R. S. J., & Kennard, C. (1994). Cortical control of saccades and fixation in man A PET study. Brain (Vol. 117). Retrieved from https://academic.oup.com/brain/article-abstract/117/5/1073/362202

Awh, E., Belopolsky, A. V, & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. https://doi.org/10.1016/j.tics.2012.06.010

Batteau, D. W. (1967). The role of the pinna in human localization. Proceedings of the Royal Society of London. Series B. Biological Sciences, 168(1011), 158–180. https://doi.org/10.1098/rspb.1967.0058

Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*(10), 904-916.

Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, *24*(4), 293–306. https://doi.org/10.3766/jaaa.24.4.5

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & psychophysics*, *29*(6), 578-584.

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

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The Physics of Optimal Decision Making: A Formal Analysis of Models of Performance in Two-Alternative Forced-Choice Tasks. https://doi.org/10.1037/0033-295X.113.4.700

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2009). The neural basis
of the speed-accuracy tradeoff. https://doi.org/10.1016/j.tins.2009.09.002

Bolognini, N., Leor, F., Passamonti, C., Stein, B. E., & L.davas, E. (2007). Multisensory-Mediated Auditory Localization. Perception, 36(10), 1477–1485. https://doi.org/10.1068/p5846

Boring, E. G. (1926). Auditory theory with special reference to intensity, volume, and localization. The American Journal of Psychology, 37, 157–188.

Broadbent, D. E. (1954). *THE ROLE OF AUDITORY LOCALIZATION IN ATTENTION AND MEMORY SPAN 1*. *Journal ol Experimental Psychology* (Vol. 47).

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *The Journal of the Acoustical Society of America*, *98*(5), 2542-2553.

Brown, M. R. G., Goltz, H. C., Vilis, T., Ford, K. A., & Everling, S. (2006). Inhibition and generation of saccades: Rapid event-related fMRI of prosaccades, antisaccades, and nogo trials. NeuroImage, 33(2), 644–659. https://doi.org/10.1016/j.neuroimage.2006.07.002

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18751–18756. https://doi.org/10.1073/pnas.0507704102

Butler, R. A., & Helwig, C. C. (1983). The spatial attributes of stimulus frequency in the median sagittal plane and their role in sound localization. *American journal of otolaryngology*, *4*(3), 165-173.

Braga, R. M., Fu, R. Z., Seemungal, B. M., Wise, R. J., & Leech, R. (2016). Eye movements during auditory attention predict individual differences in dorsal attention network activity. *Frontiers in human neuroscience*, *10*, 164.

Byrne, D., & Noble, W. (1998). Optimizing Sound Localization with Hearing Aids. Trends in Amplification, 3(2), 51. https://doi.org/10.1177/108471389800300202

Campbell, J., & Sharma, A. (2014). Cross-modal re-organization in adults with early stage hearing loss. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0090594

Carlile, S., Leong, P., & Hyams, S. (1997). The nature and distribution of errors in sound localization by human listeners. *Hearing research*, *114*(1-2), 179-196.

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The “ventriloquist effect”: Visual dominance or response bias?. *Perception & Psychophysics*, *18*(1), 55-60.

Colavita FB. (1974) Human sensory dominance. Perception & Psychophysics.;16(2):409-12.

Collins, T., Heed, T., & Röder, B. (2010). Eye-movement-driven changes in the perception of auditory space. *Attention, Perception, & Psychophysics*, *72*(3), 736-746.

Colonius, H., & Diederich, A. (2004). Multisensory Interaction in Saccadic Reaction Time: A Time-Window-of-Integration Model. Journal of Cognitive Neuroscience, 16(6), 1000–1009. https://doi.org/10.1162/0898929041502733

Colonius, H., & Diederich, A. (2010). The optimal time window of visual–auditory integration: a reaction time analysis. Frontiers in Integrative Neuroscience, 4, 11. https://doi.org/10.3389/fnint.2010.00011

Cowan, N., & Barron, A. (1987). Cross-modal, auditory-visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics*, *41*(5), 393–401. https://doi.org/10.3758/BF03203031

Crawford, T. J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., & Tetley, S. (2005). Inhibitory Control of Saccadic Eye Movements and Cognitive Impairment in Alzheimer's Disease. https://doi.org/10.1016/j.biopsych.2005.01.017

Dai, L., Best, V., & Shinn-Cunningham, B. G. (2018). Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(14), E3286–E3295. https://doi.org/10.1073/pnas.1721226115

Davis, K. A., Ramachandran, R., & May, B. J. (2003). Auditory Processing of Spectral Cues for Sound Localization in the Inferior Colliculus. JARO - Journal of the Association for Research in Otolaryngology, 4(2), 148–163. https://doi.org/10.1007/s10162-002-2002-5

Deutsch, J. A., Deutsch, D., Broadbent, D. E., Chow, K. L., Ham-Burg, D. A., & Morrell, F. (1963). *ATTENTION: SOME THEORETICAL CONSIDERATIONS 1*. *Psychological Review* (Vol. 70).

Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., & Woldorff, M. G. (2013). Cross-Modal Stimulus Conflict: The Behavioral Effects of Stimulus Input Timing in a Visual-Auditory Stroop Task. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0062802

Driver, J., & Baylis, G. vC. (1998). Attention and visual object segmentation. In *The Attentive Brain* (pp. 299–325).

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53–78. https://doi.org/10.1348/000712601162103

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in cognitive sciences*, *2*(7), 254-262.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*(3), 272–300. https://doi.org/10.1037/0033-295X.87.3.272

Dyckman, K. A., Camchong, J., Clementz, B. A., & McDowell, J. E. (2007). An effect of context on saccade-related behavior and brain activity. NeuroImage, 36(3), 774–784. https://doi.org/10.1016/j.neuroimage.2007.03.023

Eimer, M., Cockburn, D., Smedley, B., & Driver, J. (2001). Cross-modal links in endogenous spatial attention are mediated by common external locations: Evidence from event-related brain potentials. *Experimental Brain Research*, *139*(4), 398–411. https://doi.org/10.1007/s002210100773

Eimer, Martin, & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. https://doi.org/10.1017/S004857729897086X

Ettinger, U., Ffytche, D. H., Kumari, V., Kathmann, N., Reuter, B., Zelaya, F., & Williams, S. C. R. (2008). Decomposing the Neural Correlates of Antisaccade Eye Movements Using Event-Related fMRI. Cerebral Cortex May, 18, 1148–1159. https://doi.org/10.1093/cercor/bhm147

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology*, *67*, 641-666.

Foulsham, T. (2015). Eye movements and their functions in everyday tasks. In Eye (Basingstoke) (Vol. 29, pp. 196–199). Nature Publishing Group. https://doi.org/10.1038/eye.2014.275

Frens, M. A., Van Opstal, A. J., & Van Der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Perception & Psychophysics, 57(6), 802–816. https://doi.org/10.3758/BF03206796

Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42*(SUPPL. 1). https://doi.org/10.3109/14992020309074627

Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing Research*, *343*, 191–201. https://doi.org/10.1016/J.HEARES.2016.08.012

Glisky, E. L. (2019). Changes in Cognitive Function in Human Aging. In Brain Aging (pp. 3–20). CRC Press. https://doi.org/10.1201/9781420005523-1

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Haynes, D. S., & Labadie, R. F. (2008). Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+ processing. *Ear and hearing*, *29*(1), 33-44.

Haber, L., Haber, R. N., Penningroth, S., Novak, K., & Radgowski, H. (1993). Comparison of nine methods of indicating the direction to objects: Data from blind adults. *Perception*, *22*(1), 35-47.

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of cognitive neuroscience*, *15*(1), 20-29.

Hanauer, J. B., & Brooks, P. J. (2003). Developmental change in the cross-modal Stroop effect. *Perception and Psychophysics*, *65*(3), 359–366. https://doi.org/10.3758/BF03194567

Hanes, D. P., & Wurtz, R. H. (2001). Interaction of the Frontal Eye Field and Superior Colliculus for Saccade Generation. Journal of Neurophysiology, 85(2), 804–815. https://doi.org/10.1152/jn.2001.85.2.804

Hausler, R., Colburn, S., & Marr, E. (1983). Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests. Acta Oto-Laryngologica. Supplementum, 400, 1–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6316714

Heitz, R. P. (2014, June 11). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. Frontiers in Neuroscience. Frontiers Research Foundation. https://doi.org/10.3389/fnins.2014.00150

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006, February). Basal ganglia orient eyes to reward. Journal of Neurophysiology. American Physiological Society. https://doi.org/10.1152/jn.00458.2005

Horn, D. L., Davis, R. A. O., Pisoni, D. B., & Miyamoto, R. T. (2005). Development of visual attention skills in prelingually deaf children who use cochlear implants. *Ear and Hearing*, *26*(4), 389–408. https://doi.org/10.1097/00003446-200508000-00003

Howard, I. P., & Templeton, W. B. (1966). Human spatial orientation.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. Perceptual and motor skills, 37(3), 967-979.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*(2), 52-65.

Jeffress, L. A. (1948). A place theory of sound localization. Journal of Comparative and Physiological Psychology, 41(1), 35–39. https://doi.org/10.1037/h0061495

Jerger, S., Martin, R. C., & Pirozzolo, F. J. (1988). A developmental study of the auditory stroop effect. *Brain and Language*, *35*(1), 86–104. https://doi.org/10.1016/0093-934X(88)90102-2

Johnson, B. W., & Hautus, M. J. (2010). Processing of binaural spatial information in human auditory cortex: neuromagnetic responses to interaural timing and level differences. *Neuropsychologia*, *48*(9), 2610-2619.

Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160101. https://doi.org/10.1098/rstb.2016.0101

Klein, C., & Ettinger, U. (2019). Eye Movement research summary. https://doi.org/https://doi.org/10.1007/978-3-030-20085-5

Knight, S., & Heinrich, A. (2017). Different measures of auditory and visual stroop interference and their relationship to speech intelligibility in noise. *Frontiers in Psychology*, *8*(MAR). https://doi.org/10.3389/fpsyg.2017.00230

Koehnke, J., Culotta, C. P., Hawley, M. L., & Colburn, H. S. (1995). Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear and hearing*, *16*(4), 331-353.

Lange, L. (1888). New experiments on the process of the simple reaction to sensory impressions. *Philosophische Studien*, *4*, 479.

Lavie, N. (2000). Selective attention and cognitive control: Dissociating attentional functions through different types of load. In *Attention and performance XVIII* (pp. 175–194). Retrieved from https://books.google.com/books?hl=en&lr=&id=kO_baYlSVbwC&oi=fnd&pg=PA175&dq=Selective+attention+and+cognitive+control:+Dissociating+attentional+functions+through+different+types+of+load.+In&ots=ps6KMLerU0&sig=zmHJSQvPbBIfliu1MRTaYfZyVq4

Lavie, N. (2005). Distracted and confused?: Selective attention under load. https://doi.org/10.1016/j.tics.2004.12.004

Lavie, N., & De Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception and Psychophysics*, *65*(2), 202–212. https://doi.org/10.3758/BF03194795

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–197. https://doi.org/10.3758/BF03213897

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467-477.

Lewald, J., & Ehrenstein, W. H. (1998). Auditory-visual spatial integration: a new psychophysical approach using laser pointing to acoustic targets. *The Journal of the Acoustical Society of America*, *104*(3), 1586-1597.

Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *105*(6), 3454–3463. https://doi.org/10.1121/1.424672

Luce, R. D., & Luce, R. D. (1991). Representing Response Times as Random Variables. Response Times, 1–46. https://doi.org/10.1093/acprof:oso/9780195070019.003.0001

Lukas, S., Philipp, A. M., & Koch, I. (2010). Switching attention between modalities: further
evidence for visual dominance. Psychological Research PRPF, 74(3), 255–267. https://doi.org/10.1007/s00426-009-0246-y

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle:
The duplex theory of sound localization revisited. The Journal of the Acoustical Society
of America, 111(5), 2219. https://doi.org/10.1121/1.1471898

Maioli, C., & Falciati, L. (2015). Covert preparation of a manual response in a ‘go’/‘no-go’ saccadic task is driven by execution of the eye movement and not by visual stimulus occurrence. *Frontiers in Human Neuroscience*, *9*(OCT). https://doi.org/10.3389/fnhum.2015.00556

Majdak, P., Goupell, M. J., & Laback, B. (2010). 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, perception, & psychophysics*, *72*(2), 454-469.

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human
listeners. The Journal of the Acoustical Society of America, 87(5), 2188–2200. https://doi.org/10.1121/1.399186

Mason, R., Ford, N., Rumsey, F., & De Bruyn, B. (2001). Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction. *Journal of the Audio Engineering Society*, *49*(5), 366-384.

Mastroianni, G. R. (1982). The influence of eye movements and illumination on auditory localization. *Perception & psychophysics*, *31*(6), 581-584.

Mcdowell, J. E., Dyckman, K. A., Austin, B. P., & Clementz, B. A. (2008). Neurophysiology
and neuroanatomy of reflexive and volitional saccades: Evidence from studies of humans. https://doi.org/10.1016/j.bandc.2008.08.016

Middlebrooks, J. C. (1999). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, *106*(3), 1493-1510.

Middlebrooks, J. C., & Green, D. M. (1991). SOUND LOCALIZATION BY HUMAN LISTENERS. Retrieved from www.annualreviews.org

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. Cognitive
Psychology, 14(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Montagne, C., & Zhou, Y. (2016). Visual capture of a stereo sound: Interactions between cue reliability, sound localization variability, and cross-modal bias. *The Journal of the Acoustical Society of America*, *140*(1), 471-485.

Montagne, C., & Zhou, Y. (2018). Audiovisual interactions in front and rear space. *Frontiers in psychology*, *9*, 713.

Moray, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Morey, ., Morey, C., Eaves, R. D., Shelton, S., & Lufti-Proctor, J. &. (2014). The role of modality: Auditory and visual distractors in Stroop interference. *Journal of Cognitive Psychology*, *26*(1), 15. https://doi.org/10.1080/20445911.2013.859133

Munoz, D. P. (2002). Commentary: Saccadic eye movements: overview of neural circuitry. Progress in Brain Research, 140, 89–96. https://doi.org/10.1016/S0079-6123(02)40044-1

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. Nature Reviews Neuroscience, 5(3), 218–228. https://doi.org/10.1038/nrn1345

Musicant, A. D., & Butler, R. A. (1984a). The influence of pinnae-based spectral cues on sound localization. The Journal of the Acoustical Society of America, 75(4), 1195–1200. https://doi.org/10.1121/1.390770

Musicant, A. D., & Butler, R. A. (1984b). The psychophysical basis of monaural localization. Hearing Research, 14(2), 185–190. https://doi.org/10.1016/0378-5955(84)90017-0

Neill, W. T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 444–450. https://doi.org/10.1037/0096-1523.3.3.444

Norman, D. A., & Bobrow, D. G. (1975). *On Data-limited and Resource-limited Processes*. *COGNITIVE PSYCHOLOGY* (Vol. 7).

Oldfield, S. R., & Parker, S. P. A. (1984). Acuity of Sound Localisation: A Topography of Auditory Space. I. Normal Hearing Conditions. Perception, 13(5), 581–600. https://doi.org/10.1068/p130581

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*(4), 203-205.

Pinek, B., & Brouchon, M. (1992). Head turning versus manual pointing to auditory targets in normal subjects and in subjects with right parietal damage. *Brain and cognition*, *18*(1), 1-11.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological review*, *83*(2), 157.

Proksch, J., & Bavelier, D. (2002). Changes in the spatial distribution of visual attention after early deafness. *Journal of Cognitive Neuroscience*, *14*(5), 687–701. https://doi.org/10.1162/08989290260138591

Raab, D. H. (1962). DIVISION OF PSYCHOLOGY: STATISTICAL FACILITATION OF SIMPLE REACTION TIMES*. Transactions of the New York Academy of Sciences, 24(5 Series II), 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Rae, B., Heathcote, A., Donkin, C., Averell, L., Brown, S., & Wales, S. (2014). The Hare and the Tortoise: Emphasizing Speed Can Change the Evidence Used to Make Decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(5), 1226–1243. https://doi.org/10.1037/a0036801

Ragot, R., Cave, C., & Fano, M. (1988). Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation. Bulletin of the Psychonomic Society, 26(4), 350–352.https://doi.org/10.3758/BF03337679

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-64 Choice Reaction Time. Psychological Review, 111(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Rayleigh, Lord. (1875). On Our Perception of the Direotion of a Source of Sound. Proceedings of the Musical Association, 2(1), 75–84. https://doi.org/10.1093/jrma/2.1.75

Rinkenauer, G., Osman, A., Ulrich, R., Müler-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. Journal of Experimental Psychology: General, 133(2), 261–282.https://doi.org/10.1037/0096-3445.133.2.261

Rinkenauer, G., Osman, A., Ulrich, R., Müler-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readinesspotential. Journal of Experimental Psychology: General, 133(2), 261–282. https://doi.org/10.1037/0096-3445.133.2.261

Roelofs, A. (2005). The visual-auditory color-word Stroop asymmetry and its time course. *Memory and Cognition*, *33*(8), 1325–1336. https://doi.org/10.3758/BF03193365

Roffler, S. K., & Butler, R. A. (1968). Factors That Influence the Localization of Sound in the Vertical Plane. The Journal of the Acoustical Society of America, 43(6), 1255–1259.https://doi.org/10.1121/1.1910976

Sandel, T. T., Teas, D. C., Feddersen, W. E., & Jeffress, L. A. (1955). Localization of Sound from Single and Paired Sources. The Journal of the Acoustical Society of America, 27(5),842–852. https://doi.org/10.1121/1.1908052

Schroger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology, 35(6), 755–759. https://doi.org/10.1111/1469-8986.3560755

Seeber, B. (2002). Localization in the anechoic environment and with virtual acoustical directional presentation investigated with a laser-pointing method. In *Proc. of ARO 2002 (Abstracts), 25. Annual Midwinter Research Meeting, St. Petersburg Beach, Florida, USA*.

Shelton, B. R., & Searle, C. L. (1980). The influence of vision on the absolute identification of sound-source position. Perception & Psychophysics, 28(6), 589–596. https://doi.org/10.3758/BF03198830

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. https://doi.org/10.1016/j.tics.2008.02.003

Shinn-Cunningham, B. G., & Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends in Amplification*, *12*(4), 283–299. https://doi.org/10.1177/1084713808325306

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception and Psychophysics*, *69*(5), 673–686. https://doi.org/10.3758/BF03193770

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7-10.

Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception and Psychophysics*, *59*(3), 389–403. https://doi.org/10.3758/BF03211906

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. Neuroreport, 11(9), 2057-2061.

Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford University Press

Spencer, N. J., Hawley, M. L., & Colburn, H. S. (2016). Relating interaural difference sensitivities for several parameters measured in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *140*(3), 1783-1799.

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. MIT Press.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. Journal of Cognitive Neuroscience, 1(1), 12–24. https://doi.org/10.1162/jocn.1989.1.1.12

Stevens, S. S., & Newman, E. B. (1936). The Localization of Actual Sources of Sound. The American Journal of Psychology, 48(2), 297. https://doi.org/10.2307/1415748

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*, 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*(1), 83–90. https://doi.org/10.3758/BF03211619

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the “ventriloquism effect”. *Perceptual and motor skills*, *36*(3_suppl), 1171-1184.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 571–590. https://doi.org/10.1080/14640748508400920

Tomko, L., & Proctor, R. W. (2017). Crossmodal spatial congruence effects: visual dominance in conditions of increased and reduced selection difficulty. Psychological Research, 81(5), 1035–1050. https://doi.org/10.1007/s00426-016-0801-2

Treisman, A. M. (1960). Contextual Cues in Selective Listening. *From Perception to Consciousness*, 20–23. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0004

Treisman, A. M. (1969). *STRATEGIES AND MODELS OF SELECTIVE ATTENTION **. *Psychological Review* (Vol. 76).

Trujillo, M., & Razak, K. A. (2013). Altered cortical spectrotemporal processing with age-related hearing loss. *Journal of Neurophysiology*, *110*(12), 2873–2886. https://doi.org/10.1152/jn.00423.2013

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., & Wouters, J. (2006). Horizontal localization with bilateral hearing aids: without is better than with. The Journal of the Acoustical Society of America, 119(1), 515–526. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16454305

Van Ee, R., Van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, *29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

Van Veen, V., Krug, M. K., & Carter, C. S. (2008). The Neural and Computational Basis of Controlled Speed-Accuracy Tradeoff during Task Performance.

Venskytis, E.J., Clayton, C., Montagne, C., Zhou, Y., (2019) Audiovisual Interactions in Stereo Sound Localization for Individuals With Unilateral Hearing Loss. Trends in Hearing, 23,1-17. https://doi.org/ 10.1177/2331216519846232

Vliegen, J., Van Grootel, T. J., & Van Opstal, A. J. (2004). Dynamic sound localization during rapid eye-head gaze shifts. *Journal of Neuroscience*, *24*(42), 9291-9302.

Wallace, M. T., Perrault, T. J., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *Journal of Neuroscience*, *24*(43), 9580-9584.

Warren, D. H. (1970). Intermodality interactions in spatial localization. Cognitive Psychology, 1, 114–133.

Welch, R. B., Warren, D. H., Bertelson, P., Easton, R. D., Frcides, D., Gogel, W. C., … Smothergill, D. W. (1980). *Immediate Perceptual Response to Intersensory Discrepancy*. *Psychological Bulletin* (Vol. 88).

Wightman, F. L., & Kistler, D. J. (1993). Sound localization. In *Human psychophysics* (pp. 155-192). Springer, New York, NY.

Yang, C.-T., Altieri, N., & Little, D. R. (2018). An examination of parallel versus coactive processing accounts of redundant-target audiovisual signal processing. Journal of Mathematical Psychology, 82, 138–158. https://doi.org/10.1016/J.JMP.2017.09.003

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

Yost, W. A., Zhong, X., & Najam, A. (2015). Judging sound rotation when listeners and sounds rotate: Sound source localization is a multisystem process. The Journal of the Acoustical Society of America, 138(5), 3293-3310.

Zhou, Y., Balderas, L., & Venskytis, E. J. (2018). Binaural ambiguity amplifies visual bias in sound source localization. *The Journal of the Acoustical Society of America*, *144*(6), 3118-3123.

APPENDIX A

GROUP COMPARISONS OF AO VARIABILITY, VISUAL BIAS AND RESPONSE TIME

Table III.II: Groups comparisons of AO Variability and Visual Bias

ANOVA	AO Variability		Visual Bias (same)		Visual Bias (opposite)	
	Level	Timing	Level	Timing	Level	Timing
Group	F=6.46; p<0.01	F=17.7; p<0.001	F=13.07; p<0.001	F=9.8; p<0.001	F=7.22; p<0.01	F=25.06; p<0.001
Uncompensated vs. Compensated	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.01
Uncompensated vs. Normal Hearing	p<0.01	p<0.01	p<0.01	p<0.01	p<0.05	p<0.001
Compensated vs. Normal Hearing	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05	p<0.01

Table III.III: Groups comparisons of AO and VO RT

ANOVA	AO Start RT		AO Fixation RT		VO RT
	Level	Timing	Level	Timing	
Group	F=6.16; p<0.01	F=8.23; p<0.01	F=16.71; p<0.001	F=15.24; p<0.002	F=1.39; p=0.27
Uncompensated vs. Compensated	p=0.5	p=0.49	p=0.34	p=0.5	
Uncompensated vs. Normal Hearing	p<0.01	p<0.01	p<0.001	p<0.001	
Compensated vs. Normal Hearing	p<0.01	p<0.05	p<0.01	p<0.01	